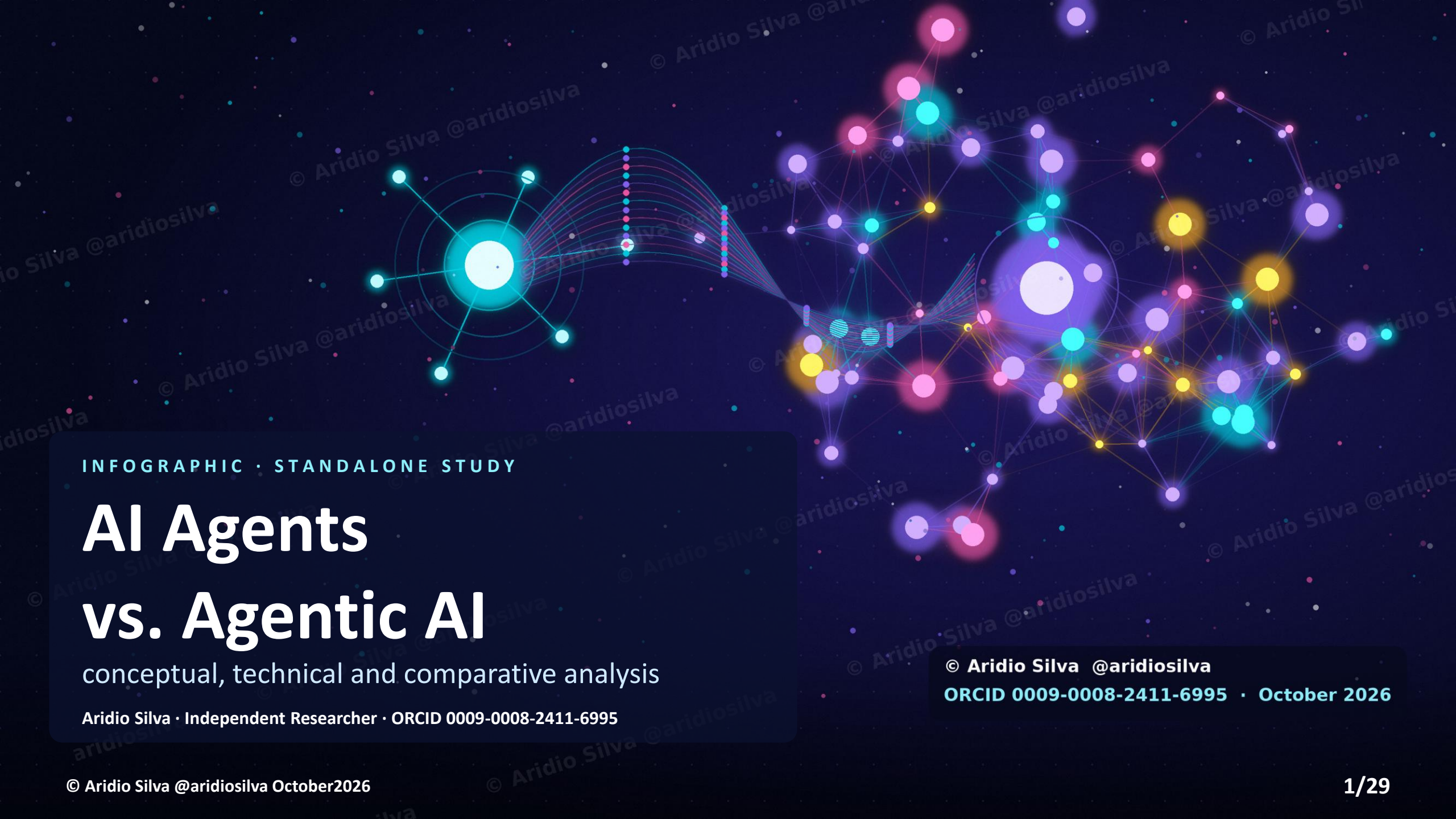




INFOGRAPHIC · STANDALONE STUDY

# AI Agents vs. Agentic AI

conceptual, technical and comparative analysis

Aridio Silva · Independent Researcher · ORCID 0009-0008-2411-6995

© Aridio Silva @aridiosilva

ORCID 0009-0008-2411-6995 · October 2026

# How the sources were verified



**Page or PDF opened and checked**

Anthropic (2024) · OpenAI (2026) · Shavit et al. (2023) · Sapkota et al. (2025, abstract and DOI)



**Official text read in search results**

Wooldridge and Jennings (1995) · Anthropic (n.d.) · Cemri et al. (2025) · OpenAI (n.d.) · Linux Foundation · Google Cloud · Microsoft · OWASP · Yao et al.



**Secondary source only †**

Google "Agents" whitepaper (summaries) · complementary roles of MCP and A2A (Heise) · OWASP ASI codes (third parties)



## IMPORTANCE

Each source gets a verification level: A, B or C.



## POSITIVE IMPLICATIONS

Readers see how well each claim is supported.



## NEGATIVE IMPLICATIONS

Level C (†) limits certainty: Google, MCP × A2A and ASI codes.

# No consensus: five sources, five uses



| Source                         | How it defines                                                                     | Nature       |
|--------------------------------|------------------------------------------------------------------------------------|--------------|
| Wooldridge and Jennings (1995) | Autonomy, social ability, reactivity and pro-activeness                            | Properties   |
| Anthropic (2024)               | "Agentic" is the genus: workflows follow predefined code; agents let the LLM drive | Architecture |
| OpenAI (2026)                  | A system that independently accomplishes tasks on the user's behalf                | Functional   |
| Shavit et al. (2023)           | Degree of adaptable autonomy for complex goals; a property, not a category         | Spectrum     |
| Sapkota et al. (2025)          | Agent (module) vs. agentic AI (coordinated multi-agent)                            | Taxonomy     |



## IMPORTANCE

Only one of the five sources separates "AI agent" from "agentic AI" as categories.



## POSITIVE IMPLICATIONS

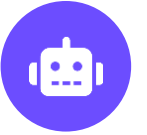
Knowing the definitions lets you pick the best fit for each context.



## NEGATIVE IMPLICATIONS

Without a standard, the same label can hide very different products.

# AI agent: model, tools, rules



A system that independently accomplishes tasks on the user's behalf (OpenAI, 2026). A simple chatbot, a single-turn LLM and a classifier are not agents: they do not control the workflow.



## Model

The LLM that reasons and decides



## Tools

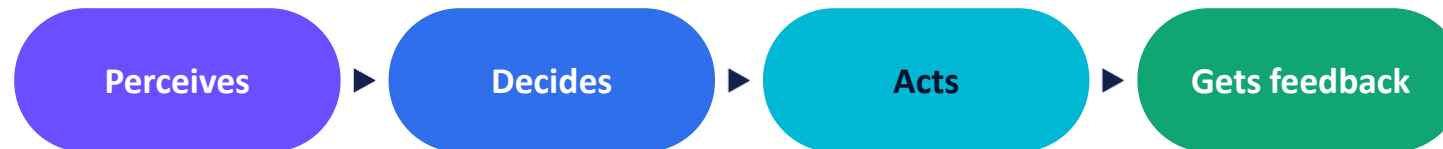
Data, actions and even other agents



## Instructions

Rules, routines and guardrails

## The agent loop



Until a stop condition: goal reached or iteration limit (Anthropic, 2024).



## IMPORTANCE

The basic artifact: what you build, test and govern.



## POSITIVE IMPLICATIONS

Automates workflows with ambiguous decisions and unstructured data.



## NEGATIVE IMPLICATIONS

Higher cost and latency and a risk of compounding errors (Anthropic, 2024).

# Agentic AI: a degree, not a box



1

## Goal complexity

How hard and varied the goals the system can achieve are

2

## Environmental complexity

Many domains, actors, long horizons and tools

3

## Adaptability

Reacts to novel or unexpected circumstances

4

## Independent execution

Reaches the goal with little human supervision

*"Agenticness" is a property, not a category. It does not imply consciousness: goals are still set by humans (Shavit et al., 2023).*



### IMPORTANCE

"Agentic" is gradual: there is no binary line between an agent and a current system.



### POSITIVE IMPLICATIONS

Better quality, time savings and scale, if well designed.



### NEGATIVE IMPLICATIONS

Raises failures, vulnerabilities and abuse; evaluation is still a nascent field.

# Three readings of the difference



1

## Taxonomic

Sapkota et al. (2025)

Distinct categories: agent = module for a specific task; agentic AI = coordinated specialized agents with persistent memory.

2

## Gradual

Shavit et al. (2023)

"Agentic" is a degree of adaptable autonomy. There is no binary line between agents and current systems.

3

## Architectural

Anthropic (2024)

"Agentic system" is the genus; workflows and agents are species, depending on who controls the path.

**Synthesis (author's analysis):** "agent" is the artifact; "agentic" is the property or paradigm. Always state the definition you adopt.



## IMPORTANCE

Avoids debating the same term with different meanings.



## POSITIVE IMPLICATIONS

A shared vocabulary speeds up projects, purchases and evaluations.



## NEGATIVE IMPLICATIONS

Single agent × multi-agent is one author's proposal, not a standard.

# Sapkota et al. (2025) taxonomy



| Aspect       | AI agent                                                                                               | Agentic AI                                                             |
|--------------|--------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|
| Nature       | Modular, LLM-driven system for task-specific automation                                                | Multi-agent collaboration paradigm                                     |
| Mechanisms   | Tool integration, prompt engineering and reasoning                                                     | Dynamic task decomposition, persistent memory and coordinated autonomy |
| Applications | Customer support, scheduling and data summarization                                                    | Research automation, robot coordination and medical decision support   |
| Challenges   | Hallucination, brittleness, emergent behavior and coordination failure, examined across both paradigms |                                                                        |
| Proposals    | ReAct loops, RAG, coordination layers and causal modeling                                              |                                                                        |



## IMPORTANCE

The only taxonomy examined that treats the terms as distinct categories.



## POSITIVE IMPLICATIONS

Helps choose between modular automation and multi-agent coordination.



## NEGATIVE IMPLICATIONS

It is a conceptual review (preprint and Information Fusion), not an experiment.

# From chatbot to agentic AI: a spectrum



*More autonomy, cost and complexity → Order and grouping: author's synthesis, not an official scale from any source.*



## IMPORTANCE

Helps choose the right level of autonomy for each problem.



## POSITIVE IMPLICATIONS

Starting simple cuts cost and eases evaluation.



## NEGATIVE IMPLICATIONS

Each step up raises cost ( $\approx 15\times$  tokens in multi-agent) and risk.

# Timeline (1/3): 2022–2024



**2022 · Oct**

ReAct (Yao et al.): language models interleave reasoning and action.

**2024 · Sep**

Google publishes the "Agents" whitepaper †.

**2023**

Shavit et al. (OpenAI): "agenticness" and seven governance practices.

**2024 · 19 Dec**

Anthropic publishes "Building effective agents": workflows × agents.



## IMPORTANCE

The vocabulary and patterns are born: LLM agents, agenticness, workflow × agent.



## POSITIVE IMPLICATIONS

An open conceptual base for academia and companies.



## NEGATIVE IMPLICATIONS

Terms still unstable; Google's milestone comes from a secondary source (†).

# Timeline (2/3): 2025



**2025 · Mar**

MAST: 14 failure modes in multi-agent systems (Cemri et al.).

**2025 · Jun**

Google donates the A2A protocol to the Linux Foundation.

**2025 · 9 Dec**

OWASP lists 10 agentic risks; the Agentic AI Foundation is born.

**2025 · 15 May**

Sapkota et al. propose the agent × agentic AI taxonomy (preprint).

**2025 · 1 Oct**

Microsoft merges AutoGen and Semantic Kernel into Agent Framework.



## IMPORTANCE

The debate turns into taxonomy, open standards and a risk list.



## POSITIVE IMPLICATIONS

Open protocols (A2A, MCP) and a risk list for agents.



## NEGATIVE IMPLICATIONS

Documented multi-agent failures and agent-specific risks.

# Timeline (3/3): 2026



**2026 · 15 Jan**

OpenAI: practical guide to agents (date shown on the page).

**2026 · 10 Aug**

Anthropic updates "Building effective agents": the 2024 tooling has changed.

**2026 · 3 Jun**

OpenAI shuts down Agent Builder and Evals.



## IMPORTANCE

Focus shifts from defining to operating: practical guides and shifting platforms.



## POSITIVE IMPLICATIONS

Practical guides and documentation updated by vendors.



## NEGATIVE IMPLICATIONS

Discontinued products and fast-changing tooling.

# Anthropic: simplicity and patterns



Workflows follow predefined paths; agents let the LLM drive the process. Five workflow patterns:

Prompt chaining

Routing

Parallelization

Orchestrator-workers

Evaluator-optimizer

## Three principles

- **Simplicity:** use the simplest solution possible.
- **Transparency:** show the planning steps.
- **Agent-computer interface:** document and test the tools.

≈15×

the tokens of a chat in the multi-agent research system (internal evaluation)



### IMPORTANCE

Defines workflow × agent and recommends the simplest possible solution.



### POSITIVE IMPLICATIONS

Composable patterns and open MCP, donated to the Linux Foundation.



### NEGATIVE IMPLICATIONS

Agents cost more and compound errors; the company warns the 2024 tooling has changed.

# OpenAI: minimal parts and guardrails



Three components: model, tools and instructions. Start with a single agent and split only when needed.

## Manager pattern

A central agent calls specialists as tools.

## Decentralized pattern

Peer agents hand execution off to one another.

## Layered guardrails

Relevance

Safety

Personal data

Moderation

Tools

Rules

Output

Human intervention when failures pass a threshold or the action is high-risk.



### IMPORTANCE

Combines a practical build guide and a governance white paper.



### POSITIVE IMPLICATIONS

Layered guardrails and well-defined human intervention.



### NEGATIVE IMPLICATIONS

Unstable platform: Agent Builder and Evals shut down on 3 Jun 2026.

# Google: architecture and interoperability



**Tools** APIs, data and real-world actions

**Orchestration** Planning, memory and reasoning in a loop

**Model** The generative model that decides

*Three layers per the "Agents" whitepaper (Wiesinger, Marlow and Vuskovic, 2024) †*

Complementary to MCP, per the trade press (Heise, 2025) †.

## A2A

Agent2Agent protocol for agent-to-agent communication. Announced in April 2025, donated to the Linux Foundation in June. ADK has native support.



### IMPORTANCE

Emphasis on layered architecture and interoperability between agents.



### POSITIVE IMPLICATIONS

A2A is an open protocol under the Linux Foundation.



### NEGATIVE IMPLICATIONS

Whitepaper read only through summaries (†); identity governance not verified.

# Comparison (1/2): concept and architecture



| Dimension    | Anthropic                                           | OpenAI                                                        | Google                                                                      |
|--------------|-----------------------------------------------------|---------------------------------------------------------------|-----------------------------------------------------------------------------|
| Core concept | "Agentic system" as the genus; workflows × agents   | Agent independently does tasks; "agenticness" as a degree     | Agent extends the model with reasoning and access to external information † |
| Components   | Augmented LLM: retrieval, tools and memory          | Model, tools and instructions                                 | Model, tools and orchestration layer †                                      |
| Multi-agent  | Orchestrator-workers; research ≈15× a chat's tokens | Manager and decentralized patterns; start with a single agent | A2A between agents; ADK with native A2A support                             |



## IMPORTANCE

All three converge on tools, an execution loop and human oversight.



## POSITIVE IMPLICATIONS

Simple, composable patterns lower project risk.



## NEGATIVE IMPLICATIONS

Different definitions make products hard to compare.

# Comparison (2/2): governance and standards



| Dimension               | Anthropic                                                  | OpenAI                                                             | Google                                           |
|-------------------------|------------------------------------------------------------|--------------------------------------------------------------------|--------------------------------------------------|
| Security and governance | Extensive testing in sandboxed environments and guardrails | Layered guardrails; human intervention; seven governance practices | Not verified in a primary source                 |
| Open standards          | MCP, donated to the Agentic AI Foundation                  | AGENTS.md †, donated to the same foundation                        | A2A, donated to the Linux Foundation in Jun 2025 |
| Volatility              | 2024 tooling has changed (company notice)                  | Agent Builder and Evals shut down on 3 Jun 2026                    | Not verified                                     |



## IMPORTANCE

Governance and open standards decide trust for production use.



## POSITIVE IMPLICATIONS

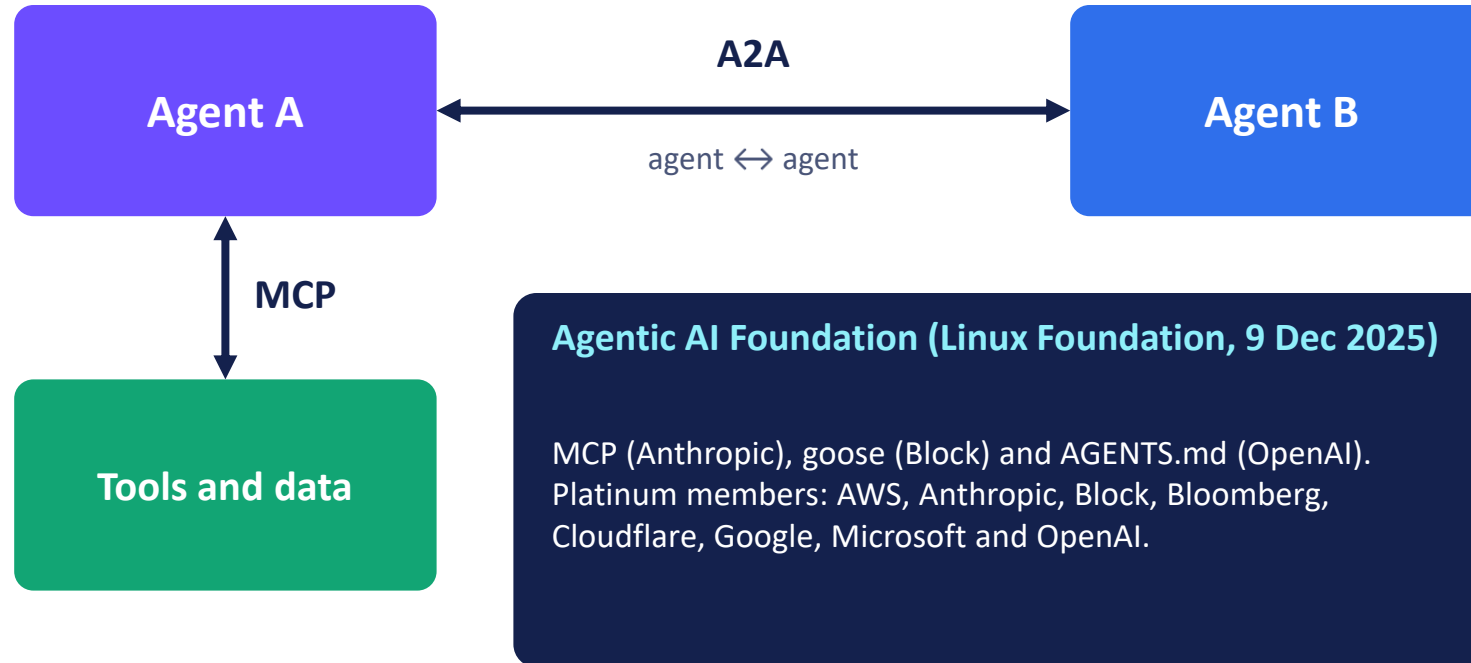
MCP and AGENTS.md at the AAIF; A2A also under the Linux Foundation.



## NEGATIVE IMPLICATIONS

Platforms change fast: tooling (2024) and products (2026).

# Open standards: MCP, A2A and AGENTS.md



## IMPORTANCE

Open standards reduce dependence on a single vendor.



## POSITIVE IMPLICATIONS

Interoperability between tools and between agents.



## NEGATIVE IMPLICATIONS

More connections widen the attack surface (author's analysis).

MCP links agents to tools; A2A links agents to each other, in complementary roles (Heise, 2025) †.

# Strengths and limits (1/2)



| Topic            | Strength or suitable use                                              | Limitation or risk                                                           | Source               |
|------------------|-----------------------------------------------------------------------|------------------------------------------------------------------------------|----------------------|
| Open-ended tasks | Problems with no predictable number of steps, in trusted environments | Higher cost and compounding errors; test in sandboxes                        | Anthropic (2024)     |
| Cost             | Better performance on broad research                                  | Agents use $\approx 4\times$ a chat's tokens; multi-agent $\approx 15\times$ | Anthropic (n.d.)     |
| Evaluation       | Testing subtasks in isolation helps                                   | Over long sequences, rare failures compound; evaluation is a nascent field   | Shavit et al. (2023) |
| Human oversight  | Requiring approval limits serious failures                            | With many fast approvals, review becomes a rubber stamp                      | Shavit et al. (2023) |
| AI monitor       | Operates at a speed and cost humans cannot match                      | Can double the cost; the same prompt injection can hijack the monitor        | Shavit et al. (2023) |



**Importance:** the same capabilities that make agents useful raise cost, risk and evaluation difficulty.

# Strengths and limits (2/2)



| Topic                | Strength or suitable use                         | Limitation or risk                                                          | Source                             |
|----------------------|--------------------------------------------------|-----------------------------------------------------------------------------|------------------------------------|
| Security             | A dedicated risk taxonomy, built by 100+ experts | Shavit et al.'s practices do not cover hijacking of agents by attacks       | OWASP (2025); Shavit et al. (2023) |
| Multi-agent          | Collaboration and task decomposition             | 14 failure modes in 3 categories; gains over a single agent are often small | Cemri et al. (2025)                |
| Technical challenges | Proposals: ReAct, RAG, coordination layers       | Hallucination, brittleness, emergent behavior, coordination failure         | Sapkota et al. (2025)              |



**Importance:** security and multi-agent failures are the risk fronts documented in the sources.

# Security: OWASP Top 10 for agents



Published on 9 Dec 2025 by the OWASP GenAI Security Project, with 100+ experts. Four of the ten categories:

**ASI01**

## Goal hijack

An attacker redirects the agent's goal, usually through text it reads

**ASI02**

## Tool misuse

The agent uses tools unsafely or beyond what was intended

**ASI06**

## Memory and context poisoning

Poisoned data distorts the agent's future decisions

**ASI10**

## Rogue agents

Goal drift, collusion or emergent behavior



### IMPORTANCE

Risk taxonomy for agents that plan, use tools and act under delegation.



### POSITIVE IMPLICATIONS

Gives security teams a shared risk language.



### NEGATIVE IMPLICATIONS

ASI codes and names confirmed only in third-party sources (†).

*Names and ASI codes confirmed in third-party sources †.*

# Multi-agent failures (MAST)



14

failure modes

## Specification

Poorly defined tasks and roles

## Inter-agent misalignment

Coordination and communication failures

## Task verification

Results not properly checked



### IMPORTANCE

Shows that adding agents does not guarantee a better result.



### POSITIVE IMPLICATIONS

The taxonomy lets teams diagnose failures by category.



### NEGATIVE IMPLICATIONS

Requires tests and oversight of specification, coordination and verification.

*Gains over single agents are often small on popular benchmarks (Cemri et al., 2025).*

# Governance: seven practices (Shavit et al.)



1

Assess suitability for the task

2

Constrain actions and require approval

3

Set default behaviors

4

Ensure activity legibility

5

Monitor automatically

6

Ensure attributability

7

Ensure interruptibility

Indirect impacts: adoption races, labor displacement, offense × defense balance and correlated failures.



## IMPORTANCE

Seven practices for safe agents with an identifiable responsible party.



## POSITIVE IMPLICATIONS

A basis for sharing responsibility among developer, deployer and user.



## NEGATIVE IMPLICATIONS

The authors say they are insufficient and do not cover hijacking by attacks.

# Glossary (1/3)



| Term           | Definition                                                                                               | Source               |
|----------------|----------------------------------------------------------------------------------------------------------|----------------------|
| AI agent       | A system with a model, tools and instructions that independently accomplishes tasks on the user's behalf | OpenAI (2026)        |
| Agentic AI     | Systems that pursue complex goals with limited direct supervision                                        | Shavit et al. (2023) |
| Agenticness    | Degree to which a system adaptably achieves complex goals in complex environments                        | Shavit et al. (2023) |
| Agentic system | Anthropic's umbrella term covering workflows and agents                                                  | Anthropic (2024)     |
| Workflow       | LLMs and tools orchestrated through predefined code paths                                                | Anthropic (2024)     |
| Augmented LLM  | An LLM with retrieval, tools and memory; the basic block of agentic systems                              | Anthropic (2024)     |

*† Verified only in a secondary source.*

# Glossary (2/3)



| Term                         | Definition                                                                             | Source               |
|------------------------------|----------------------------------------------------------------------------------------|----------------------|
| <b>Orchestrator-workers</b>  | A central LLM splits tasks dynamically, delegates to workers and synthesizes           | Anthropic (2024)     |
| <b>Manager pattern</b>       | A central agent coordinates specialized agents called as tools                         | OpenAI (2026)        |
| <b>Decentralized pattern</b> | Peer agents hand execution off to one another by specialty                             | OpenAI (2026)        |
| <b>Guardrails</b>            | Layered safeguards: relevance, safety, personal data, moderation and rules             | OpenAI (2026)        |
| <b>Human intervention</b>    | Control passes to a person when a failure limit is exceeded or the action is high-risk | OpenAI (2026)        |
| <b>Interruptibility</b>      | Ability to shut an agent down safely, including sub-agents it created                  | Shavit et al. (2023) |

*† Verified only in a secondary source.*

# Glossary (3/3)



| Term            | Definition                                                                       | Source                                    |
|-----------------|----------------------------------------------------------------------------------|-------------------------------------------|
| Attributability | Ability to trace an agent's actions to the responsible user                      | Shavit et al. (2023)                      |
| ReAct           | An approach that interleaves reasoning and action in language models             | Yao et al. (2022)                         |
| MCP             | Anthropic's protocol that integrates LLMs with an ecosystem of third-party tools | Anthropic (2024); Linux Foundation (2025) |
| A2A             | Protocol created by Google for agent-to-agent communication                      | Google Cloud (n.d.)                       |
| AGENTS.md †     | Per-project instruction format for coding agents                                 | Linux Foundation (2025)                   |
| MAST            | Taxonomy of 14 failure modes of multi-agent systems                              | Cemri et al. (2025)                       |

† Verified only in a secondary source.

# Limitations of this study



- The Google "Agents" whitepaper was not opened in the original; the description comes from third-party summaries (level C).
- The complementary role of MCP and A2A, the OWASP ASI codes and AGENTS.md come from secondary sources.
- Two official pages (Anthropic, n.d.; OpenAI, n.d.) and the Shavit et al. PDF carry no date in the excerpt read.
- Meta, xAI and other frontier developers were not covered; Microsoft appears only in passing.
- Products and dates change fast: the state described is that of 13 October 2026.



## IMPORTANCE

Stating limits protects against conclusions beyond what the sources support.



## POSITIVE IMPLICATIONS

Level C and † items are flagged throughout the study.



## NEGATIVE IMPLICATIONS

Left unverified: the Google whitepaper original and the Sapkota et al. pages.

# Conclusion



## An agent is the artifact; agentic is the property.

- 1 Start with the simplest solution (Anthropic; OpenAI).
- 2 Test end to end and by subtask before granting autonomy (Shavit et al.).
- 3 Keep people in high-risk actions; plan for interruption and tracing (OpenAI; Shavit et al.).
- 4 Treat content the agent reads as potentially hostile (Shavit et al.; OWASP +).
- 5 State the definition you adopt: there is no standard for the two terms.



### IMPORTANCE

The terms shape purchasing, architecture and governance decisions.



### POSITIVE IMPLICATIONS

Efficiency and scale when the system is simple, evaluated and supervised.



### NEGATIVE IMPLICATIONS

Cost, multi-agent failures, security risks and unstable platforms.

# References (1/2)



**ANTHROPIC.** *Building effective agents.* Engineering at Anthropic, 19 Dec 2024. [anthropic.com/engineering/building-effective-agents](https://anthropic.com/engineering/building-effective-agents) [A]

**ANTHROPIC.** *How we built our multi-agent research system.* Engineering at Anthropic, n.d. [anthropic.com/engineering/multi-agent-research-system](https://anthropic.com/engineering/multi-agent-research-system) [B]

**CEMRI, M. et al.** *Why do multi-agent LLM systems fail?* arXiv:2503.13657, 2025. DOI 10.48550/arXiv.2503.13657. [arxiv.org/abs/2503.13657](https://arxiv.org/abs/2503.13657) [B]

**GOOGLE CLOUD.** *Agent2Agent protocol is getting an upgrade.* Google Cloud Blog, n.d. Title as per the page address. [cloud.google.com/blog](https://cloud.google.com/blog) [B]

**HEISE.** *Agent Development Kit and A2A: Google wants to network AI agents.* heise online, 2025 (year inferred). [heise.de/-10360579](https://heise.de/-10360579) [C †]

**LINUX FOUNDATION.** *Formation of the Agentic AI Foundation (AAIF): MCP, goose and AGENTS.md.* 9 Dec 2025. [linuxfoundation.org/press](https://linuxfoundation.org/press) [B]

**MICROSOFT.** *AutoGen Update (discussion 7066).* GitHub, 2025; date per Visual Studio Magazine, 1 Oct 2025. [github.com/microsoft/autogen/discussions/7066](https://github.com/microsoft/autogen/discussions/7066) [B]

**OPENAI.** *A practical guide to building agents.* Page dated 15 Jan 2026. [openai.com/business/guides-and-resources](https://openai.com/business/guides-and-resources) [A]

*Accessed on 13 Oct 2026. Levels: [A] opened and checked · [B] official text read in search · [C †] secondary source only.*

# References (2/2)



**OPENAI.** *Introducing AgentKit*. n.d.; update of 3 Jun 2026 on Agent Builder and Evals. [openai.com/index/introducing-agentkit](https://openai.com/index/introducing-agentkit) [B]

**OWASP GENAI SECURITY PROJECT.** *OWASP Top 10 for Agentic Applications for 2026*. 9 Dec 2025. [genai.owasp.org](https://genai.owasp.org) [B]

**SAPKOTA, R.; ROUMELIOTIS, K. I.; KARKEE, M.** *AI agents vs. agentic AI: a conceptual taxonomy, applications and challenges*. Information Fusion, 2025. DOI 10.1016/j.inffus.2025.103599; [arXiv:2505.10468v5](https://arxiv.org/abs/2505.10468v5) [A]

**SHAVIT, Y. et al.** *Practices for governing agentic AI systems*. OpenAI, 2023 (year inferred). [cdn.openai.com/papers](https://cdn.openai.com/papers) [A]

**WIESINGER, J.; MARLOW, P.; VUSKOVIC, V.** *Agents*. Google, whitepaper, 2024. Original not opened. [kaggle.com/whitepaper-agents](https://kaggle.com/whitepaper-agents) [C †]

**WOOLDRIDGE, M.; JENNINGS, N. R.** *Intelligent agents: theory and practice*. The Knowledge Engineering Review, v. 10, n. 2, p. 115-152, Jun 1995. DOI 10.1017/S0269888900008122. [Cambridge Core](https://www.cambridge.org/core) [B]

**YAO, S. et al.** *ReAct: synergizing reasoning and acting in language models*. arXiv:2210.03629, 2022. Data checked against the reference list of Shavit et al. [arxiv.org/abs/2210.03629](https://arxiv.org/abs/2210.03629) [B]

*Accessed on 13 Oct 2026. Levels: [A] opened and checked · [B] official text read in search · [C †] secondary source only.*