BBB National Programs

Docket no. NIST-2025-0035

Comments in Response to RFI to the Center for AI Standards and Innovation (CAISI) at NIST on Agentic AI

Introduction

BBB National Programs appreciates the opportunity to provide comments in response to NIST's Request for Information regarding security considerations for agentic AI systems, which are endowed with planning, tool use, memory, and/or the ability to autonomously execute tasks within defined boundaries, present meaningful opportunities for innovation and efficiency across sectors.[1] These systems introduce new governance, security, and accountability considerations that warrant thoughtful, risk-based evaluation.[2]

BBB National Programs is a nonprofit organization that has worked with industry leaders and government entities since 1971 to establish and enforce standards that guide best practices in privacy, child-directed marketing, technology, advertising, consumer warranty claims, dispute resolution, and, more recently, AI-enabled systems. As the home of independent industry self-regulation in the U.S., BBB National Programs operates third-party accountability mechanisms designed to promote innovation, competition, and marketplace trust.

BBB National Programs has convened stakeholders and developed applied standards for responsible AI deployment in the employment context, including *Principles for Trustworthy AI in Recruiting and Hiring* and associated independent certification protocols for AI-enabled hiring technologies.[3] These initiatives demonstrate how voluntary, risk-based frameworks, coupled with independent review, can translate high-level AI governance principles into operational safeguards within specific deployment contexts.[4] BBB National Programs also previously submitted comments to the National Telecommunications and Information Administration (NTIA), on the topic of AI accountability.[5]

Our recommendations to NIST on security considerations for agentic AI systems are four-fold:

[1] McKinsey & Company. (2025) *Deploying Agentic AI with Safety and Security: A Playbook for Technology Leaders.* https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/deploying-agentic-ai-with-safety-and-security-a-playbook-for-technology-leaders

[2] National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0). U.S. Department of Commerce. https://www.nist.gov/itl/ai-risk-management-framework

[3] BBB National Programs. (2023). Principles for trustworthy AI in recruiting and hiring. https://bbbprograms.org/programs/all-programs/ai/principles-for-trustworthy-ai-in-recruiting-and-hiring

[4] Brookings Institution. (2023). Governing the AI transition: Lessons from the 1996 Telecommunications Act. https://www.brookings.edu/articles/governing-the-ai-transition-lessons-from-the-1996-telecommunications-act

[5] Lexology. (2023). BBB National Programs Comments on AI Accountability, Response to RFI by NTIA. https://www.lexology.com/library/detail.aspx?g=4f6a62bd-c6a6-494b-8afc-34524ed60ccc#:~:text=Building%20AI%20Accountability,and%20accountability%20of%20AI%20actors.

(1) recognize and prioritize the role of independent third-party accountability mechanisms in verifying responsible use of agentic AI systems. Doing so can play a pivotal role in strengthening security, governance, and public trust in such systems;[6]
(2) encourage the development of risk-based, tiered oversight frameworks for agentic AI uses, which align safeguards and documentation requirements with the real-world impact of system deployment;[7, 8]
(3) support the translation of high-level AI governance principles into operational, measurable standards — including guidance on tool-use controls, memory governance, human-in-the-loop thresholds, and incident response — to ensure that agentic AI systems are secure by design;[9, 10] and
(4) promote adaptive and continuous monitoring structures, including post-deployment review and clearly delineated responsibility across the AI supply chain, so that accountability evolves alongside increasingly autonomous technologies.[11, 12]

BBB National Programs is well positioned to contribute to the development of third-party accountability mechanisms tailored to Agentic AI. Drawing on decades of experience administering certification programs, dispute resolution systems, and co-regulatory frameworks, BBB National Programs has developed practical expertise in translating emerging policy expectations into measurable, enforceable standards.

Building on this experience, BBB National Programs can work independently and/or with respected, standards-setting bodies such as NIST to:

- Convene multi-stakeholder working groups to develop uniform, applied operational requirements for agentic AI systems, including guidance and clear guardrails on tool-use governance, memory management, audit logging, access controls, role delineation across the AI supply chain, and human-in-the-loop escalation thresholds;[2, 9]
- Develop voluntary certification frameworks and/or written commitments by industry, aligned with the NIST AI Risk Management Framework, providing structured and measurable criteria for responsible deployment of agentic systems proportionate to risk;[2]

[6] Organisation for Economic Co-operation and Development. (2019). *OECD principles on artificial intelligence.* https://oecd.ai/en/ai-principles

[7] European Union. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* https://artificialintelligenceact.eu

[8] OECD AI Policy Observatory. (2023). *Risk-Based AI Governance Resources.* https://oecd.ai/en/

[9] OWASP GenAI Security Project. (2026). *Top Risks and Mitigations for Agentic AI.* https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/

[10] IBM. (2026). *A guide to agentic AI security.* https://www.ibm.com/think/insights/agentic-ai-security

[11] Association for Computing Machinery (ACM). (2025). *Systemic Risks Associated with Agentic AI: A Policy Brief.* https://www.acm.org/binaries/content/assets/public-policy/europe-tpc/systemic_risks_agentic_ai_policy-brief_final.pdf

[12] Alsharif Abuadbba, A., Sultan, N., Nepal, S., & Jha, S. (2026). *Human society-inspired approaches to agentic AI security: The 4C framework.* arXiv. https://arxiv.org/pdf/2602.01942

- Administer independent review and oversight processes, including periodic reassessment, structured documentation requirements, complaint intake mechanisms, and dispute resolution, where appropriate;[6]
- Promote transparency while safeguarding proprietary information through confidential review processes coupled with public trust marks or certifications that signal adherence to recognized standards;
- Serve as a regulatory auxiliary, complementing government oversight by incentivizing early adoption of best practices and providing structured compliance pathways in technically complex and rapidly evolving domains.

Independent accountability mechanisms are not a substitute for government enforcement. Rather, they function as complementary governance tools — strengthening compliance incentives, reducing uncertainty, and enhancing consumer and enterprise confidence in increasingly autonomous systems.[11]

As agentic AI systems become more integrated into commercial and societal infrastructure, recognition of independent, risk-based accountability mechanisms can help operationalize security principles in practice while supporting responsible innovation.

Recommendation 1: Support Independent and Third-Party Accountability Mechanisms

As agentic AI systems gain autonomy and integration capabilities, layered accountability becomes increasingly important. Internal governance processes are essential, but independent certifications, structured assessments, and third-party oversight can provide additional validation that recognized standards are being met.[6]

Leading policy frameworks consistently emphasize accountability and external validation as central to trustworthy AI governance. The OECD AI Principles underscore transparency and accountability as foundational pillars. The NIST AI Risk Management Framework highlights governance, documentation, measurement, and continuous improvement as core functions of responsible AI deployment.[2, 6] Policy research has similarly noted that independent audits and certification mechanisms can complement agency oversight in technically complex domains.

Agentic AI systems — which may execute code, access external tools, maintain persistent memory, and initiate multi-step actions — heighten the need for such layered review.[9, 10] Independent accountability mechanisms can:

- Verify that appropriate access controls, monitoring protocols, and escalation procedures are in place;
- Clarify lines of responsibility among developers, integrators, and deployers;[13, 14]
- Provide transparent trust signals to enterprise customers and consumers; and

[13] Global Partnership on Artificial Intelligence (GPAI). (2023). *Responsible AI governance and policy recommendations.* https://www.oecd.org/en/about/programmes/global-partnership-on-artificial-intelligence.html
[14] European Parliamentary Research Service. (2025). *Artificial Intelligence and Civil Liability.* https://www.europarl.europa.eu/RegData/etudes/STUD/2025/776426/IUST_STU(2025)776426_EN.pdf

- Encourage early adoption of best practices in advance of formal regulatory requirements.

BBB National Programs has a distinguished track record in the independent accountability space. BBB National Programs' Children's Advertising Review Unit (CARU) operates the first Federal Trade Commission-approved Children's Online Privacy Protection Act (COPPA) Safe Harbor program. The organization also serves as a watchdog in other key domains, including digital advertising privacy.

As an independent third party, BBB National Programs also provides certifications, with accompanying trust marks, and consumer-to-business dispute resolution services in furtherance of international data transfer mechanisms, including the Global Cross Border Privacy Rules (CBPR) program and the EU-U.S. Data Privacy Framework. Finally, BBB National Programs is home to the trusted business-to-business dispute resolution program for advertising claims, the National Advertising Division (NAD) and the National Advertising Review Board. Though these existing accountability mechanisms differ widely from one another in subject matter and structure, they all share basic characteristics and require industry commitments to a shared goal, which together work to foster trust, transparency, and accountability in the marketplace.

Recommendation 2: Adopt Risk-Based, Tiered Oversight Frameworks for Agentic AI

Not all agentic AI systems pose equal risk. Systems that operate in high-impact domains — such as employment, healthcare, finance, or public services — warrant greater scrutiny than lower-risk automation tools.

Risk-based (and risk-tiered) governance is well established in global AI policy discussions. For example, Singapore's Model AI Governance Framework for Agentic AI (released in January 2026) operates on a risk-based approach that assesses and mitigates risks, particularly focusing on high-risk, autonomous, and tool-enabled agents. The framework emphasizes pre-deployment testing and scaling oversight based on the agent's autonomy and potential impact. The OECD AI Policy Observatory highlights risk-proportionate governance as a best practice.[8]

Academic analysis further suggests that agentic systems capable of autonomous action may warrant heightened oversight due to their integration capabilities and unpredictability.[11, 12]

For Agentic AI, a tiered oversight model could include:

- Enhanced documentation and review requirements for high-impact systems;
- Periodic reassessment for adaptive or learning-enabled agents;
- Escalation requirements where systems have authority to execute tools or trigger consequential actions.

Such an approach aligns safeguards with real-world impact while preserving space for innovation in lower-risk applications.

Recommendation 3: Move from High-Level Principles to Operationalized Standards

High-level AI governance principles provide important direction, but without operational clarity they risk inconsistency in application.[15, 16] In the agentic AI context, measurable standards are particularly important given the complexity of tool integration and autonomous behavior.[2, 9]

Operational standards may include:

- Defined tool access controls and authorization documentation;
- Clearly articulated human-in-the-loop thresholds;
- Logging and audit trails for autonomous actions;
- Incident response protocols tailored to agentic behavior;
- Memory governance and retention controls.[10]

Operationalization increases consistency, strengthens auditability, and reduces ambiguity across organizations deploying these systems.

Recommendation 4: Clarify Responsibility Across the Agentic AI Ecosystem

Agentic AI ecosystems often involve foundation model developers, fine-tuners, system integrators, enterprise deployers, and downstream users. As autonomy increases, responsibility can become diffuse.

Policy bodies such as the Global Partnership on AI and the European Parliamentary Research Service have emphasized the importance of clearly defined roles across the AI lifecycle.[13, 14] Clear allocation of responsibility reduces friction between vendors and deployers, establishes expectations for monitoring and remediation, and creates more predictable compliance pathways.[11]
Independent certification or verification frameworks may further assist in clarifying these responsibilities in practice.

Finally, because agentic AI systems may adapt over time, governance mechanisms should not be static. Post-deployment monitoring, periodic reassessment, and structured update review processes are essential to ensure that safeguards remain effective as capabilities evolve.[11, 12]

[15] Berkman Klein Center for Internet & Society. (2023). Translating AI Principles into Practice. Harvard University. https://cyber.harvard.edu/projects/artificial-intelligence-initiative
[16] AI Now Institute, AI Governance Reports.