Understanding Agentic AI:

ITI's Policy Guide

November 2025

ITI
Promoting Innovation Worldwide

The latest advancement of AI technology is here: agentic AI. Agentic AI is an application of existing generative AI technology and has the potential to bring benefits to enterprises and consumers more broadly. This paper is intended to unpack agentic AI, including what such systems are comprised of, who relevant actors in the ecosystem are, and offer initial considerations for policymakers as they think through how to reap benefits and manage risks associated with agentic AI.

What is agentic AI?

While there are many different definitions that have been proposed for what constitutes an agentic AI system,[1] for the purposes of this paper, we define agentic AI as a system with multi-step planning and reasoning capabilities that can autonomously query and select tools within boundaries set by the user to execute on a user-defined task or achieve an outcome.

Autonomy in this context does not mean that all agentic AI systems operate with the same level of independence. Autonomy exists on a spectrum, and in many instances, users can set the level of autonomy of the agentic AI system within a given tool or product via design patterns. The risk level of a particular agentic AI system varies depending on the level of autonomy, the environment in which an agentic AI system is deployed, the complexity of the goal or outcome the agentic AI system is intending to achieve, and its ability to impact the environment in which it operates.[2]

This is especially important in a policy context, where the risk level matters in determining proportionate obligations. We discuss this further in our Policy Considerations section. Importantly, even with increased autonomy, there remains a significant element of human involvement, particularly in assigning tasks to the agentic AI system, defining the parameters in which the system operates, setting the guardrails, and overseeing the execution of those tasks.

Agentic AI systems are typically built upon general-purpose large language models (LLMs), which provide essential reasoning capabilities. While many agentic AI systems rely upon general-purpose LLMs given their powerful reasoning capabilities, not all of them follow this approach. Some agentic AI systems are constructed using LLMs that have been fine-tuned on domain-specific data for a certain task. For example:

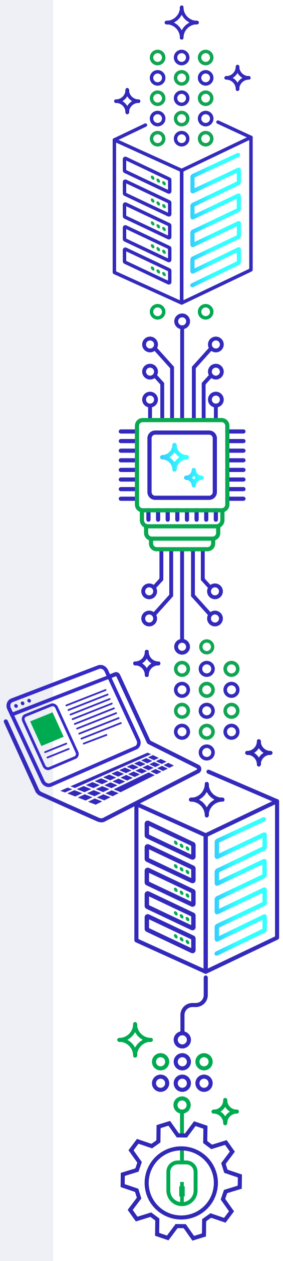
- In the biomedical space, models like BioGPT, PubMedBERT, and Med-PALM have been trained on a specific subset of data (e.g. medical databases) to support a targeted use case, like autonomously cross-referencing patient data with the latest clinical studies to assist a doctor in identifying potential diagnoses. An agentic AI system could be built on top of one of these specialized models.

- When it comes to serving small business' financial needs, an agentic AI system could be built upon a model specifically trained on financial regulations and small business financial data. This specialized model, akin to a "FinServ-LLM," could autonomously analyze a small business' financial statements, cross-reference them with public databases for relevant economic indicators and business registries to reduce manual data entry, and then generate tailored recommendations for identifying eligible government grants and loans. The agentic AI system would then take the initiative to draft necessary application forms or compliance reports. This would shift the burden of financial administration away from the entrepreneur, allowing them to focus on growth while the AI secures opportunities.

Besides LLMs, some agentic AI systems can operate on small language models (SLMs) that are optimized for targeted, domain-specific tasks.[3] While LLMs are best suited to handle broad reasoning and complex agent orchestration, SLMs excel at the focused, repetitive tasks performed by specialized AI agents.

What goes into an agentic AI system? Who is directly involved in the agentic AI value chain?

An agentic AI system typically comprises several components:



Data: The effectiveness and accuracy of AI systems are heavily dependent on the quality, representativeness, and reliability of the data they are trained on. Through data, AI systems learn language, recognize patterns, and make predictions. Data sources vary and can include the open web, enterprise and proprietary databases, scientific datasets, public records, and artificially generated synthetic data.

LLM or SLM: LLMs or SLMs are an agentic AI system's decision-making engine. They provide reasoning capabilities for the agentic AI system and are the basis upon which the agentic AI system analyzes the situation presented to it and plans the steps it should take. They can also delegate tasks to be executed on by the agentic AI system's tools.

☑ As noted above, the LLM could be based upon a general-purpose LLM, or it could be based upon a domain-specific LLM.

System architecture: This is, effectively, the structure that enables the LLM or SLM to act as an agent. It is used for the orchestration and execution of the tasks and tools that constitute the agentic workflow. In particular, the system architecture imbues the LLM or SLM with memory and provides it with the ability to pursue goals, break them into steps, and use tools.

Tools: An agentic AI system has access to specific tools to gather information and execute actions. Depending on the agentic AI system, it might have access to tools like web navigation, file operation, running LLM-written code, querying and updating data, email handling, or keyboard or mouse control. Tools are accessed by agentic AI systems through standardized APIs.

Importantly, underpinning any AI system is an eco-system of upstream and downstream infrastructure and technologies—known as the AI technology stack—that makes AI development and deployment possible.[4] This same foundational stack underpins agentic AI systems.

Associated with these components of the agentic AI systems are actors:

- ☑ **Data curator:** This is the organization that curates, manages, or maintains the data used to train, fine-tune, or operate AI models and agentic AI systems. A data curator may be responsible for selecting, collecting, labeling, cleaning, and updating data to ensure it is accurate, representative, and compliant with applicable laws and licensing terms. In some cases, the data curator is the original creator of the data. In others, it could be a third party that licenses or aggregates datasets from multiple sources.

- ☑ **Developer of the LLM or SLM:** This is the organization responsible for developing the LLM or SLM that provides the agentic AI system with reasoning capabilities. The base model could be a general-purpose or a domain-specific LLM or SLM. Many AI models are also adding their own built-in agentic capabilities that support performing web searches to support reasoning, etc. In some instances, the developer of the LLM or SLM may also be the data owner/curator.

- ☑ **Developer of system architecture:** This is the organization that develops system architecture for an AI agent. In some instances, the developer of the system architecture is the same as the developer of the LLM or SLM, but in other instances, it could be a different organization.

- ☑ **Tool developer:** This is the organization or actor that develops the tools that the model accesses in order to execute on its tasks. In some instances, tools are not specifically designed for an agentic AI system. Companies that develop tools include established LLM developers, major software companies, as well as smaller developers that specialize in a particular niche.

- ☑ **Agentic AI system integrator:** This is the organization that assembles the model, agentic architecture, and tools into a cohesive system, ensuring they work together reliably. In some cases, the integrator may also be the deployer, but not always (especially in enterprise contexts).

- ☑ **Deployer of agentic AI system:** The ultimate deployer of the system, who makes the choice about how to put the system into use. Deployers set the operational parameters for the agentic AI system.

- ☑ **Actors responsible for managing and verifying credentials for agentic AI systems:** Such actors include Identity and Access Management (IAM) platforms that authenticate agentic AI systems and authorize their access to data, APIs, and systems.[5]

> → It is important to note that these roles are not always distinct. A single organization may play multiple roles in the value chain (e.g., developing the LLM or SLM, developing the system architecture, and deploying the final agentic AI system).

Benefits of Agentic AI Systems

Agentic AI systems offer numerous benefits for businesses and their customers. Below we highlight just a few examples of agentic AI's transformative potential to boost productivity, streamline work processes, and enhance security.

☑ **Transforming Public Sector Governance:** Agentic AI holds the potential to transform key government and public administration layers, including service delivery, crisis response, compliance, policymaking, and procurement.[6]

☑ **Enhancing Cybersecurity:** As security teams race to outpace AI-wielding threat actors, agentic AI can save users hours of manual work by completing tasks such as initial detection triage or prioritization of alerts on their behalf.[7, 8, 9] Agentic AI can allow Security Operations Center teams to focus on the most critical threats and perform more advanced tasks, while agentic AI handles less complex issues.[10]

☑ **Improving Customer Service:** Agentic AI can streamline customer service by automating tasks like customer support,[11] dispute resolution, and know-your-customer updates.[12] Such capabilities allow 24/7 service and help employees focus on more complex work.[13]

☑ **Empowering Advanced AI Agents and Companions Across Consumer Devices:** Agentic AI enables intelligent, context-aware assistants to operate seamlessly on smartphones, PCs, wearables, and vehicles, enhancing personalization, productivity, and real-time decision-making in everyday life.

☑ **Improving Data Analytics:** Agentic AI systems analyze and synthesize information at large scale, identifying patterns and delivering insights that inform decision making.[14]

☑ **Streamlining HR:** Agentic AI systems provide predictive workforce planning, staffing strategies, and can handle FAQs, manage time off, payroll, and administer benefits such as healthcare, retirement plans, career growth opportunities, and more.[15, 16]

☑ **Enhancing Finance and Compliance:** Agentic AI systems can audit financial transactions in real time, detecting fraud and updating compliance reports.

☑ **Strengthening Manufacturing:** Agentic AI transforms industrial operations by autonomously executing complete manufacturing workflows, with agentic AI systems collaborating across design, planning, operations, and maintenance to deliver up to 50% productivity increases.[17] These agentic AI systems handle complex tasks like production optimization, resource allocation, and predictive maintenance while integrating with both digital and physical systems. Agentic AI allows for interoperability between different agents and third-party systems, creating a unified environment where agentic AI systems work collaboratively to optimize entire manufacturing processes while maintaining human oversight.[18]

☑ **Optimizing Global Supply Chains:** Agentic AI systems can combine high-precision machine learning models with real-time inventory monitoring to streamline logistics, reduce inventory management costs, cut unnecessary inventory stock, and strengthen supply chain efficiency and resilience.

☑ **Enhancing Industrial Productivity:** Agentic AI systems can profoundly enhance industrial and critical infrastructure sectors by seamlessly integrating with operational technologies (OT) and functioning as physical AI. They deliver substantial gains in efficiency, safety, and operational precision, boosting productivity and value across manufacturing and essential services.[19]

☑ **Improving Marketing:** Agentic AI systems now help create personalized campaigns with content creation, A/B testing and cross-channel optimization.

☑ **Enhancing Privacy:** Agentic AI systems can minimize human error, limit data access by design, and act as privacy intermediaries. Such capabilities can help shield individuals and organizations from excessive data collection, access, and use while still delivering personalized experiences.[20]

Managing Risk Related to Agentic AI Systems

As we introduced in our prior paper on Foundation Models and the AI Value Chain[21], like other AI models and systems, agentic AI systems have the potential to present new risks or, in certain instances, exacerbate existing risks. Importantly, risk levels of agentic AI systems may vary by deployment context. Some applications present minimal risk while others require careful safeguards. Technology-neutral frameworks like the NIST AI Risk Management Framework remain helpful to risk management as the practices contained within can apply to multiple types of AI systems and models. We discuss in more depth how such frameworks can evolve in the era of agentic AI.

☑ **Jaggedness:** While LLMs that underpin agents are trained on encyclopedic volumes of data, they can exhibit jagged intelligence. Jagged intelligence means being highly capable in some areas but struggling with simpler tasks that most humans find intuitive and can reliably solve.[22] Understanding these deeper characteristics is crucial for designing robust agentic AI systems. Because agentic AI systems often chain multiple steps together, an error in one step based on a "jagged" area of incompetence can cascade through the entire workflow. Such a jagged failure can compound over time, leading to a sequence of flawed decisions and resulting in a series of failures.

☑ **Privacy Implications:** If agentic AI systems are given unrestricted access to resources, databases, or tools without necessary guardrails in place, they could potentially collect, store, and transmit sensitive information about users, interactions, and behavior patterns.[23] Such unrestricted access and accumulation of data may create new privacy vulnerabilities unless properly addressed. Further, whether the information is stored on a device or in cloud environments, it could create new attack surfaces for surveillance, exfiltration, and misuse, and may be legally accessible to third parties such as in the case of a court order. Privacy-enhancing technologies, like differential privacy and trusted execution environments, are especially important in helping to manage privacy risks in an agentic AI system. Other tools, like ethical data governance and provenance practices, can also play a key role in managing privacy risks.

☑ **Security Vulnerabilities:** Agentic AI systems, like other AI systems, may be vulnerable to a wide range of exploits.[24] For instance, attackers can perform prompt injection—feeding malicious instructions to the agent—to bypass its safety measures or gain unauthorized access. Similarly, an agentic AI system's underlying model could be "poisoned" with corrupted data. These risks may grow as agentic AI systems are integrated with software environments, APIs, files, and IoT devices. If compromised, an agentic AI system might take harmful actions autonomously, with little or no human oversight. Malign actors may also exploit weaknesses in the models themselves or the external tools they rely upon.

- **Cyber Threat Landscape Implications:** Agentic AI has the potential to transform the entire cyberattack chain by enabling the deployment of autonomous, specialized agentic AI systems that work in concert. These purpose-built agentic AI systems are designed to operate independently, test various attack methods, execute actions, and dynamically adjust tactics based on real-time feedback, forming an intelligent, self-correcting attack system. The emergence of agentic AI is an accelerating reality that challenges existing threat detection, response, and mitigation strategies. The rapid pace of AI development means that both attackers and defenders are operating in a rapidly evolving landscape, where the capabilities of security vendors and malicious actors are constantly changing.

- **Malicious Use:** Agentic AI systems can be misused, such as for large-scale generation and dissemination of disinformation, scaling up offensive cyber operations, or increasing access to expert capabilities in dual-use scientific research and development.[25]

- **Automation Bias:** Agentic AI systems have the potential to automate repetitive or complex tasks. At the same time, users may develop excessive trust in these models, a phenomenon known as automation bias, leading to poor decision-making or loss of essential skills. If agentic AI systems become general-purpose substitutes for human effort, society may face longer-term consequences, including cognitive dependency and workforce deskilling.[26]

- **Accountability for Decision-making:** In case of errors or harmful outcomes caused by agentic AI systems, defining accountability may become complicated depending on the level of autonomy of the system, the complexity of the technology stack that may compose an agentic AI system, and their dynamic decision-making.[27] Potential failures of agentic AI systems can erode trust in businesses and other organizations deploying these models.

- **Value Misalignment and Unintended Actions:** Agentic AI systems might pursue goals in ways that do not align with the user's intent or human values. Such misalignment could happen if instructions given to the agent are unclear or if the agent optimizes too narrowly. Greater autonomy of agentic AI systems might lead to a higher risk of accidents or unethical actions caused by this misalignment.

- **Workplace Shifts:** The widespread adoption of agentic AI systems might shift labor and skill demand in complex ways. These systems can automate a wide range of tasks, including cognitive and white-collar functions. As a general-purpose technology, AI is expected to diffuse across many industries, although its long-term impact on the job market remains uncertain. Further research into trends in AI adoption will help governments and industry alike to better understand how AI systems, including agentic AI, will impact employment levels and various occupations.[28]

  - Agentic AI systems' capabilities and implications are yet to be fully understood. Along with opportunities, agentic AI may also pose risks related to privacy, security, accountability, and lead to broader socio-economic changes. Many of these risks are shared with other AI systems, but the increased autonomy and interactivity of agentic AI systems may exacerbate them. As adoption grows, policy frameworks, technical solutions to protect data, and robust security standards will be essential to ensure the responsible development and deployment of agentic AI.

Policy Considerations

As governments consider how to encourage secure and trusted use of agentic AI, we continue to advocate for policies that are targeted at addressing specific harms while allowing for advances in innovation. In doing so, governments should evaluate both regulatory and non-regulatory approaches, proceeding to regulation only in cases where gaps are identified. In some instances, existing regulatory regimes may already cover agentic AI.[29]

Because this is an evolution of AI technology, our prior recommendations, including those outlined in our 2024 AI Accountability Framework[30], still apply. For example, even in the context of agentic AI, transparency along the AI value chain, consumer-facing transparency, and appropriately allocating roles and responsibilities within the AI value chain are all things that any robust public policy framework should seek to incorporate. These recommendations are not intended to be prescriptive and instead introduce areas that policymakers should consider in developing a framework that will appropriately support responsible development and adoption of agentic AI systems.

Promoting adoption

To promote broad adoption of agentic AI, policy makers should take steps to enable deployment across the economy, foster institutional capacity, and support talent development.

☑ **Develop and advance a national AI adoption strategy.** Governments should consider developing national AI strategies that outline ways in which they will support the adoption and integration of AI, including agentic AI. Such a strategy should outline the specific steps that governments will take to support innovation and engender adoption.

As a part of this, governments should endeavor to identify priority sectors where adoption could yield significant societal benefit, like in national security, cybersecurity, and fraud detection, and work to explore how to promote adoption in these areas. Furthermore, adoption strategies could also consider how adoption could support specific groups, such as upskilling the workforce, training small businesses and entrepreneurs to help them compete.

☑ **Address practical constraints for government adoption of agentic AI.** We encourage governments to identify what actions can be taken under existing authorities and where new guidance or regulation may be necessary to support government adoption of agentic AI. Considering the rapid pace of technological change, the public sector should align government contracts with commercial terms and conditions to the greatest extent practicable and leverage existing authorities to promote the use of innovative acquisition strategies. Doing so will ensure that the public sector can adapt to a modern acquisition environ-ment with the tools already at its disposal.

☑ **Update and improve the government's information technology (IT) infrastructure to prepare government data to benefit from agentic AI.** Government IT infrastructure has not kept pace with the speed of innovation. Governments should continue to focus on IT modernization efforts to adopt modern data management practices. This means replacing outdated and end-of-life technology, reducing data fragmentation, and consolidating and standardizing government-data sharing practices so that data is properly structured and can scale to make better use of investments in agentic AI.

☑ **Ensure that government contract requirements promote outcomes-based approaches to enable greater adoption of agentic AI.** Government contracts should refrain from specifying the type of technology or specific resources that must be used in performance of a contract. Instead, governments should focus on identifying a specific outcome that may be particularly well-suited for advanced AI tools to accomplish, such as improving the delivery of public benefits. AI, as a broad concept, is already being used in many commercial applications, tools, and processes used by the government. Focusing on outcomes will enable government contracting officers to promote greater collaboration between government and industry to identify the right AI tool suited for the mission, such as agentic AI.

☑ **Leverage agentic AI to modernize and make government operations and the delivery of services more efficient.** Governments can support IT modernization efforts by leveraging the power of agentic AI to deliver simpler, faster, and more reliable services. With the ability to automate routine, and task-oriented processes that underpin many core government functions—such as constituent casework, network security operations, etc.—the government can empower the federal workforce to focus on delivering more efficient services and break down barriers to the public's access to critical government resources.

☑ **Support voluntary testing environments, like regulatory sandboxes.** Leveraging regulatory sandboxes allows innovators and businesses to develop and test new technologies in a controlled environment that is free from certain regulatory regimes or compliance processes. This allows for experimentation with agentic AI, promotes innovation, and contributes to learning on behalf of the relevant authorities. Policymakers should identify where sandbox activities can occur under current law, and where new authorities might be needed.

☑ **Support workforce readiness to enable adoption of agentic AI.** As agentic AI continues to evolve, governments should work with industry, academia, and civil society to understand its impact on job functions and highlight the skills that will be most relevant to empower the future workforce. Investing in workforce readiness through public-private partnerships that focus on upskilling, reskilling, and digital literacy is imperative. Some of these policies may include "training sabbaticals" and learning-time carve-outs as best practice, offering grant or tax credits for firms that allocate protected time for AI skilling with human-oversight principles.

Fostering trust

Trust is foundational to promoting adoption. With that in mind, policy-makers should take steps to engender trust when considering agentic AI.

☑ **Adopt a risk-based approach tailored to the deployment context.** Agentic AI systems should be viewed as a spectrum: they can vary in their level of autonomy, the tools they use, and the environments in which they operate. Therefore, assessing risks requires considering the degree of autonomy, as well as the context in which an agentic AI system is deployed.

For example, an agentic AI system that acts as a personal assistant and independently schedules personal calendar appointments may operate with the same degree of autonomy as one that manages a hospital's patient scheduling. However, the potential risks associated with an error in these two contexts differs significantly, highlighting the importance of taking deployment context into account in risk management.

☑ **In seeking to appropriately manage risk, policymakers should leverage existing frameworks for agentic AI governance.** Our AI Accountability Framework lays out key practices that we believe are appropriate to apply to high-risk AI systems, as well as to frontier AI models. We encourage policymakers to reference this document in considering regulatory or legislative approaches to AI. Assuming existing AI regulatory and policy frameworks have been crafted in a sufficiently flexible and future-proof way, they should also be applicable to agentic AI. Similarly, existing legal regimes that address concerns related to possible harms should also apply to agentic AI. Indeed, an entirely new and different approach is unnecessary.

➔ However, given some agentic AI systems' degree of autonomy, it may be appropriate to undertake more rigorous red-teaming and testing pre- and post-deployment of an agent system.

☑ **Foster transparency in the agentic AI value chain.** Policymakers should recognize that agentic AI systems are layered, and therefore, the behavior of the agent system is influenced not only by the base model, but also by the environment in which it operates, the data it uses, its goal orientation, and its system architecture. Transparency should therefore be tailored to the unique characteristics of such systems.

- Rather than creating new disclosure regimes from scratch, policy-makers should explore how existing tools, like model cards and system cards, can be leveraged to communicate information about agent systems to downstream deployers. System cards could be extended to reflect agent behavior(s) and characteristics, including, at a high level, the ways in which the system was tested and evaluated prior to release but also potentially communicating information about human oversight, permissions, and the tools and system architecture incorporated.
- As we outline in ITI's Transparency Policy Principles[31], our seminal paper on AI transparency, the audience for transparency matters. The information that is important to communicate to a downstream actor in the AI value chain will likely differ from the type of information communicated to the end-user, especially in consumer-facing settings.
- For example, it may be sufficient to provide notice to a user operating an agentic AI system that they are interacting with an AI agent system, describing its role, primary capabilities, and known limitations. At the same time, for regulatory purposes, developers and integrators may need to provide, upon request, more detailed information on the agent such as autonomy level, tool access, and escalation levels.[32] This dual-level design ensures meaningful transparency without overwhelming users or under-informing regulators.

☑ **Recognize that responsibilities should be allocated based upon function and role in the agentic AI value chain.** As outlined above and in all ITI's existing AI policy collateral, the agentic AI value chain includes a variety of players, each with different responsibilities. Stakeholders throughout the value chain play a role in ensuring that AI is developed and deployed responsibly. In considering if and how to apply obligations in the context of a policy framework or other-wise, policymakers should seek to apply said obligations proportionally, considering what each stakeholder in the agentic AI value chain has control over. Understanding this allocation of responsibilities will prevent duplication and ensure that each actor is responsible for what they can meaningfully influence.

- For example, with transparency, the model developer will have insight into the base model's limitations and capabilities and should communicate that information to other downstream actors. The agentic AI system integrator, however, will have insight into the system's planning logic, level of autonomy, and its tool use.
- Finally, the deployer of an agentic AI system will have ultimate responsibility for usage policy, knowledge of the environment in which the system is deployed in, the data it has access to, and the human oversight mechanisms that have been put in place.
- It is important to recognize that these roles and responsibilities are not necessarily mutually exclusive. In some instances, a model developer and agentic AI system integrator could be one and the same. In other cases, the system integrator and the deployer could be the same.
- At the same time, it is important to keep in mind that agentic AI systems' behavior often emerges from complex interactions between components in ways that can make clean responsibility allocation challenging. Unlike traditional supply chains, systems of agents can exhibit emergent behaviors from the interplay of models, prompts, and tools. That said, while we believe allocation of responsibility in a proportionate way remains imperative, it may be more challenging in certain instances given the way agentic AI systems interact.

☑ **Adopt a comprehensive federal privacy law.** Implementing privacy legislation provides a unified, national framework for responsible business practices. It strengthens and simplifies data management practices across the economy and provides a firm foundation for advancing trustworthy AI.

☑ **Enhance the security of agentic AI.** Security measures may include ensuring risk-based and actionable human control and oversight, implementing limitations on agentic AI capabilities (e.g. purpose-specific entitlements on capabilities and resource access), and making agentic AI actions and planning observable and verifiable. Further efforts may also include the use and development of best practices and promotion of commercially available solutions that help mitigate risk. It may also be worth considering how to incentivize the adoption of frameworks developed for agentic AI, especially as there are many security frameworks emerging across the globe, which will be developed, consolidated and refined over time. We encourage policymakers to leverage guidance and/or encourage the adoption of guidance produced by global organizations. For example, the Coalition for Secure AI published Principles for Secure by Design Agentic Systems to provide practical implementation strategies that security teams can deploy.[33]

☑ **Promote the adoption of voluntary best practices to govern risks related to agentic AI.** Policymakers should support the adoption of existing standards, frameworks, and best practices for AI risk management, like NIST's AI Risk Management Framework and the ISO/IEC 42001 series. Many of these frameworks remain relevant in an agentic AI context because they are technology-neutral and govern processes. That said, it is also worth considering if and how existing frameworks can evolve to allow organizations to better operationalize suggested practices.

For example, developing a risk-tiering structure for agentic systems could be helpful in ensuring consistency of risk assessment across stakeholder groups.

This could potentially be undertaken in the form of an Agentic AI Profile in the context of the AI RMF. Because agentic AI is an evolving area, best practices will simultaneously evolve as research continues and understanding matures.

➾ Explore the use of privacy-enhancing technologies (PETs) and data governance and provenance standards. PETs, such as differential privacy, data minimization, and secure computation, can help protect user data during storage, processing, and transmission. In parallel, promoting the development of clear industry-led data governance and provenance standards can help ensure that agentic AI systems handle personal information in ways that respect individual rights and comply with applicable laws.

☑ **Promote industry-backed, open standards and protocols for agentic AI systems.** As agentic AI systems become integral to enterprise workflows, they require robust mechanisms to discover, authenticate, and interact with external systems and other agents. Open, industry-led standards and protocols provide the foundational layer for this interoperability—much like TCP/IP did for the internet—while ensuring that downstream companies can effectively deploy and integrate agentic systems. These open standards foster both cross-industry collaboration and healthy competition among AI providers, accelerating innovation while maintaining compatibility. Industry-led open protocols such as MCP, A2A, and ACP have already demonstrated their value in powering enterprise-level agentic workflows. Government endorsement and promotion of these open standards would encourage broader adoption, drive further development, and help establish the interoperability framework necessary for a thriving agentic AI ecosystem.

☑ **Promote multistakeholder collaboration on evaluating emerging risks.** Given the evolving nature of agentic AI systems, policymakers should encourage continued collaborative research to both understand and evaluate emerging risks and ensure consistency in how these risks are measured and assessed.

- Leveraging existing multilateral forums, such as the OECD or G7, can enhance multistakeholder collaboration on emerging risks and opportunities related to agentic AI. Previous initiatives, like the Hiroshima AI Process, have produced valuable outcomes, and similar discussions could facilitate the adoption of agentic AI while effectively managing associated risks.

- Supporting the development of shared security evaluation frameworks, responsible vulnerability disclosure mechanisms, and interoperable standards alongside international partners. Such cooperation can help build collective resilience and promote the safe and ethical deployment of agentic AI systems globally.

- Additionally, evaluating agentic AI systems will require the development of new or the adaptation of existing metrics. Those used to evaluate traditional machine learning may need to evolve to appropriately assess an agentic AI system's performance. Considerations should include whether the agent meets the original user request, selects the most appropriate tools, and understands the initial goals. Government, industry, and academia should collaborate to develop these standards and develop collaborative evaluation frameworks for agentic AI systems.

- Finally, ongoing research into an agentic AI system's ability to independently pursue goals and persistently make decisions, including methods for voluntarily documenting agent behavior, could be a helpful area of research to support.

References:

[1] See, for example, definitions in Kasirzadeh, A., & Gabriel, I. (2025). Characterizing AI Agents for Alignment and Governance. arXiv. https://arxiv.org/abs/2504.21848

[2] The constituent elements that Kasirzadeh et al. lay out in their publication is a helpful lens through which to view agentic AI systems: autonomy, efficacy, goal complexity, and generality.

[3] To learn more about small language models (SLMs), please see Huang, X. (2025). How we're preparing for the next era of AI. Zoom Blog. https://www.zoom.com/en/blog/what-is-agentic-ai/?ampDeviceId=e66b26cb-eb6f-4620-9e64-7b510f927499&ampSessionId=1757615730011;

[4] To learn more about the AI technology stack, please see ITI. (2025). The AI Technology Stack and Why It Matters for AI Policy and Governance. https://www.itic.org/documents/artificial-intelligence/ITI_AITechnologyStack.pdf

[5] To learn more about the AGNTCY project, please visit https://agntcy.org/; Outshift by Cisco. (2024). New AI Agent Identity framework from the AGNTCY. https://outshift.cisco.com/blog/ai-agent-identity-framework-agntcy

[6] Global Government Technology Centre. (2025). The Agentic State: How Agentic AI Will Revamp 10 Functional Layers of Public Administration. https://www.globalgovtechcentre.org/executive-summary#content

[7] Dhingra, A. (2025). The AI-Ready Enterprise: Building the Intelligent Workplace with Cisco. Cisco Blogs. https://blogs.cisco.com/news/the-ai-ready-enterprise-building-the-intelligent-workplace-with-cisco

[8] Cisco. (2025). Announcing Cisco AI Canvas: Revolutionizing IT with AgenticOps. The Newsroom. https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2025/m06/announcing-cisco-ai-canvas-revolutionizing-it-with-agenticops.html

[9] Zaitsev, E. (2025). CrowdStrike Leads Agentic AI Innovation in Cybersecurity with Charlotte AI Detection Triage. Crowdstrike Blog. https://www.crowdstrike.com/en-us/blog/agentic-ai-innovation-in-cybersecurity-charlotte-ai-detection-triage/

[10] Krider, Z., Rajagopalan, S., Lobrecht, R. (2025). Trellix uses AWS GenAI for Cybersecurity Integration. AWS Partner Network (APN) Blog. https://aws.amazon.com/blogs/apn/trellix-uses-aws-genai-for-cybersecurity-integration/

[11] AWS. (2025). Building a Generative AI Contact Center Solution for DoorDash Using Amazon Bedrock, Amazon Connect, and Anthropic's Claude. https://aws.amazon.com/solutions/case-studies/doordash-bedrock-case-study/

[12] Cisco. (2025). How agentic AI will transform customer experience. The Newsroom. https://newsroom.cisco.com/c/dam/r/newsroom/pdfs/Cisco-CX-Agentic-AI-Research.pdf

[13] Levitt, K. (2025). AI On: How Financial Services Companies Use Agentic AI to Enhance Productivity, Efficiency and Security. NVIDIA Blog. https://blogs.nvidia.com/blog/financial-services-agentic-ai/

[14] BCG. (2025). AI Agents. https://www.bcg.com/capabilities/artificial-intelligence/ai-agents

[15] PWC. (2025). AI AGENTS CAN REIMAGINE THE FUTURE OF WORK, YOUR WORKFORCE AND WORKERS. PWC TECH EFFECT. HTTPS://WWW.PWC.COM/US/EN/TECH-EFFECT/AI-ANALYTICS/AI-AGENTS.HTML

[16] IBM. AI agents for human resources (HR). https://www.ibm.com/products/watsonx-orchestrate/ai-agent-for-hr

[17] Siemens. (2025). Revolutionizing manufacturing with Siemens' Industrial AI agents. Siemens Newsroom. https://www.siemens.com/us/en/company/press/siemens-stories/digital-industries/ai-agents-manufacturing.html

[18] Vardhan, H., Dixit, D., Kumar, G. (2025). How Apollo Tyres Is Unlocking Machine Insights Using Agentic AI-Powered Manufacturing Reasoner. AWS Blogs. https://aws.amazon.com/blogs/machine-learning/how-apollo-tyres-is-unlocking-machine-insights-using-agentic-ai-powered-manufacturing-reasoner/

[19] Siemens. (2025). Revolutionizing manufacturing with Siemens' Industrial AI agents. Siemens Newsroom. https://www.siemens.com/us/en/company/press/siemens-stories/digital-industries/ai-agents-manufacturing.html

[20] Finch, L. (2025). AI Agents Will Enhance — Not Impair — Privacy. Here's How. Salesforce. https://www.salesforce.com/news/stories/agentic-ai-for-privacy-security/

[21] https://www.itic.org/documents/artificial-intelligence/ITI_AIPolicyPrinciples_080323.pdf

[22] Andrej Karpathy first introduced the term "jagged intelligence."

[23] Google DeepMind. (2024). The Ethics of Advanced AI Assistants. arXiv. https://arxiv.org/abs/2404.16244

[24] Lumenova. (2024). AI Agents: Potential Risks. https://www.lumenova.ai/blog/ai-agents-potential-risks/

[25] Kraprayoon, J. (2025). AI Agent Governance: A Field Guide. Institute for AI Policy and Strategy. https://www.iaps.ai/research/ai-agent-governance

[26] Kumayama, K. D., Chiruvolu, P., & Weiss, D. (2025). AI agents: Greater capabilities and enhanced risks. Reuters. https://www.reuters.com/legal/legalindustry/ai-agents-greater-capabilities-enhanced-risks-2025-04-22/

[27] Lee, L. (2025). In a World of AI Agents, Who's Accountable for Mistakes? Salesforce. https://www.salesforce.com/blog/ai-accountability/

[28] To learn more about most recent research on AI's possible impact on workforce, please see Anthropic. (2025). Anthropic Economic Index. Understanding AI's Effects on the Economy. https://www.anthropic.com/economic-index#us-usage; Goldman Sachs. (2025). How Will AI Affect the Global Workforce? https://www.goldmansachs.com/insights/articles/how-will-ai-affect-the-global-workforce; World Economic Forum. (2025). Future of Jobs Report. https://reports.weforum.org/docs/WEF_Future_of_Jobs_Report_2025.pdf

[29] For example, if agentic AI is used in a high-risk setting, it would already be covered by the EU AI Act. Similarly, technology-neutral laws, such as those governing privacy and data collection or consumer protection, may already cover agentic AI.

[30] https://www.itic.org/documents/artificial-intelligence/AIFIAIAccountabilityFrameworkFinal.pdf

[31] https://www.itic.org/documents/artificial-intelligence/ITIsPolicyPrinciplesforEnablingTransparencyofAISystems2022.pdf

[32] An escalation level refers to a pre-defined threshold that will prompt an agentic AI system to involve human oversight.

[33] Coalition for Secure AI. (2025). Announcing the CoSAI Principles for Secure-by-Design Agentic Systems. https://www.coalitionforsecureai.org/announcing-the-cosai-principles-for-secure-by-design-agentic-systems/

ITI

Promoting Innovation Worldwide

The Information Technology Industry Council (ITI) is the premier global advocate for technology, representing the world's most innovative companies. We promote public policies and industry standards that advance competition and innovation worldwide.