NIST Internal Report
NIST IR 8607

Workshop Summary Report for "Cyber AI Profile" Hybrid Workshop #2

Katerina N. Megas
Barbara Cuthill
Christina Sames
Julie Nethery Snyder

This publication is available free of charge
from: https://doi.org/10.6028/NIST.IR.8607

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

NIST Internal Report
NIST IR 8607

Workshop Summary Report for "Cyber AI Profile" Hybrid Workshop #2

Katerina N. Megas
Barbara Cuthill
Applied Cybersecurity Division
Information Technology Laboratory

Christina Sames
Julie Nethery Snyder
National Cybersecurity Center of Excellence
MITRE

This publication is available free of charge
from: https://doi.org/10.6028/NIST.IR.8607

August 2026

U.S. Department of Commerce
Howard Lutnick, Secretary

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

Publication History
Approved by the NIST Editorial Review Board on 2026-05-14

How to Cite this NIST Technical Series Publication
Cuthill B, Megas K, Sames C, Snyder J (2026) Workshop Summary Report for “Cyber AI Profile” Hybrid Workshop #2. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Internal Report (NIST IR) NIST IR 8607. https://doi.org/10.6028/NIST.IR.8607

Author ORCID iDs
Barbara Cuthill: 0000-0002-2588-6165
Katerina Megas: 0000-0002-2815-5448
Christina Sames: 0009-0003-1817-8333
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/8607/final, including related content, potential updates, and document history.

Abstract

This report summarizes the discussion at the second “Cyber AI Profile Workshop,” in January 20506 during which focused on the Preliminary Draft of the NIST Cybersecurity Framework Profile for Artificial Intelligence (Cyber AI Profile). The workshop gathered input across government, industry, and academia to provide guidance for managing cybersecurity risks associated with AI systems and for leveraging AI to enhance cybersecurity capabilities to inform the next draft of the NIST Cyber AI Profile. NIST also provided an update on its work to develop the Special Publication (SP) 800-53 Control Overlays for Securing AI Systems (COSAiS). The report highlights key themes raised during the discussions, including governance challenges, stability of the Profile over time, AI attack surfaces, the need for consistent AI taxonomy, ideas for specific risk-based guidance and resources to support usability (e.g., use cases), and AI-enabled cyber defense opportunities.

Keywords

Artificial Intelligence; Community Profile; Cyber AI Profile; Cybersecurity Framework; Control Overlays; Control Overlays for Securing AI Systems.

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 the development of management, administrative, technical, and physical standards and guidelines for the cost-effective security and privacy of 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 to the Cyber AI Profile Workshop #2. Thank you to the participants and speakers at the Cyber AI Profile Hybrid Workshop #2, as well as our colleagues at NIST, the NIST NCCoE, and The MITRE Corporation, who offered the expertise necessary to conduct this workshop. Additionally, the thoughtful feedback stakeholders shared leading up to the event, including comments submitted to NIST on the Cybersecurity Framework Profile for Artificial Intelligence (Cyber AI Profile) Preliminary Draft, were instrumental in shaping discussions at the workshop and this workshop summary.

1. Introduction

On January 14, 2026, the National Institute of Standards and Technology (NIST) National Cybersecurity Center of Excellence (NCCoE) hosted the public workshop *Cyber AI Profile Workshop #2* to gather feedback on the Preliminary Draft of the NIST Cybersecurity Framework Profile for Artificial Intelligence (Cyber AI Profile) [1] published on December 16, 2025. This document provides a summary of the workshop's discussions, including opening remarks, panel, morning presentations, and afternoon breakout sessions. This workshop aimed to raise awareness of artificial intelligence (AI) and cybersecurity projects at NIST and obtain feedback on priority areas for organizations integrating AI into their cybersecurity infrastructure. The workshop provided a forum for stakeholders to provide input to the Cyber AI Profile, including usability, scope, content, and alignment with existing cybersecurity and AI resources. 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, but also new and modified cybersecurity risks and impacts that need to be managed. NIST is extending and adapting 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. The Cyber AI Profile is being developed in response to feedback from the cybersecurity and AI communities, indicating that there would be value in guidelines based on the CSF 2.0. The community was clear that there are opportunities for using AI to enhance cybersecurity capabilities, as well as new considerations for managing cybersecurity risk related to AI systems.

Organizations vary in how they use AI. 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 AI capabilities, such as Generative AI (genAI) or Agentic AI. Regardless of where organizations are on their AI journey, they need risk management approaches that support the realities of advancements in AI use 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 [1] to identify how organizations can leverage and adapt the CSF 2.0 [2] to manage AI-related cybersecurity risks effectively. 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 #2

The workshop's aim was to gather community feedback on the preliminary draft of the Cyber AI Profile to shape its final direction and content. By engaging stakeholders across sectors, NIST aims to ensure the development of meaningful guidelines that support organizations in managing AI-related cybersecurity risks while capitalizing on AI's capabilities.

The preliminary draft of the Cyber AI Profile [1] is the culmination of NIST research as well as an open dialogue with the public. Building on the Cybersecurity and AI Workshop Concept Paper [3] that outlined NIST's initial observations and proposed approaches, NIST collected additional feedback through Cyber AI Workshop #1 and the August and September 2025 Community of Interest working sessions to develop the preliminary draft. The Cyber AI Profile Workshop #2 agenda (see Table 1) and planned discussions focused on addressing open questions about the draft and on seeking additional public input.

The workshop also provided background on NIST's portfolio of AI and cybersecurity-related projects, including the Control Overlays for Securing AI Systems (COSAiS) [4] work.

1.3. About the Cyber AI Profile Workshop #2

The agenda for the one-day free, public, hybrid workshop is provided in Table 1. The workshop included a panel discussion and presentations on AI and cybersecurity projects at NIST. The morning portion of the workshop was open to both in-person and livestream audiences, followed by three one-hour sessions of four in-person breakout session tracks related to the three Cyber AI Profile Focus Areas (i.e., Securing AI System Components [Secure], Conducting AI-enabled Cyber Defense [Defend], and Thwarting AI-enabled Cyber Attacks [Thwart]) and COSAiS.

Table 1. Cyber AI Profile Hybrid Workshop #2 Agenda

Time	Activity and Presenters
9:00 – 9:10	**Welcome and Opening Remarks: Cherilyn Pascoe**, Director of the NIST National Cybersecurity Center of Excellence (NCCoE)
9:10 – 9:30	**About the Cyber AI Profile Project and Workshop #2: Katerina Megas**, Program Manager for the NIST Cybersecurity, Privacy, and AI Programs and Co-Lead for Cyber AI Profile
9:30 – 10:30	**Panel: AI and Cybersecurity Projects at NIST (Moderator:** Barbara Cuthill, Co-Lead for the Cyber AI Profile, NIST) • AI Risk Management Framework (AI RMF) (Martin Stanley, NIST) • Center for AI Standards and Innovation (CAISI) (Maia Hamin, NIST) • Adversarial Machine Learning (Apostol Vassilev, NIST) • Dioptra (Harold Booth, NIST)

Time	Activity and Presenters
	• Secure Software Development Framework (SSDF) AI Profile (Harold Booth, NIST) • PETs Test Bed (Gary Howarth, NIST) • Agent Identities (Ryan Galluzzo, NIST) • NCCoE Chatbot (Daniel Lee, NIST NCCoE/MITRE)
10:45 – 11:00	**Presentation: AI Accelerators: Dr. Craig Schlenoff**, Senior Advisor for Artificial Intelligence, NIST
11:00 – 11:30	**Presentation: Overview of the Cyber AI Profile Preliminary Draft: Julie Snyder**, NCCoE/MITRE, Principal, Cyber AI Profile Task Lead
11:30 – 11:50	**Presentation: Control Overlays for Securing AI Systems (COSAiS): Victoria Pillitteri**, Manager, Security Engineering and Risk Management Group, NIST

1:00 – 4:30	**Breakout Sessions**
	• Track A: Secure Focus Area
	  ○ Presenter: Noah Schiro, NCCoE/MITRE
	  ○ Facilitator: Christina Sames, NCCoE/MITRE
	  ○ Track Analyst: Michael Ekstrom, NCCoE/MITRE
	• Track B: Defend Focus Area
	  ○ Presenter: Keith Manville, NCCoE/MITRE
	  ○ Facilitator: Jonathan Keisler, NCCoE/MITRE
	  ○ Track Analyst: Ishika Khemani, NCCoE/MITRE
	• Track C: Thwart Focus Area
	  ○ Presenter: Marissa Dotter, NCCoE/MITRE
	  ○ Facilitator: John Dombrowski, NCCoE/MITRE
	  ○ Track Analyst: Daniel Lee, NCCoE/MITRE
	• Track D: COSAiS
	  ○ Presenter: Vicky Pillitteri, NIST
	  ○ Facilitator: Alicia Dawson, NCCoE/MITRE
	  ○ Track Analyst: Bryce Kearney, NCCoE/MITRE

4:40 – 5:00	**Workshop Close-out and Next Steps** • Katerina Megas, NIST • Noah Schiro, NCCoE/MITRE • Ishika Khemani, NCCoE/MITRE • Marissa Dotter, NCCoE/MITRE • Alicia Dawson, NCCoE/MITRE

The workshop provided a platform for exploring AI cybersecurity opportunities and challenges, and the potential role of the Cyber AI Profile in accelerating AI adoption by giving organizations confidence to address these challenges. Discussions during the workshop were intended to gather input from the community on securing AI components, leveraging AI for cyber defense, and thwarting AI-enabled attacks. Throughout the day, attendees engaged in dynamic conversations while speakers and panelists highlighted the importance of collaboration between the AI and cybersecurity communities to shape actionable implementation guidelines.

Materials from the workshop are available on the event page.[1] Post-event materials include a recording of the morning sessions and the presentation slides for the morning sessions and afternoon breakouts.[2]

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 to help organizations strategically adopt AI while addressing and prioritizing cybersecurity risks stemming from its advancements. Discussions included addressing risk management through a multidisciplinary approach enabled by strategic Focus Areas, as well as identifying challenges to AI-readiness and associated opportunities and risks, and forward-leaning approaches to address both. The subsections that follow summarize each presentation and discussion on the agenda.

2.1. Opening Remarks

The workshop began with opening remarks from Ms. Cherilyn Pascoe, Director of the National Cybersecurity Center of Excellence (NCCoE), and Ms. Katerina Megas, Program Manager for the NIST Cybersecurity, Privacy, and AI Programs and Lead for Cyber AI Profile. Each speaker emphasized that:

- Collaboration is important in NIST's efforts to address the intersection of cybersecurity and AI.
- The NCCoE's mission is to accelerate the adoption of secure technologies like AI to meet industry and public needs amid rapid change, new capabilities, and emerging cybersecurity risks.
- NIST guidelines only work if implementers who are building and using these technologies find the guidelines effective and applicable in the real world.
- The Cyber AI Profile is intended to help bridge taxonomy gaps, enabling clearer and consistent communication and reflecting a common view of AI activities to help organizations inform both their current and desired states.
- Collaboration and stakeholder feedback are critical to shape practical, impactful guidelines to help organizations prioritize their efforts when addressing cybersecurity challenges. Feedback on Cyber AI Profile and its usability will shape the development of a roadmap to further advance cybersecurity and AI guidance from NIST.

2.1.1. Welcome and Opening Remarks

Ms. Pascoe welcomed attendees and expressed appreciation for their engagement and collaboration. She reiterated that practical implementation guidance informed by community insights is essential to ensuring the Cyber AI Profile remains effective in addressing evolving AI

[1] The event page is available at https://www.nccoe.nist.gov/get-involved/attend-events/cyber-ai-workshop-2/post-event-materials.
[2] The afternoon breakout sessions were conducted in-person only and were not recorded.

capabilities and threats. She provided an overview of NIST's and the NCCoE's missions, which includes developing voluntary guidance on pressing cybersecurity challenges such as data protection, trusted enterprise, cybersecurity, and AI, as well as resilient embedded systems, through collaborative engagements with industry, government, and academia. Ms. Pascoe discussed how Cybersecurity Framework Community Profiles, such as the Cyber AI Profile, use the CSF to help organizations within a community understand common cybersecurity concerns and where to focus their efforts to prioritize and address cybersecurity opportunities and challenges. Ms. Pascoe noted that Profiles can also serve as a listening mechanism for NCCoE to understand shared opportunities, challenges, and concerns that reflect the community's expertise.

Ms. Pascoe concluded by acknowledging the importance of community input to develop impactful guidelines, provide insights into challenges they may be experiencing, and identify where guidance is unclear or where additional efforts could make an impactful difference in securing, using, and defending AI, resulting in a better Profile overall.

2.1.2. Setting the Stage for the Cyber AI Profile

Ms. Katerina Megas, Program Manager for the NIST Cybersecurity, Privacy, and AI Programs and Co-Lead for Cyber AI Profile, shared the Cyber AI Profile's background and roadmap, the goals and structure of the workshop, and how the Profile integrates with NIST's other AI programs and efforts, to include understanding how the broad adoption of AI may impact existing cybersecurity and privacy risks and risk management approaches.

Ms. Megas shared the gaps heard from the cybersecurity community and how their insights shaped the Cyber AI Profile's Focus Areas. The cybersecurity community wanted to understand opportunities to adopt AI, implications associated with those opportunities, and how AI is changing existing risk management measures. She observed that, in many organizations, leadership has made it a priority to understand how AI is being adopted and used, yet organizations are finding it challenging to think strategically about AI while balancing day-to-day responsibilities. The community also shared that the lack of a common taxonomy across AI publications and documentation is making it difficult for AI consumers and users to establish a consistent, clear understanding of how AI might impact them. Ms. Megas noted that the community expressed a desire to "not reinvent the wheel," to build on existing frameworks and practices, identify gaps, and determine what is new or needs to be created to address those gaps.

Ms. Megas concluded with a preview of the workshop's afternoon breakout session tracks.

2.2. Panel: AI and Cybersecurity Projects at NIST

The panel, moderated by Ms. Barbara Cuthill, NIST Co-Lead for the Cyber AI Profile, provided an overview of a selection of NIST AI projects related to cybersecurity and highlighted how these efforts connect with the Cyber AI Profile and related NCCoE activities.

2.2.1. Artificial Intelligence Risk Management Framework (AI RMF) (Mr. Martin Stanley)

Mr. Stanley stated that after releasing the Artificial Intelligence Risk Management Framework (AI RMF 1.0) [5] in 2023, the AI RMF team has maintained a long engagement with the community on what cybersecurity for AI systems means. Mr. Stanley emphasized that a key aspect included in the AI RMF is discussion on trustworthiness, and in the AI RMF, NIST is trying to understand and document how AI operates differently in a variety of familiar tasks. Mr. Stanley stated that NIST wants to understand how the community is using AI so that NIST can better identify the measures and approaches needed to establish trust in AI use. Mr. Stanley closed by stating that the AI RMF [5] looks at ways to establish consistency with the *NIST Strategy for American Technology Leadership in the 21st Century,* such that there is trust integrated into those technologies, practices, and AI uses [6].

2.2.2. Center for AI Standards and Innovation (CAISI) (Ms. Maia Hamin)

Ms. Hamin provided an overview of the Center for AI Standards and Innovation (CAISI) [7], which was reorganized and rechartered with a broader mission in early 2025. CAISI leads evaluations and assessments of United States (U.S.) and adversary AI systems, develops best practices and standards for improving AI security, and advances U.S. AI standards internationally. Ms. Hamin shared the activities completed to date to achieve these goals, highlighting work related to securing AI agents and agent systems and the security threats, risks, and challenges they face, which could hinder adoption and pose risks to public safety and national security. CAISI also develops voluntary practices to include automated benchmark evaluations of language models and AI agent systems [8]. Sharing such AI model evaluations on the state of model security, Ms. Hamin noted, can help consumers and decision-makers understand the relative security of these models. Ms. Hamin highlighted where CAISI continues to help improve the security of U.S. AI systems through the discovery of vulnerabilities and the development of guidelines and best practices. Ms. Hamin concluded by sharing how the community can get involved in CAISI, such as providing input on CAISI's recently published "Request for Information Regarding Security Considerations for Artificial Intelligence Agents" [9].

2.2.3. Adversarial Machine Learning (Dr. Apostol Vassilev, NIST)

Mr. Vassilev opened with an overview of areas considered part of adversarial AI, noting that trends have emerged that show a transition from static chatbots to agentic AI, which both give advice and take actions that can help automate workflows. Mr. Vassilev stated that while agentic AI can improve productivity, it simultaneously increases the attack surface available to attackers to exploit. Mr. Vassilev noted that over the last three to five years, there has been a recurring pattern where a large language model (LLM) is released and, within days of its release, reports emerge of successful attacks against that LLM. To understand why this pattern exists, Mr. Vassilev shared that NIST performed rigorous mathematical analysis on the robustness of guardrail mechanisms intended to thwart attacks on AI models. Results from this analysis led NIST to conclude that there are no guardrails universally robust enough to

withstand adversarial attacks and the prompts available to jailbreak systems. Mr. Vassilev stated that to protect against such attacks, proactive adoption of adaptive defense strategies is necessary. These strategies should include continuous in-house red teaming and ongoing updates to guardrails, with the goal of discouraging adversaries by raising the threshold for the effort required to exploit LLMs. To support these strategic efforts, NIST plans to update the NIST 100-2 E2025, *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations* [10], to include agentic AI and the landscape of its attacks and potential mitigations to support the AI security community. Mr. Vassilev also noted the connection to the COSAiS project and this effort.

2.2.4. Dioptra (Mr. Harold Booth, NIST)

Mr. Booth provided a description of Dioptra [11], noting that it serves as a test platform that allows evaluations of different AI systems in a reproducible manner and tracks their inputs and outputs, understanding the different models and scenarios as they relate to those inputs and outputs. Mr. Booth also gave an overview of the different use cases for Dioptra, enabling individuals, based on their areas of expertise, to obtain specific insights that can inform their work. Use cases mentioned by Mr. Booth included using Dioptra to conduct testing that supports the characterization and evaluation of an AI system prior to deployment as well as using Dioptra by researchers to understand how attacks and defenses might perform relative to a new model or training technique. Mr. Booth noted that Dioptra is available on GitHub as an open-source project, is in the process of being updated to version 1.1 (both the test platform and its associated documentation), and includes new features such as built-in support for artifact handling.

2.2.5. Secure Software Development Framework (SSDF) AI Profile (Mr. Harold Booth, NIST)

Mr. Booth also gave an overview of the SSDF AI Profile (NIST SP 800-218A, *Secure Software Development Practices for Generative AI and Dual-Use Foundation Models: An SSDF Community Profile*) [12]. Mr. Booth explained that the goal is to make the Profile broadly applicable to AI systems and to examine what is unique or new in AI system development. For example, system developers must also consider the larger scale of data used by AI systems, as well as the reliability and protection mechanisms against attacks, and data management activities such as cleaning and pre-processing. When developing AI systems, data and code must be tracked to understand what happens with data from the moment it is collected and how it is fed into training systems. Mr. Booth also noted that organizations need to be aware that reusing third-party AI models can introduce security concerns, such as providing threat vectors for adversaries to exploit NIST is assessing the timeline for updating the SSDF AI Profile published in July 2024. The SSDF [13], which serves as the basis for the SSDF AI Profile, was recently updated to version 1.2 (initial public draft [IPD]) [14], consistent with Executive Order 14306 [15]. The public comment period for the SSDF was from December 2025 to January 2026.

2.2.6. Privacy Enhancing Technologies (PETs) Testbed (Dr. Gary Howarth, NIST)

The PETs program conducts investigations into privacy-enhancing technologies, which are technical tools or controls intended to reduce privacy risks arising from the collection, sharing, and processing of potentially sensitive data. Mr. Howarth described the PETs program [16] as producing guidelines and frameworks to evaluate and manage privacy risks. Mr. Howarth shared that NIST is developing the PETs testbed at the NCCoE, which will include benchmark data and model test deployments to demonstrate how PETs are deployed and investigate trade-offs, with the goal of creating a comprehensive set of tools to guide organizations when selecting and deploying PETs and managing privacy risks. Mr. Howarth continued by stating that PETs programs, to include the NIST Collaborative Research Cycle (CRC) [17], are complementary to AI risk management efforts. The CRC program is the largest collection of synthetic data for benchmarking exercises related to privacy. Under the CRC program, NIST is analyzing the consequences of using and training AI/ML with synthetic data and is exploring the impacts on AI/ML performance.

Mr. Howarth spoke about a second program that is demonstrating how to apply privacy-preserving federated learning (PPFL). Federated learning connects ML among unconnected devices, enabling increases to both compute power and the amount of data available to train ML algorithms. PPFL introduces important privacy considerations, which NIST's privacy-preserving federated learning platform is being used to investigate. NIST is asking questions about how different privatization mechanisms affect ML performance and which security and privacy controls are needed. NIST plans to publish model code as well as a complementary privacy and cybersecurity threat modeling example that shows the types of threats that PPFL is attempting to manage and applicable privacy and cybersecurity controls. Mr. Howarth concluded by noting that reducing privacy threats requires a strong cybersecurity foundation and emphasized that using synthetic data alone does not adequately protect privacy unless the original data is also secured.

2.2.7. Agent Identities (Mr. Ryan Galluzzo, NIST)

Mr. Galluzzo opened by acknowledging that AI agents are starting to change the context and conversation around traditional cybersecurity controls, including regarding issues on agent access and identifying agents operating in organizations, how to authenticate an agent and verify it is what it claims to be, and how to authorize and manage access on an ongoing basis for agents to engage with data or take particular action, such as accessing data or resources, and necessary restrictions on those actions to address concerns that may result from data aggregation.

NIST is examining the current state of the identity management space and the tools applied today to individual users and to certain software—or machine-to-machine-based systems—and how that context is changing with systems that can ingest external or internal data sources, analyze them, and potentially take actions [18][19]. Mr. Galluzzo explained that this analysis helps NIST understand the areas of change and impact to existing standards and any solutions, including determining if existing standards can be applied or updated, if new guidance (e.g.,

Profiles) may be needed to address gaps and new context, how to build in a reference architecture that will inform configuration, or recommended best practices.

The Agent Identities work focuses on understanding and creating authorization flows and delegations pertaining to created agents, specifically determining the individual(s) who will be accountable for actions the AI system may take and how to bind the identity of that individual to the actions the agent is taking so organizations can audit, log, and understand the chains of trust related to agent actions. This work also considers human-in-the-loop scenarios, such as specifying who is notified and who decides whether to grant access when an agent tries to access data it is not authorized to use. Mr. Galluzzo concluded by noting that it is important to understand the balance between the benefits of agents operating within an enterprise and the need to protect the information an agent might access to preserve privacy, security, and confidentiality, to help inform where adaptations may be needed.

NIST is planning to propose an NCCoE collaboration project through an initial concept paper based on use cases, including how to deploy agents at scale, with the goal of launching an NCCoE demonstration project that includes industry involvement.

2.2.8. NCCoE Chatbot (Mr. Daniel Lee, NIST NCCoE/MITRE)

Mr. Lee opened with an overview of the NCCoE Chatbot [20], stating that the Chatbot was designed to increase the efficiency and ease of conducting searches on the wealth of information the NCCoE has available on its website. The NCCoE Chatbot uses a customized retrieval-augmented generation (RAG)-based LLM to assist with searching large amounts of data. The customized RAG-LLM includes a metadata layer designed to "filter down" the search index, focusing on specific keywords or information before the Chatbot conducts a semantic similarity vector comparison to produce its search results. Mr. Lee noted that including the metadata layer in the Chatbot is intended to address the lag in responses that occurs as the search index grows.

Mr. Lee shared that NIST Interagency Report (IR) 8579 (IPD), "Developing the NCCoE Chatbot: Technical and Security Learnings from the Initial Implementation" [21], provides information on the NCCoE Chatbot as well as security considerations when developing prompts to support trust. NIST IR 8579 also includes details on the reasoning behind the types of metadata information used in metadata indexing and filtering.

2.2.9. Collaboration Across Programs

The panelists also discussed the importance that NIST places on collaboration across programs. For example, there are collaboration touchpoints between the NCCoE agent identity work and the CAISI agent work. NIST also develops guidelines, such as the Cyber AI Profile, that provide a high-level understanding of cybersecurity outcomes and objectives and specific recommendations on how to achieve them, using, for example, COSAiS for specific AI types.

2.3. Presentation: NIST Strategy for American Leadership in the 21st Century: Emerging Technology Accelerators (including the "AI Accelerator") (Dr. Craig Schlenoff, NIST)

Dr. Schlenoff opened with an overview of the Administration's science and technology priorities and how those priorities are shaping emerging research on technologies such as AI and quantum information technology (QIT). NIST, guided by the Administration's priorities, developed the *NIST Strategy for American Technology Leadership in the 21st Century*, which includes AI considerations across its four areas. Dr. Schlenoff focused the discussion on specific areas related to American AI innovation and the acceleration of manufacturing productivity; protecting and securing U.S. critical infrastructure; measuring AI system performance, reliability, and security; and rapidly evaluating AI system capabilities to promote AI innovation, which aligns with work under CAISI.

Moving to the discussion on the Emerging Technology Accelerators, Dr. Schlenoff mentioned the December 2025 establishment of the "AI Accelerator," [22] which will include the "AI Economic Security Center for U.S. Manufacturing Productivity" and the "AI Economic Security Center to Secure U.S. Critical Infrastructure from Cyberthreats." Dr. Schlenoff provided overviews of each of these areas in the AI Accelerator. Dr. Schlenoff acknowledged that these areas remain broad and NIST remains consistently active in collaboration and outreach with the community to understand key challenges where impact can be made, highlighting that collaborations with industry, academia, and others are leveraged to inform conversations that identify core challenges that can be solved and work closely with those stakeholders to do so. Dr. Schlenoff noted that there are opportunities to involve academia, whether during pre-standardization development (e.g., participation in research that will inform the standards) or during standard development (e.g., through standards bodies). NIST is working with academic partners but is open to exploring increased engagement and involvement.

Dr. Schlenoff explained that NIST's strategy is to create "unprecedented AI 'gold standards'". The AI Accelerator Hub builds on work by the NIST Research Labs, combined with "spokes" that leverage public-private partnerships and collaboration. The AI Accelerator is focused on driving impact and understanding the disruptive, transformative technologies that can be implemented to deliver fundamental, significant changes in a short period of time (e.g., 3 years). Dr. Schlenoff noted that the Cyber AI Profile work and outcomes from the associated workshops will provide foundational input to the AI Accelerator Hub.

Dr. Schlenoff concluded his presentation by inviting the community to participate in these collaborations.

2.4. Cyber AI Profile Preliminary Draft

The following sections contain information shared during the presentation of the Cyber AI Profile and summaries of each of the three Cyber AI Profile afternoon breakout sessions.

2.4.1. Presentation: Overview of the Cyber AI Profile Preliminary Draft (Ms. Julie Snyder, NCCoE/MITRE)

Ms. Snyder opened with an overview of the NIST CSF 2.0 and CSF Profiles, noting that Community Profiles, such as the Cyber AI Profile, can help organizations understand the community's cybersecurity direction and collective priorities. Organizations can use Community Profiles to inform their own strategic direction and roadmap and to address gaps between their own cybersecurity activities and those prioritized in a Community Profile.

Ms. Snyder then spoke specifically about the Cyber AI Profile and how collaborative discussions informed its development. The Profile is designed to be broadly applicable, so organizations can implement it in ways that support many different AI adoption and use scenarios and a variety of roles, including developing AI systems, using AI technologies, capitalizing on cybersecurity capabilities AI can provide, and seeking to better understand and defend against adversarial use of AI to conduct attacks. This approach focuses the Profile on cybersecurity and AI-specific opportunities and concerns when designing or deploying an AI system, rather than on providing engineering guidance.

Ms. Snyder reviewed the three Focus Areas in the Profile—Securing AI System Components (Secure), Conducting AI-Enabled Cyber Defense (Defend), and Thwarting AI-Enabled Cyber Attacks (Thwart)—and shared a brief description of each, noting relationships and distinctions among the three Focus Areas. The Focus Areas reflect feedback that the NCCoE heard from the community. Ms. Snyder noted that all three Focus Areas include sample considerations, with the Defend Focus Area's considerations specifically aimed at sample opportunities that highlight how AI can enhance cybersecurity programs and activities.

The Profile's tables contain general and Focus Area-specific considerations for the CSF Subcategories, proposed priorities, and Informative References (i.e., related publicly available standards and guidelines as well as related research). Subcategory priorities are intended to reflect the relative importance of the overall cybersecurity strategy, combined with the areas most impacted within an organization. CSF Profiles, such as the Cyber AI Profile, include references to additional relevant standards and guidance that organizations and communities can incorporate into their efforts to achieve specific cybersecurity outcomes.

Ms. Snyder concluded by sharing that the afternoon breakout sessions would be exploring the general questions in the "Note to Reviewers" as well as Focus Area-specific questions.

2.4.2. Afternoon Breakout Session Track A – Secure Focus Area

This breakout session focused on the Secure Focus Area, which supplements existing cybersecurity and risk management best practices to address novel, expanded, and altered attack surfaces associated with the integration of AI systems into an organization, its ecosystems, and its infrastructure. This Focus Area covers the AI systems themselves, their supply chains, including data and machine learning infrastructure, and the other systems and data that the AI system relies on.

Track A Themes

The Secure Focus Area breakouts spent significant time on governance-related topics, particularly accountability, ownership, and risk management. One of the key challenges and concerns stressed by participants across all Track A sessions was that AI governance moves more slowly than AI innovation. Additionally, practitioners are struggling to identify and explain AI and AI risk, including finding where AI lives in their environments. Accountability, in particular, garnered attention with subtopics including AI visibility within organizations, understanding responsibilities for specific roles (e.g., developer, deployer, user), the increasing complexity of human oversight as agentic AI capabilities expand, and concerns over the relationship between accountability and liability. The lack of clarity in these areas makes it difficult to identify and manage AI risks.

Participants also focused on tactical needs, such as AI impact assessments that integrate with other risk management processes (e.g., application and data risk assessments) and that are not overly burdensome. The CSF and Cyber AI Profile are generally considered useful starting points for shaping AI risk assessments. Participants also discussed procurement challenges, including AI Bill of Materials (AIBOM), the need for guidance when purchasing AI-related technologies and components (e.g., libraries, models, data), and how to incorporate AI considerations into vendor risk assessments.

Track A Input Offered by Participants

Participants engaged in a facilitated discussion and shared insights specific to the following topic areas for the Secure Focus Area:

- Accountability for AI-driven decisions is a critical component, including how to separate or designate accountability, and the request for guidance on how to define and assign accountability.
    - Participants indicated that responsibility for AI-driven decisions typically falls on those lines of business deploying the AI but noted that responsibility may shift among different lines of business.
    - Participants discussed whether and how to separate or designate accountability, raising topics such as:
        - Clear lines of responsibility that need to exist for “developer vs. deployer vs. user” of AI systems, to include who is responsible if the systems are deployed “out of the box” or are modified.
        - Concerns about the provenance of agents developed by third parties, as well as the owner and arbiter of data that is relied upon by AI systems. One of the participants offered an example of an energy company using a chatbot to respond to inquiries with safety implications, and if, in doing so, a hallucination leads to harm, the liability will fall to the energy company even if the energy company did not develop the chatbot.
    - Participants also cautioned that the term “accountability” could be interpreted as “liability.”

- Approaches to governance are varied. While some organizations are establishing multidisciplinary governance bodies and boards with members representing a variety of expertise, including legal and marketing, other organizations are evolving governance structures that focus on roles such as a Chief AI Officer, similar to a Chief Information Officer (CIO) or Chief Privacy Officer (CPO).
- Technical challenges exist related to securing enterprise AI systems.
    - Participants noted it can be difficult to determine where AI systems, including "shadow AI," are in an enterprise and what components outside the enterprise the AI might need to access. Understanding these AI deployments is a precursor to securing the AI system.
    - Regarding vendors and vendor-provided technologies, participants stressed difficulties in managing AI risk in vendor technologies, with some vendors being more helpful in mitigating risk than others. Participants also mentioned it is not possible to clearly understand which vendors are securing or developing AI "properly," and this knowledge gap further complicates the procurement process.
    - Another area participants stressed was how to fail gracefully and revert to normal operations when AI systems do not operate as intended (i.e., cyber resiliency for operating AI, rollback of AI systems).
- Specific CSF Category topics on encryption of data in transit (PR.DS-02), authentication needs (PR.AA), and continuous monitoring (DE.CM), including which information should be logged from AI systems, also arose as areas with considerations to address.

2.4.3. Afternoon Breakout Session Track B – Defend Focus Area

This session explored the Defend Focus Area, which concentrates on identifying opportunities to use AI to support cybersecurity programs and activities and understanding the challenges that come from doing so. Participants identified opportunities to enhance cyber defense capabilities using AI, including mission assurance, predictive and proactive applications, investigation and analysis, and response and remediation. The purpose of this track was to review the Defend Focus Area content presented in the Cyber AI Profile preliminary draft and discuss feedback for creating the final version of this section.

Track B Themes

Recurring comments from the Defend Focus Area included challenges related to the testing and evaluation of AI systems for defense, leveraging AI to combat insider threats, using AI in penetration testing and continuous monitoring, monitoring AI agents (e.g., monitoring AI agents using other AI agents), leveraging AI in the CSF's Respond and Recover Functions, and ensuring that the Profile can be used by organizations of varying size and resource levels.

Track B Input Offered by Participants

Participants engaged in a facilitated discussion and shared insights specific to the following topics for the Defend Focus Area:

- Opportunities exist to emphasize how to use AI to create and use AI Security Operations Centers (SOCs).
- Participants discussed accountability in the context of "Defend", and who is responsible and accountable if an AI-enabled cyber defense tool fails in a critical moment.
- Participant discussions on AI use for Defend aligned with CSF cybersecurity outcomes and activities, to include:
    - Protect (e.g., AI-driven detections, continuous identity validation).
    - Detect (e.g., continuous monitoring; parsing, correlating, and analyzing data; quickly querying malware feeds and identifying matches for malware signatures; performing data integrity checks; detecting anomalies; issuing alerts).
    - Respond (e.g., remediating vulnerabilities; decreasing mean time in responding to AI attacks; enhancing communications and the cadence of response activities following an incident).
    - Recover (e.g., generating After Action Reports [AARs] to create a succinct narrative of an event, including timestamps; analysis of what went wrong; recommendations regarding prioritized mitigation strategies).
- The feasibility of a mapping of Defend to existing frameworks, knowledge bases, and guidance, such as MITRE ATLAS™[3], was discussed. Participants also suggested that developing guidance and/or methods for prioritization based on maturity and risk could be valuable (e.g., with greater nuance than currently provided, perhaps based on use cases, risk factors, or methods).
- There was emphasis on the need for clear language on the relationship between adopting AI for defense and the risks that may be introduced (e.g., unauthorized access to an AI cybersecurity defense tool).
- Guidance for how AI can be used to identify effective cybersecurity solutions that use AI and that could be helpful. Participants expressed that having a better understanding of existing and emerging AI tools (e.g., purchasing software that performs AI-enhanced/enabled functions) and offerings can strengthen cybersecurity operations and risk management practices.
    - Participants inquired whether NIST would propose a recommended list of tools or whether the Profile could address existing tools for various cyber capabilities

[3] ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems) is a globally accessible, living knowledge base of adversary tactics and techniques against AI-enabled systems based on real-world attack observations and realistic demonstrations from AI red teams and security groups, available at: https://atlas.mitre.org/

(e.g., automation, predictive AI, genAI) across CSF Functions, Categories, and Subcategories.

- Organizations can better defend by having an arbiter of the data with visibility into what their deployed AI systems are doing and an understanding of each AI system's training (insight into models), functions, capabilities, and actions.
- Agentic AI is increasingly used for new security-related applications. Participants discussed using agentic AI to augment junior staff, using agents to monitor other agents, and leveraging AI's predictive capabilities in SOCs to respond to alerts (e.g., an "Agentic SOC" that can take action based on the incidents it identifies through activities like stopping services, revoking credentials, or isolating devices), but find acquisitions, testing, explainability, and evaluation difficult. Participants noted that the selection process for an agent needs to consider the data the agent has access to and can operate on.

2.4.4. Afternoon Breakout Session Track C – Thwart Focus Area

This session examined the Thwart Focus Area, which addresses how attacks from AI-enhanced adversarial systems could impact the entire cybersecurity landscape, and what organizations can do to bolster their systems against these emerging threats. The purpose of this track is to review the Thwart Focus Area content presented in the Cyber AI Profile preliminary draft and discuss feedback for creating the final version of this section.

Track C Themes:

The Thwart discussions focused on sharpening the description of the Focus Area and questions regarding how to obtain and share threat information (e.g., resources for known AI-enabled attacks, sharing across government agencies). It was noted that defenders have to "get it right" every time, while adversaries only have to "get it right" once, and adversaries are looking for the holes between the "checkboxes."

Track C Input Offered by Participants:

Participants engaged in a facilitated discussion and shared insights specific to the following topic areas for the Thwart Focus Area:

- Clarifying and explaining the range of adversarial uses of AI and how the outcomes in Thwart position organizations to be more resilient against them.
- New tools are needed to raise awareness of adversarial uses of AI as envisioned in Thwart, to include:
    - Explicitly addressing the supply chain;
    - How to create collaborative awareness of AI threats using a model similar to information sharing and analysis centers (ISACs); and

    - Explore how to determine responsibility in AI and AI agent actions in chains of custody that reflect development, training, implementation, and use by different parties.

- There will be increased attack velocity from adversaries, necessitating higher response/defense velocity from organizations and potentially requiring new defenses (both AI and non-AI) to disrupt attacks.
    - Risk management becomes increasingly important, including determining and communicating what risks are acceptable and at what levels. Dynamic tools or mechanisms to maintain communications about how adversaries are using AI to conduct attacks and how those attacks are being identified and mitigated may help, such as a repository of prompt injection attacks.
    - Subcategories in the CSF Respond Function can help address this cybersecurity need, including communications on risk responses and risk management needs.

2.5. COSAiS

The following sections contain information shared during the COSAiS presentation and a summary of the COSAiS afternoon breakout session.

2.5.1. Presentation: Control Overlays for Securing AI Systems (COSAiS) (Ms. Victoria Pillitteri, NIST)

Ms. Pillitteri opened by sharing the drivers for the COSAiS project, describing the feedback from the community to develop technical implementation guidance for securing AI systems leveraging existing standards and guidelines to the maximum extent, complementing the Cyber AI Profile. She provided an overview of the NIST SP 800-53 controls catalog [23], emphasizing that any organization can leverage the broad collection of cybersecurity and privacy outcomes to determine which controls work for them to manage risk and should not be implemented in totality. Ms. Pillitteri stated that because the controls needed to manage cybersecurity and privacy risks to AI systems are largely similar to what would be needed for any type of software, the community can leverage the existing NIST SP 800-53 control catalog as a starting point. She acknowledged that since the controls are not specific to any given technology or AI systems, the COSAiS work seeks to identify subsets of controls especially relevant for AI systems to better help developers and implementers.

Ms. Pillitteri explained that control overlays can be either a fully specified set of controls that is inclusive of a baseline or a small subset of controls tailored to address a specific community or organization need, such as an environment of operation or type of system. Within these control overlays, the overlay can provide further details, such as defining specific parameter values and setting potential frequency values.

Ms. Pillitteri shared that COSAiS provides an opportunity to work with the community to complement the Profile and provide implementation-specific guidance for different types of AI systems (e.g., predictive AI, genAI, single- and multi-agent AI, and controls for security

developers). Within the COSAiS project, these existing controls will be tailored to manage risk specific to types and uses of AI systems. By developing multiple overlays, each will focus on managing the different cybersecurity and privacy risks associated with different types of AI systems, and each overlay will have a target audience and focus. The goal is for the overlays to work alone or in combination, allowing organizations the flexibility to determine which overlay(s) to apply based on their unique needs. The COSAiS can also leverage the assessment guidance that exists for the control catalog and other risk management guidance.

Ms. Pillitteri reviewed the proposed methodology in the "NIST SP 800-53 Control Overlays for Securing AI Systems Concept Paper" [24] issued during the summer of 2025 and built upon by the recently released "NIST SP 800-53 Control Overlay for Securing AI Systems: Using and Fine Tuning Predictive AI (Annotated Outline for Cyber AI Profile Workshop #2)" [25]. She discussed the contents of the annotated outline for the initial "Using and Fine-Tuning Predictive AI" overlay, emphasizing that the final decisions regarding the overlay are pending stakeholder feedback and insights.

Ms. Pillitteri emphasized that understanding the gaps between controls that organizations may currently implement, and the additional controls needed to secure AI systems, is critical to creating control overlays that are "lightweight but impactful." To develop resources to meet stakeholder needs, NIST consistently engages in outreach and collaboration with stakeholders, such as through workshops and Slack community discussions.[4] One area of feedback NIST is seeking is whether the overlays should be based on the NIST SP 800-53B [26] Moderate baseline. The COSAiS is based on several assumptions, such as an existing cybersecurity risk management program.

Ms. Pillitteri concluded with an overview of what participants can expect in the COSAiS breakout sessions and where NIST is seeking feedback, to include the foundational assumptions, potential alternatives, and the overlay structure and control selection, noting that NIST intends to offer several opportunities after the workshop to provide feedback, as that feedback is instrumental to the success of the COSAiS.

2.5.2. Afternoon Breakout Session Track D – COSAiS

This session focused on NIST's COSAiS project.

Track D Themes:

Participant discussions during the COSAiS sessions emphasized the importance of NIST making the case for why the COSAiS are needed, using clear, direct language to support the overlays. Discussions also highlighted the need to focus the overlays' guidance on small businesses, which are typically more constrained or have less access to resources to support cybersecurity initiatives and activities, and noted that the controls identified should not be too strict or inhibit innovation. Participants suggested that NIST consider using the Low-impact baseline rather than the currently proposed Moderate baseline.

[4] Information for joining the NIST COSAiS Slack Collaboration page is available at: https://csrc.nist.gov/projects/cosais/slack.

Track D Input Offered by Participants:

Participants engaged in facilitated discussion and shared insights specific to the following topic areas for COSAiS:

- Clarity on foundational concepts related to the structure of overlays, specifically why overlays include a baseline.
    - Participants expressed the value of establishing the context of a baseline in the overlays and whether the baseline serves as a starting point or is intended to represent the final control set.
    - Participants shared that it would be helpful to implementers and users of the COSAiS for NIST to consider shifting from a baseline to a modular approach for greater tailoring and adaptation by different organizations and industries with varying backgrounds, needs, environments, and AI use.
    - A request was made for additional information on the decision to develop COSAiS using existing controls rather than have NIST develop new controls specific to AI.
- To support developers and implementers, the COSAiS needs to be applicable to the current environment and also allow for innovation and the incorporation of emerging technology.
- Supply chain should be addressed in the overlays [4]. This may include guidance on requesting test data and test data artifacts from suppliers, vendors, and third-party companies that develop AI models used by other organizations. Organizations, especially smaller ones, must have mechanisms in place to establish trustworthiness and be able to enforce safety and accuracy in the AI models they obtain from others.
- There are opportunities to demonstrate the relationship between the overlays and other frameworks or guidance, including the AI RMF or European AI guidelines and standards.

3. Key Takeaways from Workshop #2

The Cyber AI Profile Workshop offered participants an opportunity to share with NIST feedback and insights throughout the day on the draft Cyber AI Profile. The themes summarized below reflect perspectives raised by workshop participants and are intended to inform NIST's continued development of the Cyber AI Profile and other NIST programs. Overall, participants expressed support for the development of the Cyber AI Profile and appreciated that it is filling gaps in guidance. They emphasized the need for both enterprise risk management [27][28][29] and implementation-level resources, especially to benefit smaller organizations or organizations overwhelmed with addressing strategic adoption, integration, and use of AI while continuing to manage an already full load of ongoing tactical and operational cybersecurity risk management activities. The following takeaways are the ideas, observations, and suggestions that NIST heard from workshop participants and received significant support from them. This

workshop was not a forum for developing consensus; rather, the takeaways represent recurrent themes that emerged during discussions and not formal positions taken by participants. This document cannot capture every thought, opinion, and suggestion provided during the workshop. These takeaways do not represent NIST recommendations or guidelines; rather, they provide important feedback to the program and serve as a basis for future conversations with the community and NIST's continued development of the Cyber AI Profile as a practical resource for organizations navigating AI-related opportunities and risks.

- **Ideas for Enhancing the Profile Content or Its Readability:**
    - General information to include: the relationship between the Cyber AI Profile and the CSF (e.g., visual depiction of the relationship versus text); content that conveys to users the "why" to support planning and implementation of capability development; the general need for the Cyber AI Profile as AI is widespread and rapidly becoming ubiquitous within organizations; and how to use in combination the Cyber AI Profile, the CSF 2.0, and the AI RMF when building a single, operational view of risk.
    - The Profile was perceived by participants as deployment-focused. Participants asked NIST to expand the content to include how to improve the cybersecurity of AI systems during development in anticipation of how the AI will be deployed and used.
    - Clarification of the definitions for the three Focus Area terms and potentially additional visuals that help distinguish the Focus Areas' scopes.
    - Formats that support sortable views, such as viewing only the Priority 1 or dropdown boxes that enable analysis of different levels of detail for the Subcategories.
    - The inclusion of additional guidance on topics such as addressing insider threats posed by both humans and AI; prioritization assistance (e.g., through use of a consequence tiering model); and AI nuances that are inherent in Priority 3 Subcategories that may not yet exist in organizations.
    - Include hyperlinks to AI resources and other guidance where possible, such as in the Informative References column, even as those references update over time.
    - Too many hurdles (real or perceived) to deploying AI often lead people to take the path of least resistance and use AI through their own accounts (e.g., personal LLMs), increasing the risk of incidents such as data leaks.
- **Guidance on Testing and Evaluation:**
    - Organizations are seeking guidance on how to answer the question, "How can we trust an AI system before giving it our data?"
    - Potential options to address testing and evaluation were presented, including performance metrics, certifications, and benchmarking.

- **Cybersecurity's Role in Trustworthiness for AI Decisions:**
    - Zero-trust and reduced trust discussions as they relate to AI decisions and actions within an organization indicated there may be a need to treat AI similarly to how human insider threats are addressed.
    - AI is often deployed in organizations without an understanding of the associated or impacted technology. Organizations must understand the technology they are deploying as well as how to securely deploy it and manage associated opportunities and risks. The notion of "shadow AI" was raised as a significant concern in the context of AI impact assessments.
    - From an integrity perspective, a lack of visibility into the data used to train AI models, or into the models themselves, is a concern. This includes a lack of visibility into how AI models are trained in products and services from suppliers and third-party vendors. Participants stated that organizations need to know what data was used in a model so they can explain AI actions and functions to end users to increase the trustworthiness of AI. This includes organizational ability to communicate clearly to leadership to establish trustworthiness related to the use of AI and AI-driven decisions (whether made by AI or users informed by AI).
- **Transparency, Integrity, and Accountability:**
    - Considerations in the Profile should highlight the need to 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.
    - AI supply chain documentation is important to secure the AI supply chain (e.g., inclusion of AIBOM, similar to a Software Bill of Materials [SBOM], to address supply chain risks associated with AI systems), and to manage risks associated with open-source models and third-party dependencies and technologies.
- **Continued Need for Human-in-the-Loop:** Human-in-the-Loop (HITL) processes and training remain critical in AI adoption to ensure effective implementation and oversight of AI systems used for cybersecurity, noting that HITL remains the current standard for AI accountability.
- **Longevity of Profile and Innovation:** The Profile should not include guidance that is too specific, as technology is moving fast, and overly specific content will prohibit the Profile from remaining usable and helpful as innovation occurs.
- **Need for Consistent AI Taxonomy:** A consistent, industry-agnostic AI taxonomy, specifically in the Profile, would enable organizations across industries to communicate clearly about AI topics.

- **Use Cases:** Use cases (e.g., operational technology [OT] cybersecurity) and illustrative examples in the Profile would be helpful.

3.1. Adjacent Considerations

Throughout the workshop, participants highlighted several important themes that extend beyond the immediate scope of the Cyber AI Profile. While these topics will not be directly addressed through the Profile, they represent valuable insights regarding community needs:

- **Tailored Support for Small and Medium Businesses (SMBs):** Participants emphasized the need to provide SMBs with tailored support, including simple acquisition questions and practical tools to address their unique challenges.
- **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 NIST Online Informative References (OLIR) [30]-based mappings and LLM integration, to make the Profile more user-friendly and applicable across diverse industries. These discussions extended beyond the level of sophistication suggested by earlier comments regarding simpler sorting capabilities to include the ability to sort COSAiS content by lifecycle stage.
- **Workforce Development and Guidance:** Participants emphasized the need for workforce training to address AI-related cybersecurity challenges, including adopting AI for cybersecurity. Challenges with the speed and velocity of AI adoption and AI attacks emphasize the importance of workforce training and awareness to recognize and defend against adversarial AI attacks. Workforce-related guidance may include decisions to align workforce roles and resources with AI capabilities and applications.
- **Procurement Process Guidance:** Practitioners need consistent guidance in the procurement process for new technologies (e.g., libraries, models, data) to support risk management decision-making related to AI. Addressing gaps in guidance enables more informed risk management decision-making by clarifying how to determine and communicate risk associated with new AI systems, identify the assurances that need to be in place, and use existing application and data risk assessment sources to adequately answer the risk questions raised by these new technologies.
- **NIST AI Project Awareness:** Creating linkages to understand the integration across multiple NIST frameworks and the integration of AI into those frameworks. Other linkages between the Cyber AI Profile and NIST's other impactful AI portfolio projects, such as those highlighted at the workshop, would help broaden awareness of and connections among these efforts.

4. Next Steps

NIST continues to follow the Cyber AI Profile Roadmap which can be found on the NCCoE Cyber AI Profile project page. This roadmap is a living document that reflects the overall steps NIST will complete to create the Cyber AI Profile. Each step of the process toward the Final Cyber AI

Profile builds on prior actions and information, including workshops and public comment periods. The Cyber AI Profile preliminary draft and Workshop #2 were informed by research conducted by the team and stakeholder comments received during previous activities in the roadmap. NIST will leverage the information collected from Workshop #2 as well as the comments submitted on the Cyber AI Profile preliminary draft, and future research and engagement activities to guide the next version of the Cyber AI Profile (the IPD) for public comment in summer 2026.

For more information on the Cyber AI Profile, Cyber AI Profile roadmap phases, or to join the Cyber AI Profile Community of Interest (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, Cybersecurity Framework Profile for Artificial Intelligence (Cyber AI Profile), NIST IR 8596 (IPRD), Gaithersburg, MD, 2025. https://nvlpubs.nist.gov/nistpubs/ir/2025/NIST.IR.8596.iprd.pdf

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

[3] National Institute of Standards and Technology, Cybersecurity and AI Workshop Concept Paper, Gaithersburg, MD, 2025. https://www.nccoe.nist.gov/sites/default/files/2025-02/cyber-ai-concept-paper.pdf

[4] National Institute of Standards and Technology, NIST SP 800-53 Control Overlays for Securing AI Systems (COSAiS), Gaithersburg, MD, 2025. https://csrc.nist.gov/projects/cosais

[5] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, Gaithersburg, MD, 2023. https://doi.org/10.6028/NIST.AI.100-1

[6] National Institute of Standards and Technology, Strategic Priorities, Gaithersburg, MD, 2024. https://www.nist.gov/director/strategic-priorities

[7] National Institute of Standards and Technology, Center for AI Standards and Innovation (CAISI), Gaithersburg, MD, 2024. https://www.nist.gov/caisi

[8] National Institute of Standards and Technology, Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations, NIST AI 800-2 (IPD), Gaithersburg, MD, 2025. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.800-2.ipd.pdf

[9] National Institute of Standards and Technology, Request for Information Regarding Security Considerations for Artificial Intelligence Agents, Gaithersburg, MD, 2026. https://www.federalregister.gov/documents/2026/01/08/2026-00206/request-for-information-regarding-security-considerations-for-artificial-intelligence-agents

[10] National Institute of Standards and Technology, Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations, NIST AI 100-2e2025, Gaithersburg, MD, 2025. https://csrc.nist.gov/pubs/ai/100/2/e2025/final

[11] National Institute of Standards and Technology, Dioptra: A Software Test Platform for Machine Learning, Gaithersburg, MD, 2024. https://pages.nist.gov/dioptra/

[12] National Institute of Standards and Technology, Secure Software Development Practices for Generative AI and Dual-Use Foundation Models: An SSDF Community Profile, NIST SP 800-218A, Gaithersburg, MD, 2024. https://csrc.nist.gov/pubs/sp/800/218/a/final

[13] National Institute of Standards and Technology, Secure Software Development Framework (SSDF) Version 1.1: Recommendations for Mitigating the Risk of Software Vulnerabilities, NIST SP 800-218, Gaithersburg, MD, 2022. https://csrc.nist.gov/pubs/sp/800/218/final

[14] National Institute of Standards and Technology, Secure Software Development Framework (SSDF) Version 1.2: Recommendations for Mitigating the Risk of Software Vulnerabilities, NIST SP 800-218 Rev. 1 (IPD), Gaithersburg, MD, 202

[15] The White House, Executive Order 14306: Sustaining Select Efforts To Strengthen the Nation's Cybersecurity and Amending Executive Order 13694 and Executive Order 14144, Washington, DC, Jun. 6, 2025. https://www.govinfo.gov/app/details/DCPD-202500669

[16] National Institute of Standards and Technology, Privacy-Enhancing Technologies (PETs) Testbed, Gaithersburg, MD, 2024. https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/pets-testbed

[17] National Institute of Standards and Technology, Collaborative Research Cycle, Gaithersburg, MD, 2024. https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space/collaborative-research-cycle

[18] National Institute of Standards and Technology, Software and AI Agent Identity and Authorization, Gaithersburg, MD, 2025. https://www.nccoe.nist.gov/projects/software-and-ai-agent-identity-and-authorization

[19] National Institute of Standards and Technology, Digital Identity Guidelines, NIST SP 800-63, Gaithersburg, MD, 2017. https://www.nist.gov/identity-access-management/projects/nist-special-publication-800-63-digital-identity-guidelines

[20] National Institute of Standards and Technology, NCCoE Chatbot Project, Gaithersburg, MD, 2025. https://www.nccoe.nist.gov/projects/nccoe-chatbot

[21] National Institute of Standards and Technology, Developing the NCCoE Chatbot: Technical and Security Learnings from the Initial Implementation, NIST IR 8579 (IPD), Gaithersburg, MD, 2025. https://csrc.nist.gov/pubs/ir/8579/ipd

[22] National Institute of Standards and Technology, NIST Launches Centers for AI in Manufacturing and Critical Infrastructure, Gaithersburg, MD, 2025. https://www.nist.gov/news-events/news/2025/12/nist-launches-centers-ai-manufacturing-and-critical-infrastructure

[23] National Institute of Standards and Technology, Security and Privacy Controls for Information Systems and Organizations, NIST SP 800-53 Rev. 5, Gaithersburg, MD, 2020. https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final

[24] National Institute of Standards and Technology, NIST SP 800-53 Control Overlays for Securing AI Systems Concept Paper, Gaithersburg, MD. https://csrc.nist.gov/csrc/media/Projects/cosais/documents/NIST-Overlays-SecuringAI-concept-paper.pdf

[25] National Institute of Standards and Technology, NIST SP 800-53 Control Overlay for Securing AI Systems: Using and Fine-Tuning Predictive AI (Annotated Outline for Cyber AI Profile Workshop #2), Gaithersburg, MD, 2026. https://csrc.nist.gov/csrc/media/Projects/cosais/documents/COSAiS-Predictive-AI-annotated-outline-Jan2026.pdf

[26] National Institute of Standards and Technology, Control Baselines for Information Systems and Organizations, NIST SP 800-53B, Gaithersburg, MD, 2025. https://csrc.nist.gov/pubs/sp/800/53/b/upd1/final

[27] Quinn SD, Chua J, Ivy N, Gardner RK, Kent K, Smith MC, Witte GA (2025) Integrating Cybersecurity and Enterprise Risk Management (ERM). (National Institute of Standards and Technology, Gaithersburg, MD), NIST Interagency Report (IR) NIST IR 8286r1. https://doi.org/10.6028/NIST.IR.8286r1

[28] Quinn S, Ivy N, Chua J, Barrett M, Feldman L, Topper D, Witte G, Gardner RK, Scarfone K (2023) Enterprise Impact of Information and Communications Technology Risk: Governing and Managing ICT Risk Programs Within an Enterprise Risk Portfolio. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 800-221. https://doi.org/10.6028/NIST.SP.800-221

[29] Quinn S, Ivy N, Chua J, Scarfone K, Barrett M, Feldman L, Topper D, Witte G, Gardner RK (2023) Information and Communications Technology (ICT) Risk Outcomes: Integrating ICT Risk Management Programs with the Enterprise Risk Portfolio. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 800-221A. https://doi.org/10.6028/NIST.SP.800-221A

[30] National Institute of Standards and Technology, National Online Informative References Program (OLIR), Gaithersburg, MD. https://csrc.nist.gov/projects/olir

Appendix A. List of Symbols, Abbreviations, and Acronyms

AAR
After Action Report

AI
Artificial Intelligence

AIBOM
AI Bill of Materials

AI RMF
Artificial Intelligence Risk Management Framework (NIST)

ATLAS™
Adversarial Threat Landscape for Artificial-Intelligence Systems

CAISI
Center for AI Standards and Innovation (NIST)

CIO
Chief Information Officer

COI
Community of Interest

COSAiS
Control Overlays for Securing AI Systems (NIST)

CSF
Cybersecurity Framework (NIST)

CSWP
Cybersecurity White Paper (NIST)

genAI
Generative AI

IoT
Internet of Things

IPRD
Interim Preliminary Draft

HITL
Human-in-the-Loop

IR
Interagency Report

ISAC
Information Sharing and Analysis Center

IT
Information Technology

ITL

Information Technology Laboratory

LLM
Large Language Model

ML
Machine Learning

NCCoE
National Cybersecurity Center of Excellence

NIST
National Institute of Standards and Technology

OLIR
Online Informative References (NIST)

OT
Operational Technology

PETs
Privacy Enhancing Technologies

PPFL
Privacy-Preserving Federated Learning

RAG
Retrieval-Augmented Generation

SBOM
Software Bill of Materials

SMB
Small and Medium Businesses

SOC
Security Operations Center

SP
Special Publication (NIST)

SSDF
Secure Software Development Framework (NIST)

U.S.
United States