March 9, 2026

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

Re: Request for Information Regarding Security Considerations for Artificial Intelligence Agents

The Cybersecurity Coalition (the Coalition) submits these comments in response to NIST's request for information on Security Considerations for Artificial Intelligence Agents.

Introduction

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.

Agentic AI marks a significant leap forward in AI – by going further to integrating multiple generative AI systems to autonomously handle complex tasks. In cybersecurity this is especially useful given the growing threats organizations are facing on a daily basis. As security teams race to outpace AI-wielding threat actors, Advanced-Agentic AI driven cybersecurity technologies can save users hours of manual work by completing tasks such as initial detection triage or prioritization of alerts on their behalf. These actions, proactively executed on behalf of the user, must be bounded by predefined goals and real-time conditions. At the same time, the increasing autonomy and operational authority of AI agents underscores the critical need to secure their development, deployment, and ongoing operation—consistent with the Secure Focus Area outlined in the Cyber AI Profile. Agentic AI can, for example, allow SOC teams to focus on the most critical threats and perform more advanced tasks, while agentic AI handles less complex issues. Different types of agents - exposure prioritization, malware analysis, threat hunting, data transformation, workflow generation, and more - are already changing the security landscape. And agentic tools are becoming widely used to assist in coding, workplace productivity, and internal operations. The Cybersecurity Coalition is very supportive of NIST's

work to standardize and provide input on aspects to consider in securing AI agents and agentic systems.

1. Framing Agentic AI: Evolution, Not a New Paradigm

Agentic AI systems represent an evolution in the complexity, autonomy, and scale of artificial intelligence capabilities. However, they should not be treated as categorically distinct from existing cybersecurity frameworks. Current NIST publications—including the Cybersecurity Framework 2.0; AI 100-2e2025; the AI Risk Management Framework; the Risk Management Framework Generative AI Profile; AI 800-1; SP 800-218A; and SP 800-53—provide a strong and adaptable baseline for governing agentic AI systems. These comments emphasize that AI, including agentic AI, is software and should be secured accordingly. At the same time, its applications require nuanced, context-specific risk management, with security incorporated by design from the outset. Any new guidance or frameworks addressing agentic AI should align with, and build upon, ongoing NIST initiatives to avoid duplication, inconsistency, or confusion. In particular, the Cybersecurity Framework Profile for Artificial Intelligence (NIST IR 8596 IRPD, the "Cyber AI Profile") and related efforts should explicitly address agentic AI where appropriate. The Cybersecurity Coalition recently submitted comments on the Cyber AI Profile, which are incorporated into these comments and are available here.[1] In addition, CAISI should seek alignment with ongoing AI agentic security work at the National Cybersecurity Center of Excellence, including the pending project on Identity and Authority of Software Agents.[2]

Organizations across sectors have invested significantly in Zero Trust architectures over the past several years, out of recognition that "once attackers breach the perimeter, further lateral movement is unhindered."[3] Agentic AI systems will face the same problem. Treating AI agents as a new category of principal, subject to the same continuous verification and least-privilege enforcement already applied to users and workloads, will reduce implementation complexity and accelerate secure adoption. For federal agencies specifically, this approach aligns with existing Zero Trust mandates under OMB M-22-09 and CISA's Zero Trust Maturity Model, allowing agencies to extend current investments rather than deploy separate AI-specific infrastructure.

These comments provide recommendations across several key areas relevant to agentic AI security, including identity, discovery, authentication, and authorization challenges; boundary protections; contextual risk; deployment diversity; evolving threat landscapes; the need for continuous governance adaptation; and the importance of supporting open, interoperable standards when necessary. The Cybersecurity Coalition commends CAISI and NIST for their continued leadership in addressing the security implications of rapidly evolving AI technologies.

2. Where Agentic Systems Break Traditional Assumptions

---

[1] https://cdn.prod.website-files.com/660ab0cd271a25abeb800460/69811473101b5ab7e16e82aa_Cybersecurity%20Coalition%20Comments%20-%20NIST%20IR%208596%2C%20Cybersecurity%20Framework%20Profile%20for%20AI.pdf

[2] https://www.nccoe.nist.gov/news-insights/new-concept-paper-identity-and-authority-software-agents

[3] https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf

As this Request for Information focuses on novel risks that arise from the use of machine learning models embedded within AI agent systems, it is important to outline where agentic systems break traditional assumptions of machine learning and AI security.

Integration

Certain risks, including explainability, are exacerbated by the environment in which they are taking place. Agentic AI is, and will be, used in countless IT and OT environments and each environment should take stock of the various risks associated with the inclusion of agents. While no two environments are the same, questions about agents integration with an environment span all types of environments. Agents span tools, APIs, vendors, users, and more.

For example, in an enterprise IT environment, an agent may integrate with email systems, document repositories, and internal knowledge bases to handle support requests or draft responses. If the agent autonomously escalates a ticket, modifies a document, or sends an external communication, explainability concerns extend beyond why the model generated a particular response to why it selected certain tools, accessed specific data sources, or triggered a particular workflow. In OT environments such as manufacturing facilities or energy infrastructure, an agent may interface with industrial control systems (ICS), sensor networks, or predictive maintenance tools. If the agent autonomously adjusts system parameters based on environmental inputs, explainability becomes a matter of operational safety. Stakeholders must be able to trace which sensor readings were relied upon, what intermediate assessments were made, and why a particular control action was taken.

Human Oversight: Autonomy and Speed

Agentic systems complicate traditional assumptions around explainability, oversight, and control. Unlike conventional software, agentic AI often has full or partial autonomy to conduct an action on behalf of a person. They can plan, execute actions, call external tools, and more with limited or no real-time human intervention.[4] Controls must account for diverse architectures and capabilities. No matter the environment, decision making without a human-in-the-loop presents various challenges with accountability and oversight. Agents should default with as little autonomy as possible, and only be granted more upon rigorous testing.

The degree of autonomy granted to an agent should be calculated based on the sensitivity of the task and the maturity of the environment. Agents should default to the least autonomy necessary. Safeguards such as scoped permissions and real-time auditing from humans can help ensure that autonomy does not outweigh human oversight.

Explainability

AI systems are often characterized as “black box” models because their internal reasoning processes are opaque or difficult to interpret.[5] Explainable AI efforts have largely emerged in

[4] https://www.ibm.com/think/topics/ai-agents
[5] https://ssir.org/articles/entry/do_ai_systems_need_to_be_explainable

response to public concern about fairness, accountability, and trust in scenarios where AI systems are used to inform high-stakes decisions. In those contexts, explainability is often framed as a transparency issue between human developers, deployers, and affected users.[6, 7] Agentic AI systems introduce a new element to explainability concerns. Unlike traditional models, agentic systems are designed to plan, reason across steps, interact with external environments, and take autonomous actions. As a result, the relevant explainability challenge extends beyond understanding a single model output to understanding a chain of decisions, intermediate steps, external tool use, and environmental interactions.[8] These differences in action traceability present a serious risk, especially as agents increasingly have access to more environments.

3. Context, Segmentation, and Environmental Risk

One of the defining features of the agentic AI threat landscape is the operational context in which agents are deployed. Unlike traditional software components, agentic systems may plan, execute, and chain actions across multiple systems, tools, and trust domains. As a result, their risk profile is highly dependent on their environment.

NIST should explicitly emphasize the application of zero-trust principles to agentic AI environments. Access to data, tools, APIs, and infrastructure should be continuously verified, narrowly scoped, and segmented according to defined roles and operational necessity. Agentic systems should not inherit broad network permissions by default, particularly where they are capable of autonomous chaining across services. Separation of privilege and duty are basic IT security principles that should apply to agentic AI systems. Authorization and access control decisions should be enforced externally to the model itself. Because models remain susceptible to prompt injection and goal manipulation, they should not be relied upon to self-enforce permission boundaries. External enforcement layers provide a necessary check on model behavior, consistent with defense-in-depth principles where no single control point is assumed fully reliable.

Visibility into deployment context is equally critical. Organizations must understand:

- Where agentic systems are deployed within their environment;
- What data, systems, and tools those agents can access;
- How agents may chain actions across systems or domains;
- What identities or credentials agents assume when operating.

Understanding the presence of agentic systems is insufficient without mapping their permissions, connections, and potential movement pathways. Any guidance should therefore emphasize maintaining meaningful human visibility and oversight of agent activity. This includes logging, monitoring, and the ability to review and intervene in agent-initiated actions where

[6] https://www.brookings.edu/articles/the-tensions-between-explainable-ai-and-good-public-policy/
[7] https://www.sciencedirect.com/science/article/pii/S0740624X21001027
[8] https://www.sciencedirect.com/science/article/pii/S1566253525006712

appropriate. Risk management should not treat agentic systems as isolated applications, but as actors embedded within a broader enterprise ecosystem.

At the same time, risk assessment must be pragmatic. Organizations should prioritize risk based on sensitivity, criticality, and potential impact rather than attempting to exhaustively model every possible interaction within complex environments. Overly burdensome mapping requirements may dilute focus and divert resources from high-impact risk mitigation.

Accordingly, NIST should recommend a risk-based prioritization approach that:

- Identifies high-impact environments (e.g., critical infrastructure, sensitive data locations, production systems);
- Evaluates agent access and action scope within those environments;
- Aligns mitigation measures with the organization's defined risk tolerance.

Agentic AI governance should therefore integrate environmental awareness, zero-trust architecture, and risk prioritization to ensure that autonomy does not translate into uncontrolled expansion of access or authority.

Network segmentation and isolation remain among the most effective mechanisms for preventing systems from acting outside their intended scope. That being said, implementers are currently experiencing difficulty in isolating AI agents, given the increasing complexity of the systems in which they operate and the models themselves. Existing segmentation, least-privilege, and isolation standards for software and AI systems should serve as the baseline for agentic AI deployments. Such standards include NIST SP 800-52, NIST CSWP 28, the NIST Cybersecurity Framework (CSF) 2.0, NIST SP 800-215, IEC 62443, among others. While these controls may not fully address the cross-domain behavior of agentic systems, they provide a good foundation.

4. Identity, Authentication, and Authorization Challenges

Agentic AI introduces new dynamics around identity, even though the fundamental principles of authentication and authorization remain the same. Unlike traditional software or static AI deployments, agentic systems may act autonomously across multiple environments, interact with diverse identity providers, invoke third-party tools, and make context-dependent decisions about access and execution. This creates new complexity in how identity is established, verified, constrained, and audited. The challenge is not redefining identity principles, but ensuring that agent identity is enforceable, observable, and enacted in a risk-informed manner.

A threshold question is whether agents should be treated merely as tools, as service accounts, or as a distinct category of digital actor. One useful construct may be to treat agents as Non-Person Entities (NPEs): entities that can initiate and sequence actions without being human principals, yet whose actions carry operational and security consequences comparable to human activity. Explicit recognition of agents as NPEs would clarify governance expectations and support more tailored control requirements. To implement NPE governance effectively,

organizations should adopt machine identity security practices that treat agent credentials with the same rigor as privileged human accounts.

This categorization challenge is particularly important because agentic systems often operate with dynamic, rather than static, permission sets. Unlike traditional user accounts or service identities that maintain relatively fixed privileges, agents may require context-dependent delegation that expands or contracts based on task, environment, or user instruction.

Risk-Based Differentiation of Agents

Not all agents present the same level of risk, and identity and authorization controls should reflect this difference. Low-privilege, narrow-scope agents, such as those limited to read-only data retrieval or reversible actions, may operate with constrained permissions and automated execution under defined guardrails. In contrast, high-impact agents capable of executing financial transactions, modifying sensitive data, or altering infrastructure configurations warrant heightened controls.

For higher-risk agents, safeguards may include stricter permission boundaries, transaction limits, mandatory human-in-the-loop review, or multi-party approval mechanisms. A tiered assurance model can help organizations align identity controls with risk exposure without imposing unnecessary friction on lower-risk use cases.

Identity Transparency and Auditability

Clear attribution and auditability are critical in agentic environments. Organizations must be able to determine whether an action was taken by a human user, an agent acting on behalf of a user, or an autonomous process operating under predefined authority. This requires unambiguous identity representation, separation between user intent and agent-executed actions, and durable logging of authentication, authorization decisions, and execution context.

Audit records should capture not only the action performed, but also the authority under which it was performed, including delegated permissions and any human approvals involved. Without transparent attribution and traceability, incident response, accountability, and governance become significantly more difficult in agent-driven systems.

Where agents are treated as NPEs with dynamic permissions, audit mechanisms should also record changes in delegated scope over time, including when privileges were expanded, constrained, or revoked. Without visibility into how agent authority evolves, organizations may be unable to assess whether an action reflected legitimate delegation or unintended privilege drift.

Agent Discovery as a Security Control

In addition to risk-based differentiation and identity transparency and auditability, NIST should treat agent discovery itself as a core security control. At present, organizations lack a consistent mechanism to locate, catalog, and validate the AI agents operating within and across their

environments. Agents may be scattered across multiple platforms, registries, internal development pipelines, and open-source ecosystems, with metadata published in agent cards, proprietary directories, or in some cases not published at all. Absent reliable answers to basic questions—what agents exist, which external services they invoke, and what permissions they hold—downstream security measures are built on partial or unreliable information. This lack of visibility also raises sovereignty concerns: organizations need the ability to independently determine and technically enforce which agents are permitted or denied, without being constrained by any single external platform or intermediary.

Guidance should encourage discovery as a security control, and recommend mechanisms that are open, interoperable, and rooted in existing Internet infrastructure, so that organizations retain this autonomy across diverse platforms and deployment models.

Agents as a New Class of Identity

Agentic systems should be incorporated into existing identity and access management frameworks as a distinct principal type. Just as organizations have extended identity governance from human users to service accounts, APIs, and workloads, AI agents represent the next category requiring identity lifecycle management, credential issuance, and access policy enforcement. This approach allows organizations to adapt proven governance processes rather than building agent-specific controls from scratch. Policy enforcement should be identity-centric rather than perimeter-centric, with access decisions that are context-aware and continuously evaluated.

Multi-Agent Communication Security

As multi-agent deployments increase, agent-to-agent communication will emerge as a distinct attack surface requiring dedicated attention. Organizations should ensure that interactions between agents, particularly those crossing trust boundaries or organizational perimeters, are subject to authentication of agent identity, authorization for the specific interaction, content inspection where feasible, and comprehensive logging. The same Zero Trust principles that organizations apply to user-to-application access should extend to agent-to-agent scenarios: never trust, always verify, and grant only the minimum access necessary for the specific interaction. Standards bodies should prioritize the development of secure communication protocols and interoperable identity frameworks for multi-agent scenarios.

5. Deployment Diversity: Cloud, Hybrid, On-Prem, and OT/ICS

Agentic AI systems will not be deployed within a single type of environment. They are emerging across cloud-native platforms, hybrid enterprise structures, on-premises systems, and across various OT/ICS contexts. Organizations must therefore consider how to prevent lateral movement, unauthorized access, or data exfiltration when agents operate across systems with differing security models. Boundary protections must be intentionally designed to account for agents that span multiple environments rather than reside within a single perimeter.

Each model has its own technical constraints, governance structures, and risk profiles. For instance, cloud-native deployments may face challenges around data governance, scalability, and access control.[9] In OT/ICS contexts such as power grids, water treatment plants, and manufacturing systems, AI integration must be balanced with high standards for operational safety and security.

Because these differing constraints affect both risk and governance requirements, implementation guidance or governance frameworks should adopt a broad lens that accounts for this diversity. It should also be applicable to any potential future applications, without needing to be majorly revised. Any governance framework should provide specific examples within different systems, as well as have broad, overarching principles that apply to agentic systems no matter where they are implemented.

6. Threat Landscape: Known Risks vs. Emerging Unknowns

The threat landscape for agentic AI systems includes the established risks associated with AI systems generally, including direct and indirect prompt injection, goal manipulation, training data poisoning, token manipulation, model extraction, excessive privilege, and related attack vectors.[10] Because agentic systems build on existing AI architectures and are deployed as software within broader digital environments, they inherit many of the traditional software and AI security risks.

As agentic AI becomes more widely deployed, additional threats may emerge that are specific to autonomous behavior, multi-step planning, tool use, and system integration. However, at present, there is limited real-world threat intelligence that differentiates agentic systems from AI systems more broadly. This gap should not delay preparation. Organizations should anticipate that agent-specific attack patterns will emerge as deployment scales, and security architectures should be designed with sufficient flexibility to incorporate new threat intelligence as it develops. Proactive measures, including agent-focused red teaming and participation in information sharing communities, can help close this intelligence gap. Industry red team testing has demonstrated that enterprise AI systems are highly susceptible to compromise within minutes of adversarial engagement, with critical vulnerabilities found in 100% of systems tested in a recently conducted threat analysis.[11]

Accordingly, recommendations addressing the threat landscape should avoid being overly prescriptive at this stage and encourage the use of defensive Agentic AI cybersecurity operations to stay ahead of the threat. Policymaking should remain adaptable and responsive to evolving evidence, supported by continued threat research and iterative updates to risk management practices as new information becomes available.

7. Attack Surface Management for Agentic Systems

[9] https://www.informationweek.com/it-infrastructure/building-secure-cloud-infrastructure-for-agentic-ai
[10] https://www.cisco.com/site/us/en/learn/topics/artificial-intelligence/ai-security-safety-framework.html
[11]https://www.zscaler.com/press/zscaler-2026-ai-threat-report-91-year-over-year-surge-ai-activity-creates-growing-oversight

Agentic AI systems expand the traditional understanding of software attack surface due to the breadth of systems, tools, and environments with which they might interact. While increased connectivity alone enlarges the potential attack surface, the more significant shift lies in how agentic systems integrate, chain actions, and rely on dynamic dependencies across domains.

Unlike conventional software components that operate within relatively fixed boundaries, agentic systems may conduct multi-step workflows across cloud services, enterprise platforms, and physical or OT environments. These integrations create additional technical points that may serve as vectors for misuse, compromise, privilege escalation, or cascading system effects.

Agentic systems also sometimes rely on layered and opaque dependencies, including dynamically invoked tools, third-party services, external data sources, and persistent memory components.[12] These dependencies may not be fully visible in static architecture diagrams and can complicate both risk assessment and incident response. As a result, agentic systems should be treated not as isolated applications, but as high-connectivity components embedded within broader enterprise ecosystems.

Guidance should therefore emphasize comprehensive organizational visibility across the full agentic stack, including agents, underlying models, connected tools, identity layers, and deployment environments. Effective governance in this context requires:

- Asset discovery to identify where agentic systems are deployed and how they are configured;
- Dependency and integration mapping across APIs, services, and data sources;
- Monitoring of agent behavior and action patterns, particularly where chaining or cross-system orchestration occurs;
- Continuous reassessment of capabilities as models are updated, retrained, or integrated with new tools.

Because frontier AI systems may evolve through version updates, configuration changes, or expanded tool access, traditional security assumptions are insufficient. Security controls and governance processes must account for the dynamic nature of agentic systems. Continuous visibility and adaptive oversight are foundational to maintaining secure deployments as agentic AI evolves.

8. Inline Enforcement for Agentic Systems

The operational tempo of agentic systems, which may chain multiple actions in fractions of seconds, requires enforcement mechanisms capable of evaluating and constraining agent actions in real time. Controls that operate out-of-band or through periodic scanning cannot match agents that execute workflows at machine speed. NIST guidance should emphasize that enforcement mechanisms must operate at or near the time of execution, with the ability to inspect, constrain, and log agent actions as they occur. Organizations that implement inline

[12]https://kanerika.com/blogs/agentic-ai-in-supply-chain/#:~:text=Agentic%20AI%20systems%20go%20beyond%20static%20predictions,social%20trends%20to%20adjust%20forecasts%20in%20real%2Dtime.

inspection at the access layer gain native visibility into agent behavior as a byproduct of enforcement, enabling real-time correlation across users, agents, workloads, and applications. This approach addresses two fundamental challenges: agent actions occurring faster than human review cycles, and telemetry fragmentation when agents operate across multiple systems.

9. Mitigation First: Managing Risk Amid Immaturity

Given the relative immaturity of agentic AI security practices, guidance should prioritize practical risk mitigation over creating new, standalone standards regimes, while leveraging and extending existing open, interoperable frameworks where appropriate. Mitigation within an agentic system deployment should be both realistic and responsible. Organizations should focus on established, high-value security controls that reduce impact even where threats are not fully understood. These include isolation of agentic components, limiting blast radius through segmentation and scoped deployment, enforcing least-privilege access to tools and data, and maintaining human-in-the-loop oversight for high-risk or irreversible actions.

To ensure security is at the highest possible level and to mitigate potential future risks, agentic AI should be developed and deployed in accordance with secure-by-design principles. Protections should be built in, not added on top. These principles also apply to the platforms, models, and tooling ecosystems that support the agentic AI ecosystem.

Importantly, protection must operate across the full lifecycle and operational chain of agentic systems, including discovery of tools and resources, identification and authentication processes, execution of tasks, and agent-to-agent communication. NIST guidance would be most effective if it clearly articulates system boundaries and guardrails, explicitly acknowledges the evolving state of best practices, and frames its recommendations as iterative and adaptable. Such an approach sets realistic expectations, strengthens security posture, and supports continued innovation rather than constraining it prematurely.

The architecture of agent discovery introduces important risk considerations for AI agent systems. Today, many discovery functions are embedded in closed, proprietary, or implementation-specific registries tied to individual platforms. This can create concentration risk: when a small number of providers mediate which agents can be seen and reached, enterprises may lack a unified, cross-provider inventory and instead rely on fragmented, platform-specific views.

In the absence of standardized, open discovery mechanisms, platforms are incentivized to operate as silos, which can hinder interoperability, complicate governance, and limit visibility across the broader ecosystem. A more resilient approach would encourage discovery mechanisms built on open, vendor-neutral standards and widely deployed internet infrastructure, enabling interoperable, decentralized, and verifiable agent discovery. Leveraging established internet infrastructure and security practices can help align agent discovery with existing operational and trust frameworks while avoiding unnecessary centralization.

NIST has the opportunity to encourage the development and adoption of open, interoperable, vendor-neutral standards for AI agent discovery and trust verification through established

multi-stakeholder standards bodies. Forthcoming NCCoE projects on software and AI agent identity and authorization could be used as testbeds to evaluate and demonstrate interoperable, standards-based discovery mechanisms in realistic enterprise environments.

10. Risk-Based Controls Aligned to Model Capability

As with other NIST guidance, any guidance or framework around agentic systems should be technology-agnostic. In a rapidly evolving field such as AI, it is crucial to provide guidance that can be applied to any application of agentic systems, including future ones. Model-specific guidance risks becoming outdated and might favor certain technologies or companies. Guidance should instead focus on core principles and risk management processes that can be applied across applications or deployments.

Recommendations should be technology-agnostic and scale based on model capabilities, decision impact, autonomy level, access scope, and any other differences. NIST guidance should promote consistency and long-term relevance while securing agentic systems.

11. Evaluation and Assessment of Agentic AI Security

Evaluation and assessment of agentic AI security warrant explicit consideration. At a baseline, organizations deploying agentic systems should conduct internal assessments addressing threats, system vulnerabilities, and contextual risk, including how autonomy, tool use, and operational environment shape potential impact. Security assessments for agentic systems should incorporate adversarial testing that accounts for agent-specific attack vectors, including goal manipulation, multi-step exploitation of autonomous decision-making, and attacks that leverage tool access or cross-system chaining. Industry experience demonstrates that enterprise AI systems frequently exhibit critical vulnerabilities when subjected to adversarial testing, with failures often surfacing within minutes of engagement. Red-teaming of goal-seeking tasks is particularly important, including testing under offline or limited-connectivity scenarios to verify fail-safe behavior when policies or cloud services are unreachable. These assessments should also account for third-party and upstream dependencies, including agent platforms, underlying models, and the broader tooling ecosystems on which agentic systems rely.

Particular attention should be paid to AI-specific supply chain risks that differ from traditional software dependencies. These include remote tooling dependencies, such as MCP servers or external backends, where server-side changes can alter agent behavior without visible changes in the agent's code, as well as model updates or version switches that may affect safety boundaries without code modifications.

At the same time, it remains an open question whether security assessments for agentic AI are materially different from those applied to other AI systems, or whether they are primarily more complex, dynamic, and continuous in nature. This area would benefit from further research and practical experience before formal standardization or prescriptive requirements are established.

12. Continuous Evolution and Iteration

The evolution of agentic AI systems is accelerating rapidly.[13] Both deployment applications and the underlying models themselves are advancing at a pace with which governance efforts will struggle to keep up. In this environment, standards and guidance must be developed early, iterated frequently, and structured to adapt alongside technological change. Static or infrequently updated frameworks will struggle to keep pace with increasingly autonomous and interconnected systems.

NIST plays a critical role as a global leader in AI security standards. By providing principled, technology-neutral, and risk-based guidance for agentic AI systems, NIST can help ensure that agentic AI innovation continues in line with security, accountability, and public trust.

[13] https://www.computer.org/publications/tech-news/trends/agentic-ai