Before the
Center for AI Standards and Innovation (CAISI), National Institute of Standards and Technology (NIST), U.S. Department of Commerce
Washington, DC

In re Request for Information Regarding Security Considerations for Artificial Intelligence Agents	Docket No. NIST-2025-0035

COMMENTS OF COMPUTER & COMMUNICATIONS INDUSTRY ASSOCIATION (CCIA)

The Computer & Communications Industry (CCIA)[1] submits the following comments in response to selected topics from the January 8, 2026 Request for Information Regarding Security Considerations for Artificial Intelligence Agents issued by the Center for AI Standards and Innovation (CAISI), which is housed within the National Institute of Standards and Technology (NIST) at the Department of Commerce.[2]

CCIA members have developed and deployed many foundational innovations in fields such as artificial intelligence (AI) and machine learning, including AI agents. A regulatory environment that supports continued innovation is essential to maintaining U.S. leadership in AI development.

[1] CCIA is an international, not-for-profit trade association representing a broad cross-section of communications and technology firms. For more than fifty years, CCIA has promoted open markets, open systems, and open networks. CCIA members employ more than 1.6 million workers, invest more than $100 billion in research and development, and contribute trillions of dollars in productivity to the global economy. A list of CCIA members is available at https://www.ccianet.org/members.

[2] 91 Fed. Reg. 698 (Jan. 8, 2026).

I. Introduction

As CCIA has documented in related filings to the U.S. Government, including to NIST,[3] AI systems do not operate in a technological vacuum and are part of a wider digital ecosystem.[4] AI's unprecedented global growth and transformative potential enables significant benefits to businesses and consumers alike, such as reducing human error and enhancing efficiency and innovation. AI is not a single technology but rather a family of related but distinct technologies with differing use cases. Applying rules designed for one type of AI or one context to another situation can hinder the development of new forms of AI and create, rather than reduce, harm.[5]

Developers, businesses, and end users deploy AI models and agents in diverse contexts.[6] Models are the underlying intelligence that trains on inputs and produces outputs based on patterns learned from data, whereas an agent is a system that uses one or more models to take actions towards a goal. If a model is a brain without a body, the agent is a worker that uses that brain (model) and tools online to complete its task. Most notably, agentic AI can be used to rapidly detect high-severity security flaws like zero-day vulnerabilities. Consequently, NIST's ongoing work on agentic AI security and privacy must address risks and safeguards across the full lifecycle of AI systems. Given the wide range of use cases and deployment environments for AI agents, federal policy should reflect a deliberate, risk-based approach consistent with emerging NIST frameworks.[7] Security, privacy, and accountability mechanisms must be

[3] *See, e.g.*, CCIA Comments In Re: Request for Information Related to NIST's Assignments Under Sections 4.1, 4.5 and 11 of the Executive Order Concerning Artificial Intelligence (Sections 4.1, 4.5, and 11) (Feb. 2024), https://ccianet.org/wp-content/uploads/2024/02/ARTIFICIAL-INTELLIGENCE-CCIA-Comments-to-NIST-on-AI-RFI.pdf.
[4] *Id.*
[5] CCIA, *Understanding AI: A Guide To Sensible Governance* (June 26, 2023), at 2, https://ccianet.org/library/understanding-ai-guide-to-sensible-governance/.
[6] *See, e.g.*, Google Cloud, *What is agentic AI?*, https://cloud.google.com/discover/what-is-agentic-ai (last accessed Mar. 9, 2026).
[7] *See, e.g.*, *The NIST AI Risk Management Framework* (Jan. 26, 2023), https://www.nist.gov/itl/ai-risk-management-framework.

adaptable to different levels of context and downstream use. Overly prescriptive or blanket regulatory approaches risk undermining the development of beneficial AI applications, including AI agents, without meaningfully improving privacy and security outcomes. For these reasons, CCIA appreciates the opportunity to share the following comments on how the federal government can achieve safe and secure agentic AI innovation.[8]

II. Security Threats, Risks and Vulnerabilities

Unlike traditional software, agentic systems face risks from tool use and non-deterministic reasoning, which is when AI produces different outputs from the exact same input or prompt. Threats may include indirect prompt injection and multi-agent dynamics. Indirect prompt injection, for example, is a security vulnerability where an attacker embeds malicious instructions into external data, and AI then processes it. This allows malicious actors to influence an agent through external data. Agents may also act on structured external data sources including product listings, reviews, metadata or other machine-readable content, which can similarly be manipulated and requires careful validation and trust boundaries. A multi-agent system is an AI system where agents coordinate to achieve a collective goal. In such systems, safety issues can occur if agents misinterpret one another's outputs or lack a shared verification protocol.

Although agentic AI introduces unique security challenges, it can also offer new opportunities to address chronic cybersecurity vulnerabilities.[9] When deployed according to secure-by-design principles, agents can provide an automated, reasoning-based edge that traditional cybersecurity tools are unable to match. Shifting toward agentic systems allows

[8] Meta, *Agents Rule of Two: A Practical Approach to AI Agent Security* (Oct. 31, 2025), https://ai.meta.com/blog/practical-ai-agent-security/.

[9] Kent Walker, *A summer of security: empowering cyber defenders with AI*, Google (July 15, 2025), https://blog.google/innovation-and-ai/technology/safety-security/cybersecurity-updates-summer-2025/.

security operations to evolve from passive monitoring to proactive, autonomous defense. Security operations centers (SOCs) are centralized functions that combine people, processes, and technology to monitor, detect, and analyze cybersecurity threats. When these SOCs use agentic AI to automate the triage, investigation, and remediation of threats as part of human-in-the-loop (HITL) systems,[10] they can be made significantly more efficient while still benefiting from human decision-making and oversight. Additionally, agentic AI can proactively search and test for system vulnerabilities, demonstrating agents' ability to find and fix exploitable zero-day concerns in code.[11]

III. Security Practices for AI Agents

Effectively securing agentic systems requires both deterministic measures and reasoning-based defenses. Cars, for instance, require safety tools like seatbelts and brakes (deterministic measures), but also a driver or system that can sense a hazard and react (reasoning-based defenses). At the model level, this includes training the agent to distinguish between legitimate user instructions and malicious commands that could be embedded in data as well as fine-tuning for safety. Examples could include ensuring that agents learn to ignore instructions found with untrusted data through what is known as "automated red teaming," which generates thousands of synthetic prompt injection attacks for the agent to avoid.

At the system level, scaffold controls (which include the software and prompt frameworks in AI that guide its reasoning and manage its memory) can use engineering guardrails, restrictions, and protocols to establish verifiable communication.[12] Input and output guardrails are filters that can act as a gateway. These can scan a user's prompts for personally

[10] Petra Kelly-Voicu, *What is Human-in-the-loop (HITL) in AI-assisted decision-making?*, 1000minds (June 2023), https://www.1000minds.com/articles/human-in-the-loop.

[11] Anthropic, *Partnering with Mozilla to improve Firefox's security* (Mar. 6, 2026), https://www.anthropic.com/news/mozilla-firefox-security.

[12] *Announcing the Agent2Agent Protocol (A2A)*, Google for Developers (Apr. 9, 2025), https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability.

identifiable information (PII) or known attack patterns and ensure that an agent cannot accidentally output a malicious script.

Strict tool restrictions can also ensure that agents only have the minimum permissions necessary.[13] For example, a read-only agent that is tasked to read and understand reports should not be capable of writing or deleting said reports through its functionality. Some protocols standardize how agents talk to each other across different organizations, ensuring that every message between them includes information that verifies the agent's identity.[14] For example, a scheduling AI agent booking an appointment with a travel agency would use a token that confirms its identity and the permission to book said session. This prevents a rogue agent from pretending to be a trusted system component.

Human oversight must be the fail-safe for consequential or irreversible actions. HITL decision-making systems allow an AI agent to propose a plan while requiring human confirmation before executing high-impact tasks. In this context, the scope of oversight should remain focused on consequential actions, such as financial transfers or system changes, instead of routine, low-risk tasks like information retrieval or comparing documents. Requiring human approval for every agent action would undermine the efficiency gains that agentic systems are designed to provide. HITL systems are essential in high-risk situations that require high accuracy, ethical considerations, or contextual understanding.[15]

Instead, human oversight should follow a risk-based and layered model. This can include pre-action review for high-impact decisions, monitoring during operation, and the ability for

[13] Aaron Brown & Matt Saner, *The Agentic AI Security Scoping Matrix: A framework for securing autonomous AI systems*, Amazon Web Services (Nov. 21, 2025), https://aws.amazon.com/blogs/security/the-agentic-ai-security-scoping-matrix-a-framework-for-securing-autonomous-ai-systems/.

[14] *The agentic commerce platform: Shopify connects any merchant to every AI conversation*, Shopify (Jan. 11, 2026), https://www.shopify.com/news/ai-commerce-at-scale.

[15] Conor Baker, *What is a Human-in-the-Loop Approach to Leveraging AI?*, Conductor (Aug. 15, 2025), https://www.conductor.com/academy/human-in-the-loop/.

humans to intervene or halt actions when necessary. Other safeguards, such as scoped delegation of permissions, constrained execution environments, and comprehensive audit trails can further reinforce accountability while enabling safe autonomy. For higher-autonomy systems where real-time human confirmation may be impractical, humans must retain ultimate accountability through post-hoc review mechanisms and clear authority to stop or modify operations. In practice, autonomy should be bounded through explicit guardrails, pre-approved action catalogues, and reversible controls to mitigate inaccuracies or unintended outcomes. Voluntary industry-led standards such as the Coalition for Secure AI (CoSAI) Security Principles, can provide a proactive framework centered on human governance and transparent verifiability in high-risk and HITL systems.[16]

IV. Accessing AI Agent Security

One significant barrier to AI agent adoption is the lack of standardized testing. The federal government should support the creation and adoption of open-source evaluation tools to empower innovation and smaller actors. This includes evaluating trajectories, like the UK AI Security Institute's Inspect AI framework, which provides a template for rigorous, large-scale testing of both agent scaffolds and capabilities.[17]

Upstream documentation is also an essential component of testing. Providers and users downstream need clear documentation regarding the safety boundaries of any models they use. However, any mandatory public disclosures must carefully avoid creating a roadmap for attackers to exploit specific model concerns or weaknesses.

[16] *Announcing the CoSAI Principles for Secure-by-Design Agentic Systems*, Coalition for Secure AI (July 16, 2025), https://www.coalitionforsecureai.org/announcing-the-cosasi-principles-for-secure-by-design-agentic-systems/.
[17] *Inspect AI: Framework for Large Language Model Evaluations*, UK AI Security Inst. (2024), https://inspect.aisi.org.uk/.

Finally, current evaluation methods for agents are in their infancy. Research should focus on automated tests that are used to stress-test other agents. This will ensure that they remain resilient, private, and secure in multiple contexts.

V. Limiting, Modifying, and Monitoring Deployment Environments

Agentic AI adoption remains primarily hindered by rigid legacy infrastructure that lacks the application programming interfaces (APIs) required for an agent to plug into an ecosystem safely. Technical constraints should be enforced via hardware-supported trusted execution environments (TEEs), which can provide isolated environments where an agent's actions are simulated or restricted to a non-critical environment before being committed. This lets developers run multiple simulations and shows how the agent will react when faced with different requests.[18]

Managing risks with counterparties also requires standardized rules. A comprehensive national privacy framework would avoid multiple different state-level requirements, for example.[19] This would promote secure interactions across organizations and jurisdictions alike. Rollbacks and negations are also important when modifying and monitoring deployment environments. Rollbacks are the process of reverting an AI to a previous state after performance degradation or safety risks. Negations refer to how AI models interpret, process, or ignore negative words. Effective systems must comprehensively log agents' reasoning and actions to allow for rollback should testing show unwanted trajectories.

Recent headlines have focused on AI agents, mainly discussing whether current architectures and capabilities could allow AI to develop consciousness.[20] However, these

[18] Schneider, Moritz, et al. *Sok: Hardware-supported trusted execution environments*, Cornell U. ArXiv (May 25, 2022), https://arxiv.org/abs/2205.12742.

[19] *See, e.g.*, Jesse Lieberfeld, *CCIA Response to House Privacy Working Group RFI*, CCIA, at 3-5 (Apr. 7, 2025), https://ccianet.org/library/ccia-response-to-house-privacy-working-group-rfi/.

[20] *See, e.g.*, Avi Chai Outmezguine, *Moltbook Looks Like Consciousness. It Isn't.*, Built In (Feb. 17, 2026), https://builtin.com/articles/moltbook-ai-consciousness-debate.

examples overstate such agents' capabilities. Instead of having independent agency, these systems pattern-match and generate content from training data, which is how AI systems like LLMs operate.[21]

This example also points to real technical concerns. Security experts have flagged risks in how systems are built, including write-access vulnerabilities and exposed API keys, which are important in understanding how agents operate within their respective systems and the vulnerabilities that must be addressed as AI agents continue to evolve. Early research environments help us understand how agents behave and interact. These design decisions, including how agents are authenticated and how digital services handle user behaviors, can educate policymakers and industry alike, shifting stakeholders towards constructive policy and engineering discussions that are appropriately risk-based.

VI. Additional Considerations

The federal government can achieve safe and secure agentic AI innovation through the following recommendations. First, the government should define what safety outcomes are required rather than mandating how the industry should build and deploy AI agents. Similar to the aviation industry's shift towards performance-based rules, this approach allows nascent agentic AI technology to mature safely and dynamically.[22] Legal clarity regarding AI agents is also essential. Regulators must understand the current legal landscape, review policies that may be creating artificial innovation barriers, and, where necessary, implement risk-based and technologically-neutral policy changes.[23]

[21] Will Douglas Heaven, *Moltbook was peak AI theater*, MIT Tech. Rev. (Feb. 6, 2026), https://www.technologyreview.com/2026/02/06/1132448/moltbook-was-peak-ai-theater/.

[22] Federal Aviation Administration, *U.S. Transportation Secretary Sean P. Duffy Announces Improvements to Recreational Aviation Safety, Expansion of Light-Sport Sector* (July 22, 2025), https://www.faa.gov/newsroom/us-transportation-secretary-sean-p-duffy-announces-improvements-recreational-aviation.

[23] *Response to Call for Views: Agentic AI and Regulatory Challenges*, CCIA (Nov. 6, 2025), https://ccianet.org/library/response-to-call-for-views-agentic-ai-and-regulatory-challenges/.

Second, the U.S. government should use its procurement power to promote systems that are secure by design. Using many different vendors with histories of robust security can help mitigate systemic risk.

Third, a pragmatic and balanced regulatory approach would minimize the risk of stifling innovations in AI technology and processes. As new technological developments rapidly change agentic AI's capabilities and use cases, policymakers should allow flexibility for different safety approaches to flourish through a cyclical framework focused on continuous improvement. Technology companies seek to operate in clear, fair, and principle-based regulatory environments.[24] Before adopting any new regulation, it is essential to determine whether existing laws can address aspects of AI that are not unique to the technology. New laws should only seek to address identified, demonstrable policy gaps, rather than duplicate or conflict with pre-existing functional frameworks.[25] For instance, existing voluntary, risk-based frameworks and consumer protection laws already provide a baseline for cybersecurity and AI agent accountability as it relates to unfair and deceptive practices.[26] Enacting overlapping and conflicting laws could lead to stifling innovation.

[24] *See, e.g.*, Christian M. Dippon & Matthew D. Hoell, *A Quantitative Evaluation: The Economic Costs of Structural Separation, Line of Business Restrictions, and Common Carrier Regulation of Online Platforms and Marketplaces*, CCIA Research Center (Mar. 18, 2022), https://research.ccianet.org/reports/economic-costs-regulation-online-platforms-marketplaces/#main-content; Engine & the CCIA Research Center, *Tools to Compete Lower Costs, More Resources, and the Symbiosis of the Tech Ecosystem* (Jan. 25, 2023), https://research.ccianet.org/reports/tools-to-compete/#main-content.

[25] *See* CCIA, *Understanding AI: A Guide To Sensible Governance* (June 2023), https://ccianet.org/library/understanding-ai-guide-to-sensible-governance/.

[26] *See, e.g.*, *Cybersecurity Framework (CSF) 2.0* (Feb. 26, 2024), https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf *and* Andrea Arias, Federal Trade Commission, *The NIST Cybersecurity Framework and the FTC* (Aug. 31, 2016), https://www.ftc.gov/business-guidance/blog/2016/08/nist-cybersecurity-framework-ftc (demonstrating how the voluntary, risk-based CSF aligns with FTC consumer protection enforcement on data security); *see also* Federal Trade Commission, *FTC Sues to Stop Air AI from Using Deceptive Claims about Business Growth, Earnings Potential, and Refund Guarantees to Bilk Millions from Small Businesses* (Aug. 25, 2025), https://www.ftc.gov/news-events/news/press-releases/2025/08/ftc-sues-stop-air-ai-using-deceptive-claims-about-business-growth-earnings-potential-refund.

Multi-stakeholder engagement is essential. Sensible governance requires a partnership between those building these innovations and those protecting the public interest. Best practices in AI governance, and technology policy more broadly, are defined by technological capabilities as well as government preference, and should not be dictated by prescriptive regulations. Meaningful best practices require consultation with industry stakeholders who design, deploy, and operationalize these systems, along with stakeholders who understand how AI functions in real-world settings.[27] The safeguards, tools, and governance approaches that best mitigate risk are those that are principle-based and readily adaptable to the constantly-evolving nature of AI and like technologies.

Consumer-facing digital services have already built considerable consensus around mitigating risk to users and other parties. A recently published international standard, ISO/IEC 25389, reflects the evolving industry consensus on mitigating online risk.[28] This constitutes the first horizontal standard for safety on online consumer services. Policy work in this space should incorporate the existing consensus around best practice: safety by design, appropriate governance, application, improvement, and transparency.

VII. Conclusion

CCIA appreciates the opportunity to provide input on privacy and security considerations for AI agents. As the Administration considers the future of American competitiveness, it is essential that NIST leverages industry-developed standards when promoting best practices for agentic AI use. By acting as a facilitator rather than a rigid regulator, the federal government can cultivate a trusted AI ecosystem that empowers and drives global innovation by focusing on

[27] *See, e.g.*, Digit. Tr. & Safety P'ship, *Best Practices for AI and Automation in Trust and Safety* (Sept. 2024), https://dtspartnership.org/best-practices-for-ai-and-automation-in-trust-and-safety/.

[28] ISO/IEC 25389:2025, Information technology — The safe framework (Edition 1, June 2025), https://www.iso.org/standard/90106.html.

alignment and balance. Through the promulgation of voluntary, consensus-based standards, policymakers can ensure the U.S. maintains its global technological leadership while mitigating concerns through risk-based regulation.

Respectfully submitted,

Jesse Lieberfeld
Policy Counsel– Privacy, Security, & Emerging Technologies
Tricia McCleary
Information Policy Manager
Computer & Communications Industry Association
25 Massachusetts Ave NW, Suite 300C
Washington, DC 20001
tmccleary@ccianet.org

March 9, 2026