Cybersecurity Coalition

January 30, 2026

Comments of the Cybersecurity Coalition to The National Institute of Standards and Technology (NIST)

Re: Request for comments on the preliminary draft NIST IR 8596, Cybersecurity Framework Profile for Artificial Intelligence.

The Cybersecurity Coalition (the Coalition) submits this comment in response to NIST's request for comments on NIST IR 8596 irpd: Cybersecurity Framework Profile for Artificial Intelligence (the Cyber AI Profile, or Profile).[1]

The Coalition is composed of leading companies with a specialty in cybersecurity products and services dedicated to finding and advancing consensus policy solutions that promote the development and adoption of cybersecurity technologies. We seek to ensure a robust marketplace that will encourage companies of all sizes to take steps to improve their cybersecurity risk management. We are supportive of efforts to identify and promote the adoption of cybersecurity best practices, information sharing, and voluntary standards throughout the global community. We support the direction of the Profile and believe it reflects many of the principles the Coalition advocated in its earlier comments.

The Coalition appreciates NIST's ongoing leadership in developing the Cyber AI Profile, and is pleased that the draft reflects many of the priorities we shared in our comments[2] to the *Cybersecurity and AI Workshop Concept Paper.* The intersection of AI and cybersecurity is a timely and significant topic for guidance from NIST that integrates and builds on the widely accepted, and widely adopted, Cybersecurity Framework (CSF). The Cyber AI Profile represents an opportunity to help organizations secure their AI systems as they work to deploy more, and more varied, AI systems. We appreciate the invitation for feedback on this draft. Our recommendations focus on sharpening clarity, improving usability, and ensuring that the Profile fully delivers on those goals.

[1] https://www.nccoe.nist.gov/projects/cyber-ai-profile
[2] https://www.centerforcybersecuritypolicy.org/insights-and-research/ai-profile-for-nist-csf-would-help-risk-management-pros

1. Foundational Strengths of the Draft Cyber AI Profile

The Coalition is pleased that the Cyber AI Profile situates AI cybersecurity within the existing Cybersecurity Framework. AI introduces new challenges regarding cybersecurity, but AI advances do not - in and of themselves - necessarily require fundamental changes to the way organizations address cybersecurity. Applying Zero Trust principals to AI is an important framework and discipline can enable organizations to better understand the context of AI uses. By anchoring AI-related considerations to CSF 2.0 outcomes, the draft reinforces the principle that AI systems are part of an organization’s broader digital infrastructure and should be governed through established, risk-based cybersecurity practices. This approach avoids unnecessary fragmentation and supports integration into existing enterprise risk management programs. Building on this strong foundation, we offer suggestions to further ensure that the Profile is fully integrated with existing NIST standards and is readily adoptable and operationalized by organizations in practice.

We also appreciate that the draft recognizes that AI system risk must be managed as a lifecycle and systems issue, not simply based on evaluating a model or specific tool. AI risk can come from any part of the lifecycle, and must be addressed as they arise - whether during design, deployment, integration, or operation. Additionally, AI systems are increasingly dependent on data, infrastructure, and third-party services that must be addressed as part of cybersecurity risk management. A systems-oriented perspective is necessary to manage real-world risk. We offer additional ideas to ensure that the entire systems-based lifecycle is addressed.

We also agree that AI-enabled tools are an increasingly essential element of cybersecurity defense. Including discussion of AI-enabled defense as a core focus reflects this reality, important where adversaries also leverage AI. Defensive AI can be enabled through effective risk management. We offer thoughts on how best to frame this within the context of the CSF.

Finally, we appreciate that the Cyber AI Profile maintains a voluntary, technology-neutral, and risk-based posture for AI systems. Avoiding prescriptive requirements on specific technologies or architectures is critical given the pace of change in AI development and operations. Ensuring that the Profile is flexible enough to support innovation, accommodate the wide variety of organizational contexts in which AI is deployed, and supports innovation will help the Profile remain relevant and flexible over time.

2. Suggestions for Broad Improvements

While the Profile contains many strong elements, it would benefit from clearer framing, stronger integration with existing NIST frameworks and other widely accepted standards, and additional exploration of the systemic aspects of AI-related cyber risk.

2.1 The Profile Needs Clearer Framing, and Focus Areas Currently Confuse the Matter

The purpose (1.1) and scope (1.2) of the Profile are strong. Integrating AI-specific recommendations into current governance frameworks is important, given the broad and accelerating integration and use of AI systems within digital and hybrid systems.

The three Focus Areas (Secure, Defend, and Thwart) are crucial use cases of AI in cybersecurity, but should be defined as use cases rather than as cross-cutting matrixed additions to the CSF pillars. As organizations continue to expand their knowledge of AI systems relating to cybersecurity, it might be confusing to focus on the use cases so explicitly, while the controls themselves should be done across all three. Introducing three additional focus areas alongside the five CSF functions requires organizations to reason about each control through multiple, overlapping lenses. This is likely to increase implementation friction, blur the conceptual boundaries among the CSF functions (Govern, Identify, Protect, Detect, Respond, Recover), and add complexity that is not warranted given the inherently flexible, technology-agnostic design of the CSF.

The Cyber AI Profile should instill confidence in the CSF. It should be stated clearly that this Profile is to highlight the differences between certain AI use-cases, but not to drastically change how it is used. AI is software and in the majority of instances should be treated as such.

2.2 The Profile Should Take a Systemic View of AI-Related Cyber Risk

The Cyber AI Profile should highlight how AI risks extend beyond system features and traditional technical bugs. Design and implementation failures can introduce security vulnerabilities that are distinct from traditional cybersecurity vulnerabilities. The data supply chain, flawed optimization objectives, and unsafe training procedures are just a few examples of

unique, systemic risks related to AI. Secure engineering practices do lead to better cybersecurity outcomes, and the Profile does a good job of outlining many important controls and practices to support this. But it should also introduce systemic risk related to design and implementation that are just as important. The Profile should include relevant controls, mapped to the CSF 2.0, to ensure that these issues are not passed by.

The Profile would also benefit from adopting an AI lifecycle-based perspective on cybersecurity risk. Threats to AI span safety, security, runtime, supply chain, model and system behavior, input manipulation, harmful outputs[3], and much more. Each of these domains is connected to the others and should be treated as such. An approach tailored to the entire lifecycle of AI would help organizations identify relevant controls and to avoid gaps between development and deployment.

The Profile could also emphasize that Zero Trust architecture provides a unifying framework across all three Focus Area use cases. Zero Trust principles apply whether organizations are securing AI systems (Secure), using AI for defense (Defend), or protecting against AI-enabled attacks (currently, “Thwart” in the draft profile). By assuming breach, verifying continuously, and enforcing least-privilege access, Zero Trust addresses the reality that AI systems introduce expanded attack surfaces, operate at machine speed, and exhibit behaviors that traditional perimeter-based security cannot adequately govern.

The Cyber AI Profile provides an opportunity to set the taxonomy of AI threats in the context of the CSF. High-level terms like “prompt-injection” can minimize differences in attacker techniques and objectives and their impacts across the AI lifecycle. Incorporating or aligning with existing AI threat taxonomies would improve clarity and strengthen the Profile’s ability to address the full range of AI threats. AI systems are frequently non-deterministic, rely on human language, and increasingly operate across multimodal inputs and outputs. Emerging use cases, including agentic and agent-to-agent systems introduce additional risks. Securing AI agents should follow a secure-by-design, defense-in-depth model, with controls enforced at both the network and application layers to protect agents across discovery, identification, and runtime interactions. AI agent identification and discovery should rely on standardized, open protocols to ensure interoperability and portability across diverse environments, while avoiding proprietary silos and vendor lock-in. To address this complexity, the Profile should clarify taxonomies that distinguish types of threats across the AI lifecycle, including for agentic AI.

3. Recommendations by Focus Area

We recommend de-emphasizing the three focus areas as a primary organizing construct within the Profile, as they risk being interpreted as complications to controls or CSF pillars rather than as descriptive use cases. If the focus areas are retained, however, the following recommendations are intended to strengthen their clarity and usefulness. In that context, they can serve as a helpful way to understand different AI-related cybersecurity use cases. Section

[3] Integrated AI Security and Safety Framework https://learn-cloudsecurity.cisco.com/ai-security-framework

2.1 provides a useful explanation of how the three areas enable and reinforce one another, and these interactions should be elevated and reflected more consistently throughout the Profile. Additionally, any distinct risk management and cybersecurity considerations associated with these use cases should be more clearly articulated, which would help further differentiate them from controls and CSF pillars.

3.1 Securing AI System Components ("Secure")

The *Secure* section (Section 2.1.1) appropriately focuses on the new threats and vulnerabilities introduced by integrating AI into existing infrastructures. Consistent with CSF 2.0 and other standards, it is important to acknowledge that securing AI systems shares many similarities with securing other types of software. At the same time, the *Secure* section should clearly emphasize what makes AI security distinct and provide examples.

The Profile does note that, compared to other computer systems, "AI behavior and vulnerabilities tend to be more contextual, dynamic, opaque, and harder to predict, as well as more difficult to identify, verify, diagnose, and document."[4] It should also discuss the kinds of AI-specific security considerations that make AI security, including within the CSF, distinct. Model supply chain threats (training data leakage, model tampering, model theft, and supply chain compromise), prompt-manipulation attacks (prompt injection, jailbreaks, context manipulation), threats related to non-deterministic and emergent behavior, multi-modal threats, and agentic and agent-to-agent threats are just some of the unique threats that should be more directly addressed given their growing prevalence and their potential to undermine system reliability and trustworthiness. In addition, risks associated with third-party AI components should be elevated and treated as first-order supply chain concerns, especially where organizations rely on externally developed models, hosted services, or API-based integrations.

The Profile would also benefit from recommending security practices that extend beyond traditional preventive and detective controls. These include continuous model evaluation and validation throughout the system lifecycle, AI-specific red-teaming to identify emergent and context-dependent failure modes, and continuous monitoring of AI systems in deployment to detect drift, anomalous behavior, and evolving threat conditions.

3.2 Conducting AI-Enabled Cyber Defense ("Defend")

The *Defend* section (Section 2.1.2) provides clear examples of how AI can enhance defensive processes by "augmenting human analysts, enhancing detection and response time, and supporting recovery." These identified use cases - Mission Assurance, Predictive and Proactive Activities, Investigation and Analysis, and Response and Remediation - are distinct and clearly articulated.

[4] https://nvlpubs.nist.gov/nistpubs/ir/2025/NIST.IR.8596.iprd.pdf

While the Profile references detection and response, attack surface analysis, and predictive threat modeling, it should provide more specific direction on how organizations can adapt these capabilities to the rapidly evolving AI threat and operational environment.

To strengthen the *Defend* section, the Profile should more clearly characterize AI-enabled defense use cases, including detection and response, attack surface analysis, and predictive threat modeling. It should provide guidance on the considerations for appropriate levels of autonomy for these capabilities, including discussion of "human in the loop" controls that may seem in tension with taking advantage of the autonomy that AI can deliver in fast-moving and information-rich situations. It should also emphasize "human-in-the-loop" approaches, in which AI systems operate within defined confidence thresholds and escalation criteria, ensuring that human oversight is maintained for high-impact or ambiguous decisions.

3.3 Thwarting AI-Enabled Cyber Attacks ("Thwart")

The *Thwart* section (Section 2.1.3) would benefit from a clearer explanation of how thwarting AI-enabled cyber attacks differs from the activities described in the *Secure* and *Defend* sections, particularly with respect to securing systems and building operational resilience. It also seems less specific to the defense of AI systems than other elements of the Profile. The Profile, and the CSF 2.0, has historically emphasized defensive measures focused on preventing, detecting, and responding to attacks. Within this context, the *Thwart* section should more clearly explain how its recommended actions align with, or extend beyond, this established framing.

We recommend explicitly clarifying whether this use case is primarily intended to support resilience, adversary disruption, detection of AI-enabled attack techniques, or another distinct objective. Alternatively, adopting revised terminology or more tightly aligning this section with existing CSF 2.0 concepts, such as *Response* and *Recovery*, may enable smoother integration into organizations' operational strategies. Additionally, focusing on concrete and measurable controls is beneficial, and there is not a clear way to measure controls of, or the success of, "thwart."

We would also support removing this focus area. While it is a useful case, "thwart" seems to be a departure (in both substance and word choice) from the defensive focus of the CSF and the goals of the Profile itself. We propose substituting the word with a less aggressive alternative, assuming the focus area is not removed.

4. Generalized Topic Gaps

4.1 Governance

Governance is underspecified and underrepresented throughout the Profile. While individual controls are mapped to the CSF 2.0 Govern function, governance as a cross-cutting capability is not emphasized at the necessary level within the Profile narrative. Although technical controls are essential, the Profile should clearly articulate that effective governance is particularly critical for AI systems given the pace of technological change, the nondeterministic nature of some AI

systems, and their increasing integration. Governance can be a mechanism that allows organizations to adapt their risk management over time, and as technology and threats advance.

It is difficult for frameworks, especially those involving rapidly advancing technologies, to keep up. The Profile should contain mechanisms that allow for adaptability without requiring official updates. Maintaining technology-agnostic approaches, for example, allows for adaptability and usability even as technologies change. Mature governance programs can help ensure that risk is well managed.

The Profile should place greater emphasis on high-level governance objectives and strategies, in addition to specific controls, and more closely align them with the CSF 2.0 Govern function, the AI RMF, and other recognized standards. Governance guidance should explicitly address risk appetite for AI-enabled systems, accountability for autonomous or semi-autonomous actions, and broader risks with cybersecurity implications, including system safety and reliability considerations where relevant.

4.2 Autonomy/Agents

AI systems have increasing autonomy, culminating in autonomous AI agents that can autonomously impact digital and physical systems. While autonomous and agentic AI are referenced, they should be more clearly defined, especially with regards to each CSF 2.0 function. We understand that there are other efforts underway to discuss the security of agentic AI systems, including CAISI's *Request for Information Regarding Security Considerations for Artificial Intelligence Agents*, the Cyber AI Profile needs to address these topics as well. As AI use cases advance, the Profile must be able to adapt - without requiring publishing a new draft. In addition to providing clear guidance on handling current and expected threats, the Profile should be adaptable and agile in the face of new and emerging threats.

Autonomous and agentic AI risks are currently underrepresented relative to their current, and expected, levels of implementation in organizations. The Profile should more explicitly articulate the range of human involvement in AI systems and explain how differing levels of autonomy affect risk, governance, and control expectations across the Profile. The Profile should also provide clear guidance on how agentic behaviors introduce different risk considerations such as:

- *Expanded threat surface:* AI agents don't just respond to inputs, they can initiate action, chain tools and processes, and work across multiple systems. These integrations mean that failures, errors, or compromises can propagate more quickly. Integrations and actions can be dynamic, and agents may make different decisions than humans would.
- *Agent as attack vector:* Agents may act as privileged API callers on behalf of users or systems, without appropriate authentication or guardrails against unintended outcomes.
- *Prompt injection:* Agents may ingest untrusted data continuously (often based on system integration) and may encounter embedded instructions or lead secrets.
- *Memory or state poisoning:* Agents maintain context (memory) across context, so malicious instructions or false beliefs may persist and influence future decisions.

- *Identity and delegation:* Agents act on behalf of users, systems, or organizations, but may not have appropriately strong agentic identity management or authentication.
- *Supply chain:* Agents rely on models, plugins, tools, data sources, and orchestration layers. This adds a level of complexity if there are failures or integrated malicious elements.

In general, and especially when referencing agentic AI, the Profile should work hand-in-hand with other NIST efforts such as the ongoing *RFI Regarding Security Considerations for Artificial Intelligence Agents.* We recognize these two efforts are separate, but also recommend greater collaboration and integration to ensure that an organization implementing agentic AI is easily able to reduce risk without becoming overburdened. It should be clear to an organization where to start and how to proceed when implementing risk management frameworks, including how the Cyber AI Profile impacts agentic AI.

4.3 Identity and Authentication

AI agents perform an increasingly broad range of tasks, including high-risk and real time activities, and identity, authentication, and access control are already foundational to effective AI cybersecurity. The rise of autonomous agents and agent-to-agent interactions will place new demands on identity management that are not addressed in the Profile, particularly in situations without a “human-in-the-loop”. To manage these risks, the Profile should emphasize managed AI identities with traceability and alignment with zero trust principles, while remaining technology-agnostic. Specific considerations should include: least-privilege access that adapts based on agent task context, continuous verification of agent identity and authorization (not just at initiation but throughout operation), time-bound credentials with automatic expiration and renewal, audit trails that capture the full chain of delegation when agents act on behalf of users or other systems, and mechanisms to rapidly revoke agent access across all connected systems when compromise is detected or suspected.

In addition, the Profile should acknowledge the emerging need to distinguish between human and AI agent identities, including how identities are assigned and governed across their lifecycles. While strong standards for AI agent identity management do not yet exist, the Profile could articulate desired outcomes, such as explicit labeling of agent identities, that allow organizations to differentiate agent actions from human actions. By framing these expectations at the outcome level, the Profile can provide near-term guidance while remaining flexible enough to incorporate future standards as they develop.

Given the adaptive and evolving nature of AI systems, continuous monitoring should be treated as a core requirement rather than a supplementary control. The Profile would benefit from greater clarity on how it will be maintained as AI capabilities and deployment models evolve, including whether interim guidance, modular updates, or targeted supplements are anticipated. We understand that multiple projects around agentic AI security and identity are in progress, and we strongly suggest integrating this work for inclusion in the Cyber AI Profile.

4.4 Interoperability Considerations

To support adoption and reduce duplicative compliance efforts, the Profile should be accompanied by explicit mappings to other relevant NIST frameworks and widely recognized international standards.

4.4.1 Stronger Integration with Existing NIST Frameworks

While other NIST frameworks such as the AI RMF, SP 800-53, and the Privacy Framework are referenced, the Cyber AI Profile should more explicitly explain how it works together with these other frameworks. A Profile should complement the CSF 2.0 and other frameworks within the NIST ecosystem, not compete. The Profile should clearly reference and align with relevant NIST and external efforts addressing agentic AI, including ongoing NIST work on AI risk management (including the AI Risk Management Framework).

A holistic risk-management approach would involve every framework, but it is unrealistic for many organizations, especially SMEs, to do an in-depth review of each of them. We recommend providing a clear mapping for where the Cyber AI Profile directly corresponds to each of the other relevant frameworks. In addition, there should be clear guidance for organizations on how to implement the entire range of frameworks and how they fit together. In particular, additional discussion of the AI RMF, 800-53, and the Privacy Framework would be useful; further discussion of complementary frameworks, including ISO 42001, Singapore's AIVerify, and ETSI's ETSI TS 104 223. Demonstrating how these domestic and international frameworks work together would not only aid organizational adoption but also reinforce NIST's leadership in promoting globally interoperable approaches to AI risk management.

We also recommend ongoing communication and cross-referencing with other projects at NIST to ensure that the Cyber AI Profile contains the most relevant and up-to-date information. For example, there is currently a *Request for Information Regarding Security Considerations for Artificial Intelligence Agents*[5] put out by the Center for AI Standards and Innovation (CAISI) of NIST. We recognize that these are separate efforts within NIST, and we will submit comments for the other RFI, but we strongly encourage ensuring that the messaging within the Cyber AI Profile is consistent with any guidance on relevant topics and includes relevant topics such as the implications of agentic AI.

Discussion of additional work would also benefit the Profile, as they identify novel failure modes and security risks unique to agentic systems.

4.4.2 Alignment with Internationally Recognized Standards

Beyond simple cross-referencing, the Profile should support interoperability through clear mappings, consistent terminology, and alignment of controls, enabling organizations to leverage

[5]https://www.federalregister.gov/documents/2026/01/08/2026-00206/request-for-information-regarding-security-considerations-for-artificial-intelligence-agents

existing risk management processes. Such interoperability is particularly critical for multinational enterprises, supply-chain risk management, and government procurement, where overlapping frameworks are unavoidable. Maintaining these mappings over time would reduce duplicative compliance efforts, accelerate adoption, and reinforce U.S. leadership by positioning the Profile as a globally usable, standards-aligned foundation for AI cybersecurity. In addition to vaguely referencing other controls, the Profile should include explicit references and mappings to:

- SP 800-53,
- ISO/IEC 27001/27002,
- CSA CCM,
- MITRE frameworks,
- Relevant CISA/NSA guidance,
- Other relevant international standards.

Interoperability supports adoption, reduces duplication, and strengthens U.S. leadership on the world stage. Whether or not the Profile itself is regularly updated, it should also acknowledge that organizations will need to continuously adapt to account for evolving technologies, threat models, and deployment patterns, using the Profile as a foundation rather than a static set of requirements.

4.4.3 Zero Trust Architecture

The Profile should explicitly recognize Zero Trust architecture as a foundational framework for securing AI systems and AI-enabled operations. Zero Trust principles—verify explicitly, use least-privilege access, and assume breach—are particularly well-suited to AI security challenges. AI systems operate at machine speed across distributed environments, process sensitive data continuously, and introduce non-deterministic behaviors that traditional perimeter-based security cannot adequately address. Zero Trust provides the architectural foundation for:

- Continuous verification of AI system identity and authorization, not just at deployment but throughout operation
- Least-privilege access for AI agents and models, ensuring they receive only the minimum permissions required for their current operation
- Inline inspection of AI traffic (including encrypted interactions) to detect threats and enforce policies in real time without relying on perimeter controls
- Micro-segmentation that limits lateral movement if an AI system is compromised
- Continuous monitoring and validation of AI system behavior against established baselines

The Profile currently references Zero Trust in passing but does not adequately emphasize its role as a cross-cutting architectural principle essential for all three Focus Areas. Organizations implementing the Profile should apply Zero Trust principles across AI asset management (Identify), access controls (Protect), threat detection (Detect), incident response (Respond), and

recovery procedures (Recover). This architectural approach enables organizations to secure heterogeneous AI deployments—vendor-managed models, internal models, embedded AI, and emerging components—through consistent controls rather than attempting to secure each AI system individually.

5. Generalized Cautions

As NIST continues to develop the Profile, it is important to recognize that some AI-related security considerations do not map cleanly to existing CSF functions or categories. In particular, decisions related to data sets, model development, and system integration can significantly influence downstream cybersecurity risk in ways that extend beyond traditional control-based approaches. Given the pace of technological change, the Profile will also need to evolve rapidly and remain technology-agnostic.

The Profile should also exercise caution when addressing emerging practices such as AI-specific software bills of materials (AI BOMs). While the value of increased transparency into model components, training data sources, and dependencies is clear, these practices are still maturing. Premature endorsement, without well-defined best practices or implementation guidance, could create confusion or impose burdens that outweigh near-term benefits. While standards for AI-BOMs are maturing, the requirements for visibility into AI libraries and models is a present-day necessity that can be enabled via automation. We recommend that NIST characterize AI BOMs as an evolving area of practice and coordinate closely with relevant stakeholders before making recommendations.

The Profile should also explicitly acknowledge that AI-related cybersecurity risk is highly context-dependent, varying based on system purpose, deployment environment, and potential impact. This variability should be addressed carefully within existing governance and risk management sections, rather than through rigid classifications or categorizations that could inhibit adoption or fail to scale as the technology evolves. The Profile should also consider uses of AI that are not known across an entire environment. The automated discovery of unmanaged AI deployments across multi-cloud environments, for example, would ensure that all of those deployments are adequately dealt with.

6. Conclusion

The Cybersecurity Coalition appreciates NIST's continued leadership in addressing the cybersecurity implications of artificial intelligence and for developing a Cyber AI Profile that builds on the widely adopted CSF 2.0. The draft reflects meaningful progress toward integrating AI into established, risk-based cybersecurity practices while remaining voluntary, technology-neutral, and adaptable to diverse organizational contexts.

We believe that clarifying the Profile's structure and framing, strengthening its treatment of systemic and lifecycle-based AI risks, elevating governance considerations, and improving

integration with existing NIST frameworks and internationally recognized standards will significantly enhance its usability and adoption. With these refinements, the Cyber AI Profile can serve as a practical, scalable foundation for organizations seeking to manage AI-related cyber risk without unnecessary complexity or fragmentation.

We encourage NIST to continue coordinating closely across its internal efforts, as well as with partners such as CISA, the Office of the National Cyber Director, and international standards bodies, to ensure consistent guidance and interoperability. The Coalition looks forward to continued engagement with NIST as the Profile evolves and stands ready to support its ongoing development and implementation.

APPENDIX - line edits

Section	Page #	Line #	CSF Core Element	Comment	Suggested Change
2.1	7-8	369-404		The current framing of the Focus Areas (Secure, Defend, Thwart) is interpreted as additional, matrixed CSF pillars rather than illustrative use cases.	Describe Focus Areas as common AI-related cybersecurity use cases, and can be used as contextual lenses for organizations to interpret and prioritize relevant CSF outcomes based on their AI deployment and threat environment. Potentially add “These Focus Areas are illustrative and should be mapped to existing CSF 2.0 functions and categories rather than implemented as a standalone control set.” Revise Section 1.3 and Section 2 to reflect this.
1.1, 1.2	3-5	251-325		The Profile should be clearer about its role in relation to the CSF.	Explicitly state that the Profile complements CSF 2.0 and does not introduce new CSF functions, pillars, or requirements. We suggest at the end of Section 1.1, with reinforcement in Section 1.2.
2.1.1	9-10	409-436		The draft emphasizes technical vulnerabilities but underplays systemic risks introduced through AI system design, data	Add “AI-related cybersecurity risk may arise not only from exploitable technical vulnerabilities, but also from system design decisions, data dependencies, optimization objectives, and integration choices. These risks may manifest across the

				dependencies, and optimization choices. Current guidance treats data as part of the broader supply chain but does not elevate data integrity to the level of software or hardware integrity.	AI system lifecycle and should be addressed using a systems-based, lifecycle-aware approach. Organizations should assess AI cybersecurity risk across the full lifecycle of an AI system, including design, data sourcing, training, integration, deployment, operation, maintenance, and decommissioning. Controls should be selected to prevent gaps between development and operational environments.”
2.1.1	9-10	409-436		The Secure section does not sufficiently distinguish AI-specific security characteristics from traditional software security.	Add “Compared to traditional software, AI systems may exhibit non-deterministic behavior, opaque decision processes, and context-dependent vulnerabilities. These characteristics introduce security risks that may be more difficult to predict, test, and document, including model tampering or theft, data poisoning, input manipulation, emergent behavior, and multimodal exploitation.”
2.1.1/General	9-10	409-436		Third-party AI components should be elevated as first-order supply chain risks. The draft mentions supply chain risks, but largely in reference to other documents.	“Third-party AI components, including externally developed models, hosted inference services, plugins, and APIs, should be treated as first-order supply chain dependencies and incorporated into supplier risk management practices.”
General				Autonomous and agentic AI systems are underdefined and underrepresented given their current and expected deployment. The draft should include guidance on how to manage varying levels of autonomy in AI-enabled defensive systems.	Add significant discussion throughout the draft of how different levels of autonomy, including agentic AI systems, impact these considerations.

2.1.3	12-13	496-536		The relationship between the Thwart Focus Area and existing CSF Respond and Recover functions is unclear. “Thwart” should be clarified or revised, and the relation clarified.	Add “This Focus Area addresses activities intended to detect, disrupt, or mitigate AI-enabled attack techniques and to enhance system resilience. These activities generally align with CSF 2.0 Respond and Recover functions and should be implemented accordingly.” Revise the section as appropriate; the term could also be replaced with something less leading.
General				Governance is treated primarily through individual controls rather than as a cross-cutting capability essential for AI risk management. In addition to mapping CSF core elements, the draft should include overall discussion of governance.	“Effective governance is particularly critical for AI systems due to their potential autonomy, non-deterministic behavior, and rapid evolution. Governance mechanisms enable organizations to adapt cybersecurity risk management practices over time without relying on prescriptive technical controls.”
General				The Profile does not sufficiently address AI-specific identity and access management challenges, particularly for agents.	Add additional discussion of identity management for AI systems, especially agents.
General				The Profile mentions Zero Trust in some places but does not adequately emphasize it as a foundational architectural principle for AI security. Zero Trust is particularly well-suited to AI challenges: AI systems operate at machine speed, process sensitive data continuously,	Add a dedicated subsection (4.4) emphasizing Zero Trust architecture as a cross-cutting principle applicable to all three Focus Areas and all six CSF Functions. The Profile should explain how Zero Trust principles (verify explicitly, use least-privilege access, assume breach) address AI-specific challenges and provide guidance on applying Zero Trust across AI asset management, access controls, threat detection, incident response, and recovery. Organizations should

				introduce non-deterministic behaviors, and expand attack surfaces through integrations and tool access. Zero Trust provides the architectural foundation for continuous verification, least-privilege access, inline inspection, and micro-segmentation—all essential for securing heterogeneous AI deployments.	understand that Zero Trust is not optional for AI security at scale—it's the architectural foundation that enables consistent controls across diverse AI deployments.