

Discussion Essay: Topics for Community Discussion at the NIST Cyber AI Profile Spring 2026 Community of Interest (COI) Working Session #2 – Extending the Technical Content

Introduction

The NIST [Cyber AI Profile](#) is intended to help organizations strategically adopt AI while addressing and prioritizing cybersecurity risks stemming from its advancements. In December 2025, NIST NCCoE released the [Preliminary Draft](#) of the NIST Cyber AI Profile (NIST IR 8596). In January 2026, the NIST NCCoE Cyber AI Profile team hosted a [workshop](#) to obtain feedback on the Preliminary Draft and to identify cybersecurity priorities as AI adoption continues to grow. Feedback from the January 2026 workshop and ongoing analysis of the 1,400+ comments received on the draft confirmed support for the development of the Cyber AI Profile and validated that organizations need practical guidance that builds on existing cybersecurity practices while addressing the cybersecurity risks related to AI development and use. Participants emphasized the need for both enterprise risk management and implementation level resources, particularly for organizations managing AI adoption, integration, and use alongside existing cybersecurity operations. [Key themes](#) heard from participants during the workshop have been instrumental in both creating the next draft of the Profile and identifying the community's areas of priority as AI continues to evolve.

As a result of this analysis, the team identified key topics for discussion with the COI. The NIST NCCoE is holding a [series of virtual COI working sessions](#) to carry out these discussions. This read-ahead document focuses on two topics that reflect feedback received on the Profile and presents proposed ways forward that will be discussed with the community during COI [Session 2: Extending the Technical Content](#), which will be held on May 5, 2026, from 1:00–4:00pm EDT. Input from the COI will help validate our proposed direction and identify any implementation considerations that should inform the next draft of the Profile.

About the COI Working Sessions

The Spring 2026 COI working session series is designed to support targeted, facilitated discussions to further refine the Cyber AI Profile. The working sessions will cover community feedback on adapting cybersecurity practices to AI, strengthening the Cyber AI Profile in key technical areas, and exploring revised Profile delivery formats to enhance usability for different roles in the AI ecosystem. The working sessions are designed to:

- Collect feedback from the COI regarding the proposed approaches in the discussion essay for each session
- Gather insights regarding current practices and challenges
- Identify key characteristics for effective approaches to addressing the discussion topics

During the working sessions, which are held on Zoom, participants will have the opportunity to share feedback by actively engaging in discussion and using Slido polls. The Preliminary Draft and this discussion essay should be reviewed prior to the working sessions as they provide critical context, an overview of the topics, questions for discussion, and proposed paths forward.

The May 5 working session will focus on the following:

- Incorporating and supporting additional Agentic AI considerations and use cases
- Addressing the application of Zero Trust (ZT) principles in the context of AI systems

Overview of Discussion Topic and Potential Options

The NCCoE team received a significant number of comments, along with substantial suggested changes, related to the topics of Agentic AI and Zero Trust. To better understand the community's feedback and thoughtfully plan potential revisions, the NCCoE team will continue focused discussions on both important topics. These discussions will help clarify key concerns, identify areas of consensus and divergence, and ensure that any updates are practical and aligned with real-world implementation needs. More detailed information on the context, feedback, and specific comments for each topic is provided in the subsections below.

Agentic AI

Context

AI agents and agent-based applications are being rapidly adopted across enterprises, consumer markets, and operational environments. These systems can pursue goals, make decisions, interact with tools and services, and take actions with varying levels of autonomy, introducing distinct cybersecurity risks and risk management challenges. At the same time, agentic AI is becoming an increasingly important capability for applying AI to strengthen cybersecurity risk management programs, where agents can support security analysis, monitoring, and response. The Cyber AI Profile will help organizations address governance, oversight, access control, monitoring, validation, and resilience needs, all of which become especially significant when AI systems are empowered to act more independently or influence operational decisions. Further emphasizing Agentic AI in the Profile could support a broadly available, consistent, and common approach to managing emerging AI-related cybersecurity risks while enabling more informed, secure, and responsible adoption of this rapid development of AI applications.

Overview of comments

Analysis of the comments related to Agentic AI indicated that the Cyber AI Profile requires further revision to address Agentic AI considerations. Feedback generally fell into the following primary themes:

- Requests for significant revisions, including the addition of a new, separate section dedicated to Agentic AI security
- Requests to incorporate additional considerations and examples specifically tailored to Agentic AI within the Sample Focus Area Considerations and Sample Opportunities
- Requests to revise only selected CSF Categories and Subcategories, with varying levels of emphasis, by adding considerations and examples specifically tailored to Agentic AI

Proposed options forward on Integrating Agentic AI in the Profile:

1. **Option 1: Create a separate section dedicated to Agentic AI security to address relevant use cases.** This option would add a distinct section (for example, a new section in Section 1.2) to the draft Cyber AI Profile focused specifically on Agentic AI, including a range of relevant topics such as agent-based systems, autonomous decision-making, tool use, orchestration across multiple systems, and relevant deployment patterns. A dedicated section could describe the unique cybersecurity

considerations associated with Agentic AI, such as expanded attack surfaces, goal hijack, insecure tool access, privilege misuse, unintended tool usage, and challenges related to monitoring and human oversight. This approach would allow the Profile to clearly acknowledge the growing importance of Agentic AI.

- Option 2: Update the draft to emphasize Agentic AI throughout the Cyber AI Profile, at the Subcategory level for each Focus Areas.** This option would take an approach to make Agentic AI more prominent by revising the draft so that Agentic AI is treated as a cross-cutting theme throughout the Profile rather than as a standalone topic. General Considerations, Sample Focus Area Considerations and Sample Opportunities in each Focus Area would include examples or considerations that specifically address AI agents’ applications (see a mock-up example for PR.AA-01 below).

CSF 2.0 Core: PROTECT	General Considerations	Focus Area Proposed Priorities & Considerations		
		Secure	Defend	Thwart
PROTECT (PR)	Safeguards to manage the organization’s cybersecurity risks are used			
Identify Management, Authentication, and Access Control (PR.AA)	Access to physical and logical assets is limited to authorized users, services, and hardware and managed commensurate with the assessed risk of unauthorized access			
PR.AA-01: Identities and credentials for authorized users, services, and hardware are managed by the organization	<p>General Considerations: Give AI systems unique and traceable identities and credentials to better track their activity.</p> <p>General Agentic AI Considerations: Non-human identities with unique, traceable credentials and tightly scoped access rights with enforcing least-privilege access to data, models, tools, and APIs.</p> <p>Example Informative References: NIST SP 800-53, Rev 5: AC-01; AC-02; AC-14; IA-01; IA-02; IA-03; IA-04; IA-05; IA-06; IA-07; IA-08; IA-09; IA-10; IA-11</p>	<p>Proposed Priority: 1</p> <p>Sample Focus Area Considerations: AI systems may need their own identities and credentials (i.e., AI service level accounts) to interact with a broader system. Organizations need traceability between AI systems and their actions.</p> <p>Agentic AI Considerations: Unique non-human identities with traceability and securely storing secrets in approved vaults rather than embedding them in code, prompts, or configuration files.</p> <p>Example Informative References: DASF 1, 3, 21-22, 32, 40, 48; ATLAS AML.M0005; ATLAS AML.M0019; OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: Model Access Control; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025</p>	<p>Proposed Priority: 2</p> <p>Sample Opportunities: AI catches credential misuse that previous rules might miss by flagging unusual authentication activity.</p> <p>Sample Focus Area Considerations: Assign and manage unique and traceable identities and credentials to AI defense agents to support defensive response activities.</p> <p>Agentic AI Opportunities: AI agents monitor identity and access activity for signs of credential misuse or unauthorized behavior, and automatically deploy protective actions such as step-up authentication, token revocation, account lockout, or access restriction.</p> <p>Agentic AI Considerations: High degree of autonomy in applying defense measures.</p> <p>Example Informative References: DASF 39; WASP AI Exchange: Model Access Control; ENISA Threat Landscape 2025; ATLAS AML.M0005</p>	<p>Proposed Priority: 1</p> <p>Sample Focus Area Considerations: Standard cybersecurity practices apply. (Rationale) AI-enabled cyber attacks will lower the barrier of entry to gaining access to identities and credentials, services, and hardware.</p> <p>Agentic AI Considerations: High degree of autonomy in applying defense measures in machine speed such as adaptive authentication, session termination, token revocation, and temporary account lockout. These measures help limiting the speed and scale of AI-enabled attacks.</p> <p>Example Informative References: NIST SP 800-172 – Configuration Management (3.4); NIST SP 800-172 Identification and Authentication (3.5); NIST SP 800-207 – Zero Trust Architecture; ATT&CK M1032; ATT&CK M1027; ATLAS AML.M0005; ATLAS AML.M0019; AI 100-2e2025</p>

SAMPLE TEXT PROVIDED FOR DISCUSSIONS PURPOSES ONLY

- Option 3: Revise the draft in selected CSF Subcategories by adding considerations and examples specifically tailored to Agentic AI.** This option would provide a more targeted approach by identifying the Subcategories where Agentic AI introduces the most significant or distinctive cybersecurity implications and then updating those sections accordingly. Revisions could range from brief additions of agent-specific examples to more substantial enhancements in areas such as governance, identity and access management, system monitoring, third-party dependencies, validation, and response planning. This approach would enable the Profile to address the most relevant Agentic AI issues while maintaining the existing structure. Agentic AI content is interwoven into the existing content

alongside other types of AI (see a mock-up example for PR.AA-01) without overemphasizing one type of AI. **(NIST’s preferred approach)**

CSF 2.0 Core: PROTECT	General Considerations	Focus Area Proposed Priorities & Considerations		
		Secure	Defend	Thwart
PROTECT (PR)	Safeguards to manage the organization’s cybersecurity risks are used			
Identify Management, Authentication, and Access Control (PR.AA)	Access to physical and logical assets is limited to authorized users, services, and hardware and managed commensurate with the assessed risk of unauthorized access			
PR.AA-01: Identities and credentials for authorized users, services, and hardware are managed by the organization	<p>General Considerations: Give AI systems unique and traceable identities and credentials to better track their activity with unique, traceable credentials and tightly scoped access rights with enforcing least-privilege access to data, models, tools, and APIs.</p> <p>Example Informative References: NIST SP 800-53, Rev 5: AC-01; AC-02; AC-14; IA-01; IA-02; IA-03; IA-04; IA-05; IA-06; IA-07; IA-08; IA-09; IA-10; IA-11</p>	<p>Proposed Priority: 1</p> <p>Sample Focus Area Considerations: AI systems may need their own identities and credentials (i.e., AI service level accounts) to interact with a broader system. Organizations need traceability between AI systems including AI agents and their actions.</p> <p>Example Informative References: DASF 1, 3, 21-22, 32, 40, 48; ATLAS AML.M0005; ATLAS AML.M0019; OWASP AI Exchange: Controls to Limit the Effects of Unwanted Behavior; OWASP Conventional Runtime Controls; OWASP AI Exchange: Model Access Control; OWASP LLM Top Ten: LLM03 Supply Chain; ENISA Threat Landscape 2025; AI 100-2e2025</p>	<p>Proposed Priority: 2</p> <p>Sample Opportunities: AI catches credential misuse that previous rules might miss by flagging unusual authentication activity. And AI agents can automatically deploy protective actions such as step-up authentication, token revocation, account lockout, or access restriction.</p> <p>Sample Focus Area Considerations: Assign and manage unique and traceable identities and credentials to AI defense agents to support defensive response activities with AI agent-enabled high degree of autonomy in applying defense measures.</p> <p>Example Informative References: DASF 39; WASP AI Exchange: Model Access Control; ENISA Threat Landscape 2025; ATLAS AML.M0005</p>	<p>Proposed Priority: 1</p> <p>Sample Focus Area Considerations: Standard cybersecurity practices apply leveraging AI agent-enabled autonomy in applying defense measures in machine speed to limit the speed and scale of AI-enabled attacks.</p> <p>(Rationale) AI-enabled cyber attacks will lower the barrier of entry to gaining access to identities and credentials, services, and hardware.</p> <p>Example Informative References: NIST SP 800-172 – Configuration Management (3.4); NIST SP 800-172 Identification and Authentication (3.5); NIST SP 800-207 – Zero Trust Architecture; ATT&CK M1032; ATT&CK M1027; ATLAS AML.M0005; ATLAS AML.M0019; AI 100-2e2025</p>

SAMPLE TEXT PROVIDED FOR DISCUSSIONS PURPOSES ONLY

NIST has identified Option 3 (revise the draft in selected CSF Subcategories by adding considerations and examples specifically tailored to Agentic AI) as the most viable option. This approach reflects the following considerations:

- The current structure of the Cyber AI Profile has been well received, and Option 3 provides a technology-neutral, balanced, and proportionate way to address the comments without significantly altering this structure.
 - Offers a technology-neutral, consistent treatment of Agentic AI among other types of AI that are addressed in the Profile by only calling out a specific type of AI when there is value in doing so and otherwise not singling out specific types of AI.
 - Responds meaningfully to community feedback while maintaining its overall structure and intent by revising selected CSF Subcategories and adding Agentic AI-specific considerations and examples where most relevant.
 - Allows the revisions to reflect varying degrees of relevance across different sections of the Profile. Not all CSF Subcategories are affected equally by Agentic AI considerations, so a targeted approach makes it possible to emphasize agentic considerations where they are most impactful, while avoiding unnecessary expansion in areas where such considerations are less essential.
 - Supports better integration with the existing content of the Cyber AI Profile. Rather than separating Agentic AI from the rest of the content or recasting

the entire document around it, this approach incorporates agentic considerations directly into the relevant cybersecurity outcomes, helping preserve consistency, readability, and alignment with the Profile structure.

- Option 2 places disproportionate emphasis on Agentic AI topic relative to other important AI cybersecurity considerations. The Profile is intended to provide structured and technology neutral recommendations to support a broadly applicable and balanced approach to AI cybersecurity risk management, not to make Agentic AI overly prominent as a cross-cutting theme throughout the Profile.
- Option 1 may present challenges for usability and coherence. Placing Agentic AI in a standalone section could make it harder to integrate those considerations consistently throughout the rest of the document and more difficult for organizations to apply the guidance in practice.

The following criteria are being considered to guide decisions regarding whether and how much Agentic AI content is necessary throughout the Profile:

- Whether the topic is broadly applicable across most Agentic AI use cases or mainly relevant to higher-risk or special applications
- Whether the topic should receive primary emphasis in the Profile, more limited treatment, or be included mainly as an example or supplemental consideration
- Whether the topic is sufficiently distinct to warrant Agentic AI-specific consideration

The following areas reflect examples of how Agentic AI considerations may be integrated across CSF Functions. The examples of technical topics below (grouped by CSF Functions) are intended to support discussions during the COI working session, and to provide insight into more detailed approaches for addressing varying degrees of emphasis and relevance to Agentic AI use cases and considerations:

Govern (GV):

- **Agentic AI governance and risk management:** establish security policies for agents, including permitted levels of autonomy, acceptable tool use, data handling requirements, and risk classification based on use case, potential impact, and access level
- **Human oversight and decision authority and accountability:** define when human-in-the-loop or human-on-the-loop review is required, especially for high-

impact, irreversible, or high-risk actions; include approval workflows, escalation paths, decision traceability and accountability, and emergency procedures

- **Policy-based control of high-risk actions:** use a central policy approach to govern agent actions, including step-up approval, transaction limits, and explicit authorization for sensitive operations
- **Access governance and periodic review:** conduct periodic reviews of agent identities, permissions, tool access, and delegated privileges to ensure that access remains appropriate over time
- **Legal, regulatory, and organizational alignment:** align agent security practices with applicable privacy, cybersecurity, records management, and sector-specific regulatory requirements

Identify (ID):

- **Agent, tool, and service identity architecture:** unique agent identity enabling clear accountability, precise access control, and reduced risk by eliminating shared identity or credentials across agents and services
- **Tool and connector inventory:** agent-accessible tools, plugins, APIs, SDKs, and external services and actions are clearly identified and documented to support transparency, governance, and effective control of agent behavior
- **Dependency and supply chain visibility:** maintain visibility into model sources, tool providers, plugins, third-party dependencies, and supporting infrastructure to better understand upstream risk
- **Environment, tenant, and trust boundary definition:** define separation across development, test, and production environments; enforce tenant boundaries; and identify where cross-environment or cross-tenant interaction could create risk
- **Use case and action mapping:** use cases such as autonomous decision-making, sensitive data access, external tool control, or operational impact are clearly identified to enable security oversight and protections applied in proportion to their risk

Protect (PR):

- **Identity, authentication, and authorization:** agent operates with the minimum access required, using time-bound privileges or zero-standing privileges, tightly

scoped permissions where feasible to further reduce exposure and limit the impact of misuse.

- **Credential lifecycle management:** prefer short-lived credentials and automate issuance, rotation, renewal, revocation, and disablement; eliminate long-lived static keys where possible
- **Secrets management:** secrets are protected from exposure in storage and during use, kept out of prompts, code, and unsecured configuration, redacted from outputs and logs, and accessible only through enforced policy controls
- **Tool use and execution security:** agents operate within approved tool, action, and execution boundaries, such as with validated tool use, isolated runtimes, and restricted file system, process, and network access to reduce unintended behavior and contain potential impact
- **Prompt injection and instruction-hijacking defenses:** treat all external content as untrusted, separate system, developer, user, and tool contexts, apply input sanitization and policy checks, and require robust gating before tool calls or sensitive actions are executed
- **Data protection:** minimize the data sent to models, encrypt data in transit and at rest, enforce tenant isolation, detect and redact sensitive data where appropriate, and apply retention and deletion controls
- **Output safety and action validation:** use policy-based output filtering, structured outputs, schema validation, and confirmation or secondary validation for high-risk or high-consequence actions
- **Supply chain integrity protections:** use signed artifacts, pinned versions, dependency scanning, and secure endpoints for models, tools, and supporting components

Detect (DE):

- **Logging and auditability:** maintain logs of prompts, retrieved content, tool calls, actions taken, applicable user and agentic identities, approvals granted, and security-relevant decisions to support traceability and accountability
- **Immutable and correlated audit trails:** correlate agent actions to requests from users and other agents, workflow context, tool usage, and approval events, and preserve audit records in a tamper-resistant manner where appropriate

- **Continuous monitoring and anomaly detection:** detect unusual permission use, token abuse, suspicious prompt patterns, excessive or unexpected tool calls, policy violations, and indicators of possible exfiltration or misuse
- **Behavioral observability:** monitor for deviations from expected agent behavior, such as unexpected escalation of privileges, unusual action sequences, or actions outside normal operating patterns

Respond (RS):

- **Agentic AI incident response playbooks:** develop and maintain playbooks for prompt injection, instruction hijacking, unauthorized tool use, data exfiltration, credential compromise, privilege misuse, and unintended autonomous actions.
- **Identity and access containment:** support rapid disablement or quarantine of agent identities, suspension of tool access, and isolation of affected workloads or connectors
- **Controlled emergency access:** provide break-glass mechanisms for urgent intervention, with strong approval, monitoring, and auditing controls
- **Escalation and coordination:** define clear escalation paths among security, operations, AI, privacy, and legal teams when an agent-related event has broader operational or compliance implications

Recover (RC):

- **Resilience and fail-safe behavior:** implement rate limiting, fail-closed behavior where appropriate, rollback capability, and kill-switch mechanisms to reduce the impact of agent malfunction or misuse
- **Recovery and restoration planning:** establish procedures for restoring services, revalidating agent behavior, reissuing credentials, and safely returning tools or workflows to operation after an incident
- **Validation after recovery:** test that containment and corrective actions were effective before re-enabling agent access or resuming autonomous functions
- **Exercises and continuous improvement:** conduct regular red-team exercises, simulations, and post-incident reviews to improve resilience and refine controls for future revisions of the Profile

Open questions for discussion during the working sessions:

- What considerations do agents introduce which go beyond the considerations laid out for AI systems broadly?
 - Some examples include autonomy related risks (e.g., goal drifting), tool misuse (e.g., indirect prompt injection), long-running agent (e.g., memory poisoning)
- Which important distinct types of risk are introduced by agentic AI applications that are not already identified in the current draft? What associated references are available for those types of risk?
- What significant gaps do current practices leave when addressing the unique characteristics of AI agents in each Focus Area?
- What other special considerations for AI agents are important but have not been sufficiently covered?

Zero Trust (ZT)

Context

ZT principles and best practices can be applied to AI agents, assuming by default that no user, device, model, or automated agent is trustworthy. As AI agents gain access to data, tools, and decision-making workflows, organizations must continuously verify identity, enforce least-privilege access, provide Just In Time (JIT) access, validate the security posture of the agent's execution environment, monitor behavior, and validate every action or request in context. Recognizing the relationship between ZT principles and agent security, ZT is the second technical area of discussion during the second COI working session.

ZT is important to the Cyber AI Profile because AI systems often operate across complex environments involving users, models, data sources, tools, APIs, cloud services, and third-party components, all of which can expand the attack surface and increase cybersecurity risk. A ZT approach helps address these risks by applying techniques and best practices such as continuous verification, least-privilege access, unique identity and authentication, zero standing privileges, time-bound privileges, segmentation, monitoring, and explicit control over interactions among AI components and supporting systems. These principles are especially relevant for AI deployments that access sensitive data, influence decisions, or take actions across organizational or trust boundaries. Incorporating ZT considerations into the Cyber AI Profile can therefore help organizations apply consistent and risk-informed safeguards as they adopt AI within their cybersecurity risk management

programs, while strengthening resilience, limiting unauthorized access, and reducing the potential impact of compromise.

NIST NCCoE has previously demonstrated the implementation of Zero Trust Architectures (consistent with NIST SP 800-207) in NIST SP 1800-35, which includes mappings between ZT concepts and the NIST CSF. This work can provide a foundation for applying ZT principles in AI environments.

Overview of comments

Analysis of the comments related to ZT indicated that the current draft Profile may require further revision to more clearly and consistently include ZT considerations in the context of AI systems. Community feedback suggests that the Preliminary Draft needs greater emphasis on ZT principles, particularly in areas such as Just-In-Time identity authentication and access control, segmentation, and trust boundaries across AI components and supporting infrastructure. This feedback indicates a need to further examine how ZT considerations should be reflected in the Profile so that it can better support organizations adopting AI in complex and distributed environments.

Proposed options forward on Zero Trust:

1. **Option 1: Defer incorporating revisions related to applying Zero Trust principles or considerations, since Zero Trust is addressed in other NIST publications.** This option would leave the current draft Cyber AI Profile largely unchanged with respect to Zero Trust and reference existing NIST Zero Trust publications as complementary guidelines. Under this approach, the Profile could reference those publications but would not introduce new Zero Trust-specific revisions in the Profile itself.
2. **Option 2: Add a brief section dedicated to applying Zero Trust principles.** This option would introduce a concise section in the draft Cyber AI Profile that explains the relevance of Zero Trust to AI systems and highlights key principles such as strong identity, least-privilege access, continuous verification, segmentation, and monitoring. A dedicated section could provide useful context and acknowledge the importance of Zero Trust without requiring extensive changes throughout the rest of the document.
3. **Option 3: Address the comments related to applying Zero Trust principles within selected CSF Subcategories by adding relevant considerations.** This option would incorporate Zero Trust more directly into the Profile by revising selected CSF Subcategories to include considerations relevant to applying Zero Trust principles in AI environments. **(NIST's preferred approach)**

NIST has identified Option 3 (address the comments related to applying Zero Trust principles within selected CSF Subcategories by adding relevant considerations) as the most viable option. This approach reflects the following considerations:

- Responds to community feedback while maintaining the integrity of the Profile's structure. Option 3 integrates ZT considerations directly into the CSF-based structure of the Cyber AI Profile and makes the guidance more actionable by linking ZT principles to specific cybersecurity outcomes and relevant Profile content.
- Directly addresses the feedback received requesting stronger treatment of ZT in the Profile, minimizing the likelihood of any gap in guidance to organizations on how ZT principles apply specifically in AI contexts.
- Option 3 does not treat ZT as a standalone concept; instead, it embeds ZT considerations directly within the CSF elements in the Cyber AI Profile, reducing ambiguity for users in how to operationalize ZT principles alongside AI.

Open questions for discussion during the working sessions:

- What special considerations do AI and agentic AI applications introduce when applying ZT practices beyond those already established for ZT more broadly?
- What zero trust implementation differences might occur with AI Agents?
- Which Zero Trust concepts are most applicable to AI agent security?

Some examples:

- Identity and access control (PR.AA) (e.g., agent identity and ownership, short-lived and just-in-time access tokens, limited scope of privilege [least privilege])
- Platform and data security (PR.PS, PR.DS) (e.g., agents, tools, and data registration, data integrity and provenance)
- Continuous monitoring (DE.CM) (e.g., guardrails, HITL/HOTL, audit logs)
- Governance and risk management (GV.RM) (e.g., address AI agent related risks)

Next Steps

We welcome feedback on these topics during the upcoming COI working sessions and encourage you to register using the links in the table below. In addition to this discussion essay, we also encourage participants to review the Cyber AI Profile [Preliminary Draft](#) prior to the working sessions to best prepare for the informed, targeted discussions. If you are unable to attend or want to provide written feedback, you are welcome to share it via email with cyberaiprofile@nist.gov by May 15, 2026.

Session	Description	Date/Time
Session #1: Updates to Profile Elements and Contents	This session discussed approaches to addressing feedback received regarding Profile content, including clarifying approaches and phrasing for Profile elements such as the priorities and considerations in Section 2 of the Preliminary Draft. Please also see the Session #1 discussion essay .	April 28, 2026 / 1:00-4:00 P.M. EDT
Session #2: Extending the Technical Content	This session will explore how the Profile is being strengthened in critical technical areas including Agentic AI and Zero Trust	May 5, 2026 / 1:00-4:00 P.M. EDT
Session #3: Usability of the Profile	This session will explore different delivery formats to ensure the Profile meets the needs of different stakeholders	May 12, 2026 / 1:00-4:00 P.M. EDT

These discussions are intended to inform NIST's efforts to craft the next Cyber AI Profile draft. Please stay tuned for the release of the Discussion Essay for Session #3. We look forward to hearing from you!