Agentic Artificial Intelligence Glossary

2026

Authors:
Ziyang David Fan
Executive Director, ICAP
SVP, Tech & Innovation, SVLG

Mengyu Ruby Han
Associate, Tech & Innovation
and ICAP, SVLG

Chelsea Dixon
Associate, Tech & Innovation, SVLG

SILICON VALLEY LEADERSHIP GROUP

INSTITUTE for CALIFORNIA AI POLICY

Executive Summary

Artificial intelligence is entering a new phase of practical deployment. As organizations move from experimenting with generative AI tools to integrating AI into real workflows, Agentic AI represents an important next step: systems that can interpret goals, plan across multiple steps, use tools, and support task completion across digital and, increasingly, physical environments.

For the Silicon Valley Leadership Group and our members, this conversation sits at the center of a broader effort to advance innovation, economic competitiveness, and public benefit. The region's technology ecosystem is helping shape how AI is developed, deployed, and governed, while employers across sectors are exploring how these tools can improve productivity, strengthen services, support workers, and open new opportunities for growth. Agentic AI is part of that broader transition from AI as a content-generation tool to AI as an enabling layer for business operations, workforce transformation, research, infrastructure, and more.

Agentic AI should be understood not simply as a more advanced chatbot, but as an action-oriented system that combines models, tools, memory, planning, and oversight to complete tasks across workflows. Agentic AI can interpret a goal, determine what steps are needed, call external systems, review results, and continue working until a task is complete or a checkpoint is reached.

This glossary is intended to support a shared vocabulary for thoughtful discussion among business leaders, policymakers, technical experts, legal and compliance teams, workforce partners, and civic stakeholders. Clear terminology can help stakeholders distinguish between different levels of autonomy, identify promising use cases, align expectations, and design appropriate oversight. By clarifying how agentic AI works and where it may be most useful, this glossary aims to inform practical conversations about how to leverage this technology effectively in service of economic growth, competitiveness, and broad public benefit.

Acknowledgements

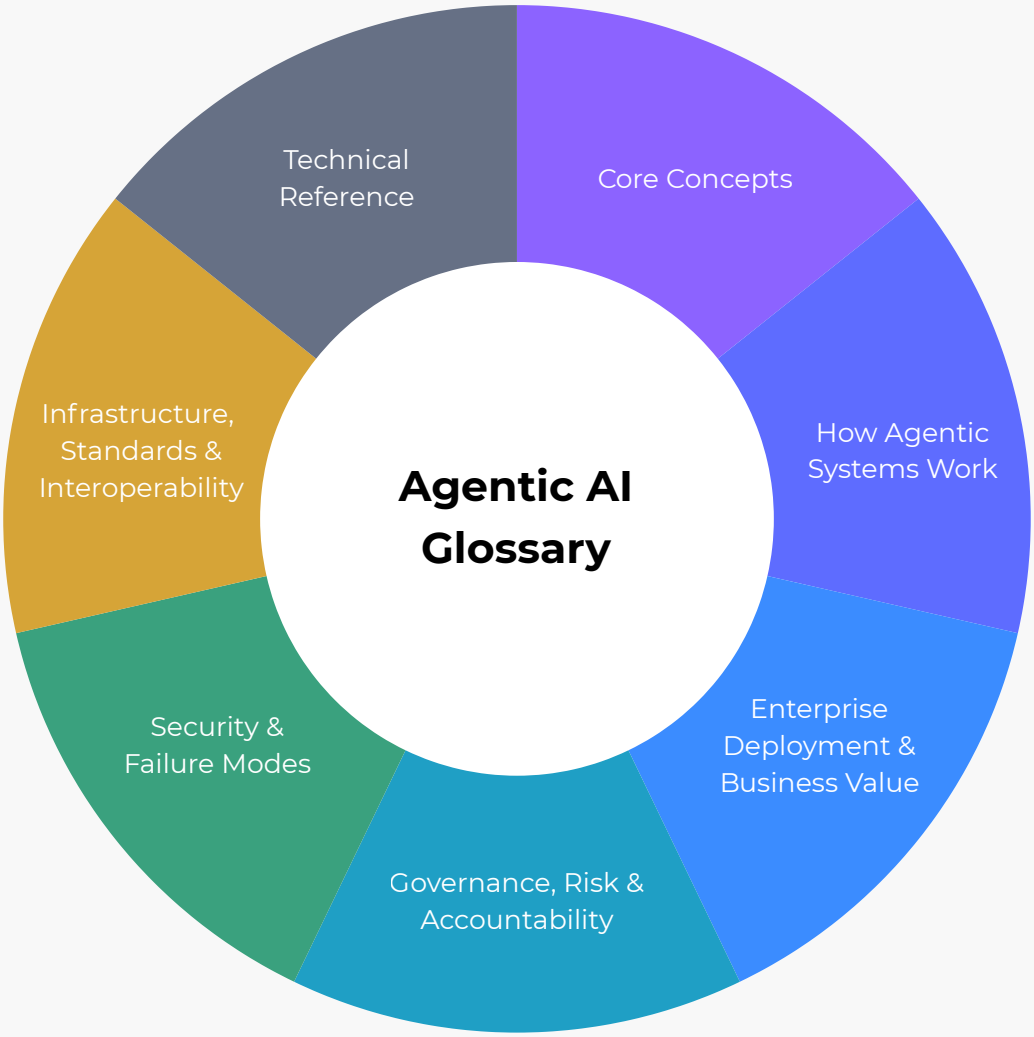
We gratefully acknowledge Lockheed Martin for their valuable contribution to this effort to foster a deeper understanding of AI within the business community. Their insightful analysis of AI's diverse impacts was instrumental in the creation of this document.

SVLG also extends its appreciation to the member companies and representatives who participated in the Agentic AI Task Force and small group workshops. Their thoughtful engagement, practical insights, and cross-sector perspectives helped shape this glossary as a resource to foster a deeper understanding of agentic AI within the business and policy community.

Special thanks are also extended to the SVLG communications team, Laura and Johnnie, for their guidance and support in bringing this document to publication. Their editorial insight and communications expertise helped ensure this resource is clear, accessible, and useful for a broad business and policy audience.

Source Orientation

This draft synthesizes terminology from enterprise AI explainers, cloud-provider documentation, AI risk-management frameworks, security guidance, emerging agent interoperability standards, and the existing April 2026 working draft. It is written for a mixed audience of executives, policymakers, legal and compliance teams, government affairs staff, technical practitioners, and external partners.



How to Use The Glossary

Core Concepts

Building a shared understanding of key agentic AI terms and concepts.

How Agentic Systems Work

Explaining how agents operate without requiring technical expertise.

Enterprise Deployment & Business Value

Evaluating use cases, understanding organizational impact, and planning enterprise adoption of agentic systems.

Governance, Risk & Accountability

Developing policies, evaluating vendors, and promoting a sustainable AI ecosystem.

Security & Failure Modes

Understanding risks, safeguards, and common agent failure scenarios.

Infrastructure, Standards & Interoperability

Preparing for implementation discussions with technical teams.

Technical Reference

Supporting deeper conversations with engineering, security, product, and standards experts.

Term Category Index

This glossary contains 70 terms organized across 7 business and policy-facing categories. The categories are ordered to help readers move from baseline concepts to deployment, governance, security, infrastructure, and technical reference material.

Core Concepts

How Agentic Systems Work

Enterprise Deployment & Business Value

Customer Service Agent

Domain-Specific Agent

Enterprise Integration

Human Augmentation

Incident Response Agent

Research Agent

Software / Coding Agent

Workflow Automation

Governance, Risk & Accountability

Agent Accountability

Assurance / Assurance Case

Audit Log

Certification

Grounding

Guardrails

High-Stakes Use Case

Human-in-the-Loop (HITL)

Human-on-the-Loop (HOTL)

Human Override

Operational Design Domain (ODD)

Permissioning

Reversibility

Traceability

Verification & Validation (V&V)

Security & Failure Modes

Blast Radius

Cascading Failure

Data Leakage

Data Poisoning / Adversarial Inputs

Excessive Agency

Hallucination

Memory Poisoning

Prompt Injection

Sandbox / Sandbox Environment

Tool Misuse

Infrastructure, Standards & Interoperability

Agent2Agent Protocol (A2A)

AgentOps

Agent Identity

Agent Interoperability

Agent Observability

API

Benchmark

Model Context Protocol (MCP)

Telemetry

Technical Reference

Context Window

Determinism / Reproducibility

Embedding

Fine-Tuning

Inference

Latency

Model Card / System Card

Token

Vector Database

Core Concept

These terms establish the baseline vocabulary for distinguishing agentic AI from generative AI, assistants, and traditional software automation.

Agentic AI
Strategy / Literacy

An AI system or architecture designed to pursue a goal, make decisions, use tools, and complete multi-step tasks with varying degrees of human oversight. For business and policy audiences, the key distinction is that agentic AI does not merely generate content; it can take action across workflows, systems, and organizational processes.

AI Agent
Strategy / Literacy

A software-based AI system that observes information, reasons about a task, and takes one or more actions to achieve a goal. An agent typically combines a model, instructions, tools, memory, permissions, and an execution loop.

AI Assistant / General Copilot
Strategy / Literacy

A user-facing AI tool that helps a person complete tasks but usually waits for user direction at each step. Assistants and copilots may become agentic when they can independently plan steps, call tools, and carry out tasks under defined permissions.

Autonomy Spectrum
Strategy / Literacy

A framework for describing how much independence an AI system has, ranging from manual human control to supervised autonomy to highly autonomous operation. Most enterprise deployments sit in the middle of the spectrum, with autonomy calibrated to task risk, reversibility, and business impact.

Digital Worker
Strategy / Literacy

A business metaphor for an AI agent configured to perform knowledge-work tasks such as research, analysis, drafting, scheduling, customer support, or data processing. The term is useful for enterprise strategy, but it should not obscure the need for accountability, supervision, and clear limits on authority.

Foundation Model
Strategy / Literacy

A large AI model trained on broad datasets and adaptable to many downstream tasks. Foundation models may be language models, vision models, audio models, or multimodal models that combine text, images, audio, video, sensor data, or other inputs. In agentic systems, foundation models often serve as the reasoning, generation, or perception layer that enables agents to interpret information, plan actions, and interact with digital or physical environments.

Generative AI
Strategy / Literacy

AI systems that create text, images, code, audio, video, or other content in response to prompts. Generative AI is often the foundation for agentic AI, but an agentic system adds planning, tool use, workflow execution, and oversight mechanisms.

Large Language Model (LLM)
Strategy / Literacy

A type of foundation model trained to process and generate language. In most current agentic AI systems, the LLM acts as the reasoning interface that interprets instructions, plans actions, and communicates with humans or software tools.

How Agentic Systems Work

These terms explain the operating model: how agents interpret goals, gather information, plan actions, use tools, coordinate with other agents, and continue through multi-step workflows.

Agentic Loop
Architecture / Operations

The repeated cycle through which an agent makes progress on a task: observe information, reason or plan, take an action, review the result, and decide what to do next. This loop is what makes agents more adaptive than one-shot chatbot interactions.

Memory
Architecture / Operations

A software-based AI system that observes information, reasonsThe mechanisms that allow an agent to retain or retrieve information relevant to a task. Memory may exist within the current prompt, in an external database, or in persistent storage across sessions; each form raises different privacy, accuracy, and governance considerations.
about a task, and takes one or more actions to achieve a goal. An agent typically combines a model, instructions, tools, memory, permissions, and an execution loop.

Multi-Agent System
Architecture / Operations

An architecture in which multiple agents collaborate, delegate tasks, or exchange outputs to complete a larger objective. Multi-agent systems can scale complex work, but they can also introduce cascading failures if one agent passes along incorrect information.

Orchestrator / Supervisor Agent
Architecture / Operations

A top-level agent or software component that decomposes work, assigns tasks to sub-agents or tools, monitors progress, handles errors, and synthesizes outputs. In policy terms, the orchestrator is often the control point where oversight, logging, and escalation should be designed.

Perceive-Reason-Act-Learn Planning

Architecture / Operations

A plain-language framework for describing agentic behavior. The agent gathers relevant information, interprets it against a goal, acts through tools or outputs, and uses feedback to adjust the next step or future workflow. In this context, “learn” usually refers to in-workflow adaptation, such as updating a plan, incorporating feedback, or storing relevant context in memory. It does not necessarily mean that the underlying AI model is being retrained.

Retrieval-Augmented Generation (RAG)

Architecture / Operations

A method that retrieves relevant documents or data from an external knowledge base and provides them to the model as context. RAG is important for enterprise agents because it helps them use current, proprietary, or domain-specific information rather than relying only on training data.

Sub-agent

Architecture / Operations

A specialized agent that performs a narrow function within a broader workflow, such as search, coding, summarization, financial analysis, or document review. Sub-agents can improve specialization but also create coordination and accountability challenges.

Task Decomposition

Architecture / Operations

The process of breaking a complex objective into smaller subtasks that can be completed, delegated, or checked separately. This is central to long-horizon workflows such as software migration, compliance review, incident response, or customer case resolution.

Tool / Plugin
Architecture / Operations

A specific callable capability made available to an agent, such as a calculator, CRM connector, code interpreter, document retrieval system, or ticketing-system integration. Tools should be permissioned and monitored because they determine what the agent can actually do.

Tool Use / Function Calling
Architecture / Operations

The ability of an agent to invoke external capabilities such as APIs, databases, web search, code execution, calendars, file systems, or enterprise software. Tool use turns an AI system from a content generator into an action-taking system.

Enterprise Deployment & Business Value

These terms explain the operating model: how agents interpret goals, gather information, plan actions, use tools, coordinate with other agents, and continue through multi-step workflows.

Customer Service Agent
Use Cases / ROI

An agent designed to answer customer questions, retrieve relevant information, update records, route cases, or escalate unresolved issues to a human. These systems can improve response time and coverage, but they require escalation rules, accuracy controls, and customer transparency.

Domain-Specific Agent
Use Cases / ROI

An agent designed for a particular function, sector, or body of knowledge, such as healthcare intake, software modernization, legal review, finance, procurement, customer service, or other regulated and mission-critical sectors. Domain specificity can improve performance but requires stronger evaluation against sector-specific risks, rules, and operational requirements.

Enterprise Integration
Use Cases / ROI

The connection of agents to business systems such as CRM, ERP, HR, cybersecurity, legal, productivity, cloud, or data platforms. Integration is where agentic AI creates business value, but it is also where access control, vendor risk, and auditability become critical.

Human Augmentation
Use Cases / ROI

The use of AI agents to support human workers rather than replace them outright. For workforce and policy discussions, this framing emphasizes productivity, reskilling, oversight, and new job design rather than simple labor substitution.

Incident Response Agent
Use Cases / ROI

An agent used to triage, investigate, report, or help remediate operational or cybersecurity incidents. These agents can reduce response time, but organizations should define when actions require human approval, especially for rollback, access changes, or external notifications.

Research Agent
Use Cases / ROI

An agent that gathers information, reviews documents, synthesizes findings, and prepares summaries or recommendations. Research agents are useful for policy, legal, market, and technical analysis, but outputs should be grounded in verifiable sources.

Software / Coding Agent
Use Cases / ROI

An agent that can write, test, debug, refactor, or document code, sometimes across multi-step development workflows. Coding agents create productivity opportunities but require review for security, quality, intellectual property, and change-management risk.

Workflow Automation
Use Cases / ROI

Software-driven execution of a business process with limited human involvement. Agentic AI extends traditional automation by allowing the system to adapt to new information, handle exceptions, and determine next steps rather than only following fixed scripts.

Governance, Risk & Accountability

These terms support conversations about oversight, responsibility, human review, documentation, and acceptable use.

Agent Accountability
Oversight / Compliance

The process of assigning responsibility for an agent's design, deployment, permissions, outputs, and actions. Accountability may involve developers, deployers, vendors, system integrators, human supervisors, and the organization operating the system.

Assurance / Assurance Case
Oversight / Compliance

A structured, evidence-based argument that a system is acceptably safe and secure for its intended use, supported by verification, validation, testing, and documentation. It connects individual controls, such as guardrails, logging, and human oversight, into a defensible whole. High-stakes sectors increasingly treat a documented assurance case, rather than monitoring alone, as the basis for approving an agent to operate.

Audit Log
Oversight / Compliance

A record of an agent's actions, tool calls, data access, decisions, approvals, and outputs. Audit logs should be designed so organizations can investigate incidents, demonstrate compliance, and understand system behavior over time.

Certification
Oversight / Compliance

Formal attestation by a recognized authority that a system meets defined safety, performance, or regulatory standards and is approved for use (for example, the FAA in aviation). Agentic AI raises two distinct challenges: certifying software produced by AI (such as code from a coding agent), where existing processes assume human-traceable rationale and deterministic review; and certifying AI that operates autonomously in safety- or mission-critical roles, where no mature path yet exists for non-deterministic, adaptive systems.

Grounding
Oversight / Compliance

The practice of connecting an agent's output to verifiable information, such as cited sources, enterprise records, live databases, or approved knowledge bases. Grounding reduces hallucination risk and supports trust, review, and compliance.

Guardrails
Oversight / Compliance

Technical, procedural, or policy constraints that limit what an agent can do. Guardrails may include system instructions, content filters, access controls, allowlists, blocklists, rate limits, human approvals, and escalation thresholds.

High-Stakes Use Case
Oversight / Compliance

A use case where agent actions or outputs can materially affect safety, rights, financial outcomes, employment, healthcare, legal obligations, critical infrastructure, national security, or public trust. High-stakes use cases generally require stronger review, testing, documentation, and human oversight.

Human-in-the-Loop (HITL)
Oversight / Compliance

A design approach in which a human must review, approve, or guide an agent before it takes certain actions. HITL is appropriate for high-stakes, safety-critical, irreversible, sensitive, or externally visible decisions.

Human-on-the-Loop (HOTL)
Oversight / Compliance

An oversight model in which an agent can operate autonomously while a human monitors progress and can intervene if needed. HOTL is useful for lower-risk or time-sensitive workflows, but it depends on effective alerts, dashboards, and intervention authority.

Human Override
Oversight / Compliance

A mechanism that allows a human operator to pause, stop, redirect, or reverse an agent's actions. Override capability is a practical governance requirement for agentic systems that interact with customers, money, infrastructure, legal obligations, or sensitive data.

Operational Design Domain (ODD)
Oversight / Compliance

The defined set of conditions under which a system is validated and approved to operate, including environment, inputs, and task scope. Performance and safety are only assured inside the ODD; behavior outside it is unspecified and a primary source of risk.

Permissioning
Oversight / Compliance

The process of defining what data, tools, systems, and actions an agent is allowed to access or execute. Strong permissioning limits the agent to the minimum access necessary for its task and reduces the potential impact of error or misuse.

Reversibility
Oversight / Compliance

The degree to which an agent's action can be undone if it is incorrect, harmful, or unauthorized. Reversible actions can often tolerate more automation; irreversible or hard-to-reverse actions should usually require stronger human approval.

Traceability
Oversight / Compliance

The ability to reconstruct what an agent did, what information it used, which tools it called, what decisions it made, and where human approvals occurred. Traceability is essential for debugging, audit, compliance, and liability analysis.

Verification & Validation (V&V)

Oversight / Compliance

Two complementary checks: verification confirms a system is built correctly and meets its specified requirements (“did we build the system right”); validation confirms it meets the real-world need (“did we build the right system”). Because agentic AI is non-deterministic and operates over an open-ended action space, traditional test-every-path approaches do not fully transfer.

Security & Failure Modes

These terms describe the risks that become more important when AI systems can retrieve information, call tools, retain memory, or take action.

Blast Radius
Security / Resilience

The scope of potential damage if an agent makes a mistake, is compromised, or behaves unexpectedly. Organizations reduce blast radius through limited permissions, sandboxing, reversible actions, rate limits, and human approval gates.

Cascading Failure
Security / Resilience

A failure pattern in which one agent's error spreads through a workflow, tool chain, or multi-agent system and causes broader harm. Cascading failures are a major reason to design checkpoints, logs, and bounded authority into agentic systems.

Data Leakage
Security / Resilience

The unintended exposure of sensitive, confidential, proprietary, or personal information through an agent's outputs, tool calls, memory, logs, or integrations. Data leakage risk increases when agents connect to multiple internal systems or external platforms.

Data Poisoning / Adversarial Inputs
Security / Resilience

Attempts to manipulate an AI system by corrupting training data, retrieval sources, user inputs, or external content. For agents, adversarial inputs can influence both what the system says and what it does.

Excessive Agency
Security / Resilience

A risk condition in which an agent is given more autonomy, permissions, tools, or scope than it needs to complete its task. Excessive agency increases the likelihood and severity of unintended actions.

Hallucination
Security / Resilience

A model output that sounds plausible but is false, unsupported, or fabricated. In agentic systems, hallucinations can cause downstream harm if the agent uses inaccurate information to take action, update records, or instruct other agents.

Memory Poisoning
Security / Resilience

A failure or attack pattern in which inaccurate, malicious, or irrelevant information is stored in an agent's memory and later influences future behavior. Persistent memory should therefore be curated, permissioned, and reviewable.

Prompt Injection
Security / Resilience

An attack in which malicious or conflicting instructions are inserted into content the agent reads, such as a webpage, email, document, or database entry. Prompt injection is especially serious for agents because the system may act on the malicious instruction through connected tools.

Sandbox / Sandbox Environment
Security / Resilience

An isolated computing environment where an agent can execute code, test actions, or run experiments without affecting production systems. Sandboxing is a core safety measure for experimentation, code execution, and high-risk tool use.

Tool Misuse
Security / Resilience

An agent's incorrect, unauthorized, unsafe, or poorly timed use of an external tool. Tool misuse can occur because of bad instructions, faulty reasoning, malicious input, insufficient permissions design, or weak monitoring.

Infrastructure, Standards & Interoperability

These terms describe the technical and standards layer that will shape how agents connect to tools, data, enterprise systems, and other agents.

Agent2Agent Protocol (A2A)

Standards / Implementation

An open protocol originally announced by Google and later placed under Linux Foundation governance to support secure communication and coordination among agents. A2A is relevant to enterprise adoption because agents from different vendors may need to discover capabilities, exchange information, and coordinate work.

AgentOps

Standards / Implementation

The operational discipline of deploying, monitoring, evaluating, securing, and improving AI agents in production. It extends MLOps and DevOps concepts to the specific challenges of goal-directed, tool-using systems.

Agent Identity

Standards / Implementation

A unique identity assigned to an agent so its permissions, actions, ownership, and logs can be managed separately from human users or other systems. Agent identity is important for access control, auditability, and incident response.

Agent Interoperability

Standards / Implementation

The ability of agents, tools, data systems, and platforms to work together across vendors or technical environments. Interoperability can reduce vendor lock-in and support enterprise-scale deployment, but it also requires common standards for identity, permissions, security, and logging.

Agent Observability

Standards / Implementation

The ability to monitor an agent's behavior, tool calls, failures, costs, latency, and escalation points. Observability helps organizations detect drift, stuck workflows, unsafe behavior, and performance issues in production.

API
Standards / Implementation

An application programming interface: a structured way for software systems to communicate. APIs are the main pathway through which agents retrieve information, call tools, and take actions in enterprise systems.

Benchmark
Standards / Implementation

A standardized test or evaluation dataset used to measure model or agent performance. For agents, benchmarks may test reasoning, tool use, task completion, reliability, latency, groundedness, and ability to recover from errors.

Model Context Protocol (MCP)
Standards / Implementation

An open standard introduced by Anthropic for connecting AI systems to external tools and data sources in a more consistent way. For policy and business audiences, MCP matters because it points toward a more interoperable agent-tool ecosystem.

Telemetry
Standards / Implementation

Operational data generated by an agent or agentic system, such as tool-call frequency, success rates, error rates, latency, resource use, and intervention events. Telemetry supports monitoring, evaluation, governance, and continuous improvement.

Technical Reference

These terms are useful for readers who need to understand cost, performance, model behavior, and supporting technical infrastructure.

Context Window
Technical Fluency

The amount of information a model can process in a single interaction. Agents must manage context carefully because long workflows can exceed what the model can reliably review at once.

Determinism / Reproducibility
Technical Fluency

The extent to which a system produces the same output given the same input and conditions. Traditional safety-critical software is largely deterministic and therefore testable against fixed expectations; agentic AI is typically non-deterministic, a central reason its behavior is harder to verify, certify, and audit.

Embedding
Technical Fluency

A numerical representation of text, images, or other data that captures semantic meaning. Embeddings let systems compare, group, classify, recommend, and retrieve related information even when exact keywords do not match. In agentic systems, retrieval is one of the most common uses because embeddings help agents find relevant context from large knowledge stores.

Fine-Tuning
Technical Fluency

The process of further training a model on specialized data to improve its performance for a specific task, domain, style, or output format. Fine-tuning is often used to specialize model behavior, improve task and domain performance, or make outputs more consistent with organizational needs. When the goal is to connect an agent to current, proprietary, or frequently changing information, retrieval-augmented generation is often used alongside or instead of fine-tuning. Fine-tuning can also raise governance questions about data quality, evaluation, bias, version control, and model maintenance.

Inference
Technical Fluency

The process of running a trained AI model to generate an output from a given input. In agentic workflows, inference happens repeatedly as the agent plans, observes, calls tools, and revises its next step.

Latency
Technical Fluency

The time between a user request or agent action and the resulting response or completed step. In business settings, latency affects user experience, operational feasibility, and cost.

Model Card / System Card
Technical Fluency

Documentation that explains a model or AI system's intended use, limitations, evaluation results, risks, and responsible deployment considerations. For agentic systems, documentation should cover not only the model but also tools, permissions, human oversight, and operating conditions.

Token
Technical Fluency

A unit of text processed by a language model, often a word or word fragment. Token usage affects context limits, cost, speed, and the feasibility of long-running workflows.

Vector Database
Technical Fluency

A database designed to store and search embeddings. Vector databases are commonly used in RAG systems to help agents retrieve relevant enterprise or domain-specific knowledge.

Defintion Notes

The following terminology choices are used consistently throughout this glossary to improve clarity and reduce ambiguity.

Agentic AI vs. AI Agents

Agentic AI refers to the broader category of systems.

AI Agent refers to an individual system or component.

Key Distinction

Agentic AI describes the ecosystem.

AI Agent describes an individual actor within it.

Autonomy vs. Agency

Agency refers to a system's ability to pursue goals, make decisions, and take actions.

Autonomy refers to the degree of independence a system has from human oversight when carrying out those actions.

Key Distinction

Agency is about capability.

Autonomy is about independence.

Chain-of-Thought

Earlier versions of this glossary included Chain-of-Thought as an agent architecture concept. This edition omits it because internal model reasoning is not equivalent to an auditable decision process.

For governance, compliance, and accountability discussions, observable artifacts—such as tool-call logs, decision records, and execution traces—provide more practical forms of transparency.

Key Distinction

Reasoning processes are not the same as governance records.

Learning

Agentic systems are often described as "learning from feedback," but that phrase can refer to several different processes.

These may include:

- Workflow Adaptation — modifying plans or actions based on outcomes
- Memory Updates — retaining relevant context for future interactions
- Model Training or Fine-Tuning — changing the underlying model itself

Because these processes operate differently, they also introduce different oversight, governance, and risk considerations.

Key Distinction

Not all learning changes the model. Adaptation, memory, and retraining are distinct processes.

Digital Worker

The term Digital Worker is retained as an enterprise-facing metaphor because it can help communicate business value and organizational impact.

However, the term should be used carefully. While it may describe how work is performed, it should not imply that accountability, responsibility, or decision authority has been delegated to the AI system itself.

Key Distinction

A digital worker can perform tasks, but accountability remains with people and organizations.