FPF | CENTER FOR ARTIFICIAL INTELLIGENCE

Defining "Agentic": Why We Need a Shared Taxonomy for AI Agents

September 2026
Stacey Gray and Jameson Spivack

The authors would like to thank Alex Rosenberger and Justine Gluck for their feedback and contributions.

Join us on October 21st in California for a Workshop on Advanced Issues in Agentic Deployment.

This summer, lawmakers in the United States released two of the first legislative proposals for regulating "AI agents": the federal AI AGENT Act and California SB 1106. The proposals offer significantly different legal definitions of agentic AI, tailored to their different purposes. The AI AGENT Act, which would establish a framework for consumer-facing agents, defines agents in terms of their relationship with a user. California SB 1106, meanwhile, was aimed at building public infrastructure for cybersecurity, and defined agents in terms of their technical architecture and capabilities.

As FPF's Center for AI has written previously, a rough consensus has emerged from AI literature about what defines whether AI is "agentic" from a technical perspective: in most cases, it involves a combination of advanced reasoning capabilities and access to consequential systems. Yet the term "agent" has different meanings to different audiences. For some, it refers to the powerful "agent swarms" seen in recent headlines about frontier labs, with all of the safety and alignment concerns they raise. For others, an "agent" describes a broad range of simple, automated systems, such as an LLM-powered web scraping tool. And for many others, "agentic-ness" is defined not by the technical design of the system at all, but instead by the level of human involvement (as with the levels of autonomy in vehicles) or by the degree to which the agent's behavior is attributable upstream to a particular owner or user (as with "principal-agent" law). The term "agent" can also refer to a particular relationship or legal status, such as in the phrase "authorized agent" used in the California Consumer Privacy Act, which is applied well beyond the context of AI.

Almost all large enterprises, from banks to retailers, are figuring out how to deploy AI agents in both internal and customer-facing contexts to automate and streamline recurring, labor-intensive work. However, lacking common definitions and frameworks for thinking about what constitutes an agent can be a significant communication and governance challenge, especially when legal

frameworks apply obligations based on a binary "in or out" scope. Governing agents effectively requires a shared language for agents and agentic behavior. This blog post explores this issue by examining technical literature, U.S. legislative efforts, and other major global frameworks.

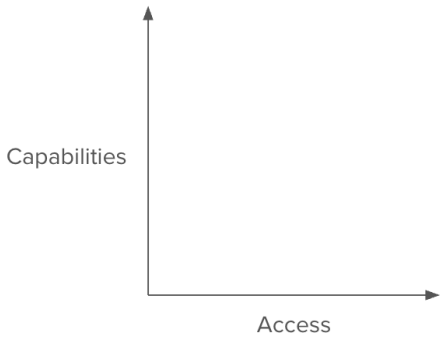
1. What's in a Name? Agentic AI Definitions in Technical Literature

In popular discourse, speakers use the terms "AI agents" or "agentic AI" to refer to many things, from a chatbot assistant that browses the Internet to a sophisticated system of LLMs and infrastructure that analyzes a user's computer files and accesses external accounts on their behalf. In fact, the conflation of agents with other existing AI systems has led some to joke that agents are just "three (LLMs/bots/prompts) in a trench coat." Jokes aside, the imprecise use of terminology obscures the fact that there are real, substantive differences that distinguish frontier "agents" from automated tools and LLMs more generally.

As FPF has written previously, much of the technical and governance literature characterizes AI agents as a significant evolution of existing AI, with definitions generally converging around their *capabilities* and their *access*. OpenAI, for example, characterizes agentic AI systems by "the ability to take actions which consistently contribute towards achieving goals over an extended period of time, without their behavior having been specified in advance." The World Economic Forum, meanwhile, emphasizes agents' *autonomy* and *authority*, the latter defined as "the granted permissions and access rights to perform specific actions within defined boundaries." In other words, AI agents can break down tasks into multiple steps with some degree of autonomy, and carry out those tasks across a range of tools and applications. Anthropic's Claude Cowork, for example, can plan and execute work based on user prompts, using local files and external systems like Google Drive, and without constant user approval.[1]

These two prongs defining agentic AI—capabilities and access—matter for how agents are governed, because they are the factors that raise unique challenges compared to other kinds of AI, and which may require novel governance approaches. Agentic systems exist along a spectrum, with different tools exhibiting varying levels of autonomy, adaptability, action-taking, and data access depending on the use case, and a shared conception of agentic AI should take these features into account.

[1] Some literature further distinguishes between "AI agents" and "agentic AI," with agentic AI referring to the higher-level system that coordinates multiple agents and their interactions with other tools, and agents referring to specific software that actually carries out tasks. This post will, unless otherwise noted, refer to AI agents and agentic AI as part of one system.

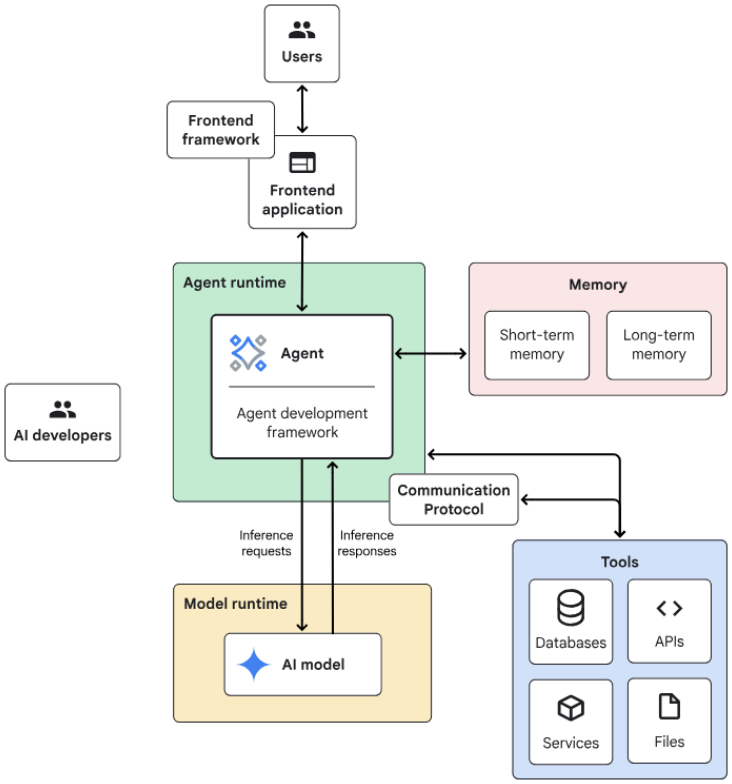


A simple way to conceptualize AI agents for the purpose of anticipating risks is by evaluating both their capabilities and the systems and tools to which they have access.

Capabilities: autonomy, adaptability, and action-taking

A distinguishing feature of agentic AI is the degree of latitude a system has in deciding *how* to accomplish a goal. A scripted bot that executes predetermined steps, even if it does so without human review, is not agentic because it does only what it was configured to do. On the other hand, a system that receives an open-ended objective and independently determines the path to achieve it—by breaking the goal into subtasks, sequencing and prioritizing those subtasks, and revising the plan based on intermediate results—is agentic because it is autonomous and responds to changing conditions. As legal scholar Noam Kolt puts it, LLMs are "copilots" and agents are "autopilots."

Most systems deployed today sit somewhere between the scripted bot and the fully-autonomous agent. Many "agentic" tools can plan multi-step tasks, but seek user approval or confirmation before consequential steps, a safeguard meant to prevent serious mistakes. Others operate independently, but within a narrow domain like software bug detection, and cannot adapt outside of their specific context. Where an AI system falls on this spectrum matters for governance because autonomy and adaptability are what make agent behavior harder to predict and audit—oversight models built for deterministic automation do not necessarily transfer to non-deterministic systems that may behave in unpredictable ways.



An illustration of agentic AI's components and the tools which it may access. Source: Google's Cloud Architecture Center

Agentic AI is also distinguished by the kinds of actions it can take on behalf of a user. Today's agents can click buttons, fill out forms, search and update databases, send messages, and execute code, with developers continually building additional tools and integrations. Agents are able to serve these functions by observing their environments, deciding what action to take, translating that decision into a call to a tool (such as navigating the web or operating on files), and then evaluating the results before repeating the cycle. When it comes to governing agentic systems, of course, not all actions carry the same weight—an agent that drafts and posts social media content based on rigid criteria may pose less of a threat than an agent that manages business systems, makes financial decisions that would typically require auditing, or makes other decisions with legal effects like hiring. A risk-based, proportionate approach to governance is tailored to the kinds of actions agentic systems can take, and how consequential these actions might end up being.

Access to systems and tools

Agents are enabled, or constrained, in their decision-making and action-taking by the systems and tools to which they have access. The ability to access systems external to the AI model itself is not exclusive to agents; retrieval-augmented generation (RAG), for example, is used to connect

LLMs to additional datasets in order to generate more relevant, high-quality output. Where traditional RAG provides read-only access, however, more agentic systems generally have the ability to write or act on a user's data or files. An LLM utilizing RAG could read a user's inbox to understand context before replying to a query; an agent, on the other hand, could draft and send an email on their behalf. An LLM might search the Internet and flag sales for the user; an agent, operating through shared protocols and standards, might make purchases. This may entail users granting agents access to authenticated accounts, such as on banking or retail websites, which has led to litigation over whether platforms can block agents that scrape their sites without authorization. Governance frameworks can recognize the level of privileged access—both read and write—that agents have, as well as how the ability to combine data from across previously-siloed systems may grant them additional power.

2. Emerging U.S. Legislative Definitions and Global Regulatory Trends

The two recent U.S. agentic AI bills represent distinct—but not necessarily contradictory—approaches to conceptualizing agentic AI, tailored to the risks each bill addresses. The federal AI AGENT Act defines agents by their relationship with a user, applying to "custodial user agents" that are expressly authorized by, and loyal to, the user. The definition doesn't focus on what an agent can do, but is scoped to cover software-based agents that interact with online platforms on the user's behalf "in a transparent, documented, scope-limited, and revocable manner." This aligns with the overall thrust of the discussion draft, which is generally a consumer protection and competition bill aimed at ensuring agents are loyal to users and can interoperate across online platforms under fair and reasonable terms.

Meanwhile, California SB 1106, which did not proceed but could be re-introduced in 2027, provided a much more granular definition of agents, reflecting both the capabilities and access prongs of agentic AI discussed in the literature:

> *(1) "Agentic artificial intelligence" or "agentic AI" means an artificial intelligence system that can pursue multistep or abstract goals by independently breaking them into tasks or delegating them, adjusting its own behavior based on results rather than waiting for user instruction or consent, and that can act in the world, including by executing code, using external tools, making purchases, accessing accounts that require authentication, or operating outside its own.*
>
> *(2) "Agentic artificial intelligence" does not include a system that only responds to direct prompts and instructions from the user and lacks network egress or shell access.*

This definition—which was amended from a previous, even-more-technical definition—distinguished agents from other forms of AI by emphasizing their autonomy, adaptability, and problem-solving. In contrast to the AI AGENT Act's focus on consumer protection and competition, SB 1106 was more broadly about state agencies conducting risk

analyses for AI and critical infrastructure, and cataloging high-risk AI uses. In each bill, agents are definitionally scoped to match the particular governance issues they seek to address.

In other emerging global frameworks, policymakers have also so far largely converged on a conception of agents emphasizing their technical capabilities and level of access (see table below). For example, Singapore's InfoComm Media Development Authority (IMDA), in its Model AI Governance Framework for Agentic AI, recognizes a lack of definitional consensus but nonetheless posits that agents share common features such as independent planning, decision-making, and action-taking to achieve a user-defined goal. The IMDA further defines agentic AI systems as software systems containing one or more agents that often combine nondeterministic and deterministic components. Similarly, in a request for information about agentic AI security, the U.S.' Center for AI Standards and Innovation (CAISI), housed within the National Institute of Standards and Technology (NIST), does not explicitly define AI agent systems but characterizes them as "capable of planning and taking autonomous actions that impact real-world systems or environments."

Notable Global Legislative and Regulatory Definitions of "Agentic AI"

Legislation	"Agent" definition
California SB 1106	"**Agentic artificial intelligence**" or "**agentic AI**" means an artificial intelligence system that can pursue multistep or abstract goals by independently breaking them into tasks or delegating them, adjusting its own behavior based on results rather than waiting for user instruction or consent, and that can act in the world, including by executing code, using external tools, making purchases, accessing accounts that require authentication, or operating outside its own. "Agentic artificial intelligence" does not include a system that only responds to direct prompts and instructions from the user and lacks network egress or shell access.
U.S. AI AGENT Act	**Custodial user agent** means a software-based agent that is expressly authorized by a user to interact with a large online platform provider on that user's behalf in a transparent, documented, scope-limited, and revocable manner.
U.S. National Institute for Standards and Technology (NIST), Center for AI Standards and Innovation	*Request for Information Regarding Security Considerations for Artificial Intelligence Agents*: **AI agent systems** are capable of planning and taking autonomous actions that impact real-world systems or environments. AI agent systems consist of at least one generative AI model and scaffolding software that equips the model with tools to take a range of discretionary actions.
Singapore InfoComm Media Development Authority	*Model AI Governance Framework for Agentic AI*: There is no consensus on what defines an **AI agent**, but there are certain common features – agents usually possess some degree of independent planning, decision-making, and action-taking (e.g. searching the web or creating files) over multiple steps to achieve a user-defined goal. Agentic AI systems are software systems consisting of one or multiple AI agents that may operate individually or collaboratively.

	Legal Responsibility for AI Agents (Discussion Paper): **Agentic AI systems** are software systems consisting of one or multiple AI agents that may operate individually or collaboratively. In practice, many agentic AI systems combine nondeterministic (e.g. LLM-based planning to determine what tool to call) and deterministic components (e.g. rules-based access controls on tools or partially scripted workflows).
European Commission Proposal for the Cloud and AI Development Act	'**AI agent**' means an AI system or a coordinated set of AI systems, that can perceive and act upon their environment, with a degree of autonomy, using tools as needed to achieve specific goals and adapt to changing inputs and contexts.
France National Commission on Informatics and Liberty (CNIL) and AI and Digital Council (CIANum) Note	*Agentic AI and protection of personal data: an equation with multiple unknowns for users*: **Agent-based AI** refers to a set of computer programs based on generative AI models that are capable of making autonomous decisions, orchestrating complex actions, and interacting with third-party services, with or without human validation. The architecture of an agent-based AI system generally relies on the interaction between several components: (i) An **"orchestrator" agent**, at the core of which the generative AI model enables the user to interact with the system using natural language. … (ii) **"Specialized" agents**, coordinated by the orchestrator agent, which can perform complex tasks requiring specific skills (code generation, text processing, online payments, etc.). (iii) All of these agents can then be connected to external services such as applications, databases, or web search tools, which they can utilize to process requests. *[Translated from French.]*

What's Next: Cross-Cutting Agentic Governance Issues

AI systems with independent judgment and privileged access raise a recurring set of questions that regulatory and governance frameworks must answer—and that privacy, legal, and governance teams deploying agents must address. Certain sectors, like finance and banking, that are at the forefront of agent adoption are already dedicating significant resources to addressing these questions. The following governance issues, among others, are relevant across sectors and applications, though in practice may require different approaches depending on relevant risks.

- **Identity**: When an agent acts, the parties on the other side of the interaction—such as platforms or merchants—need a way to know which agent is acting, who operates it, and which actor delegated it. Today there is no standard way to identify AI-driven traffic, and existing identity and access management tools don't apply neatly to agent identities. Many early agent infrastructure efforts have centered around identity, including proposals for agent IDs, credentialing, and cryptographic signatures.

- **Accountability**: Visibility into agent activity tends to break down when agents act across systems, over time, and through chains of delegation. Maintaining traceability requires dedicated logging and monitoring infrastructure, keeping records of what an agent observed, decided, and did, attributable to a responsible party.

- **Authorization and consent**: Existing consent frameworks assume a human being is clicking through a static, fixed-purpose application. Agents challenge this assumption,

requiring authorization that covers ongoing delegation whose specific actions cannot all be enumerated in advance.

- **Access and data minimization**: Agents perform better when they have broader access to a user's files and systems, which may put agent design in direct tension with data minimization. Establishing appropriately-scoped agent access will require novel approaches to setting least-privilege permissions, applying zero-trust principles, and allowing agents to convey intent, which may prove challenging given agents' autonomous and non-deterministic nature.

- **Reliability and assurance**: Even as agents' *capabilities* improve, they're not necessarily becoming more *reliable*. In other words, agents may increasingly be able to complete more types of tasks, and more quickly, but may not be able to do so accurately or consistently. Researchers also currently lack effective tools for measuring agents' reliability.

Join us in California on October 21st

These issues are the focus of FPF's AI Governance Workshop on Advanced Issues in Agentic Deployment on October 21. The workshop will bring together privacy, legal, AI governance, and engineering leaders to work through the practical challenges of deploying agents, including identity, accountability, authorization, and data access. We will also share an early look at FPF's draft Agentic AI Policy Taxonomy resources, which aim to provide a shared vocabulary for these conversations.

If your organization is building, deploying, or governing AI agents, we hope you will join us. Space is limited; please request to attend here.