

Market Sharing about AI Agent - Erica Yu

1. Definition about AI Agent

1.1 Gartner definition:

AI agents are **autonomous** or **semiautonomous software entities** that use AI techniques to **perceive, make decisions, take actions** and achieve goals in their digital or physical **environments**.

NOT an AI Agent...

- Large Language Models (they are models)
- Set of instructions that carries out a specific task (like subroutines)
- Automated software functions (like procedures)
- Automated procedures from an RPA workflow (processes)
- Conversational assistants (or any other assistant)
- An interface to an assistant (like an assistant UX)
- ...

2. Growing Market momentum:

2.1 Andrew Ng Snowflake Build 2024:



“Agentic AI is the next major thing in AI.”

“Don’t wait for GPT-5; building an AI Agent with GPT-4 can already achieve—or even surpass—the capabilities we

Key takeaways:

- **Shift to Agentic Workflows:** Traditional AI relies on static, one-off prompts, while agentic workflows enable AI to self-reflect, use external tools, plan tasks, and collaborate, improving reasoning and performance.
- **Iteration Boosts Performance:** Even weaker models (e.g., GPT-3.5) can outperform stronger ones (e.g., GPT-4) when optimized through iterative workflows.
- **Real-World Applications:** AI Agents will enhance customer support, legal analysis, healthcare diagnostics, and software development, driving efficiency.

expect from future models.”

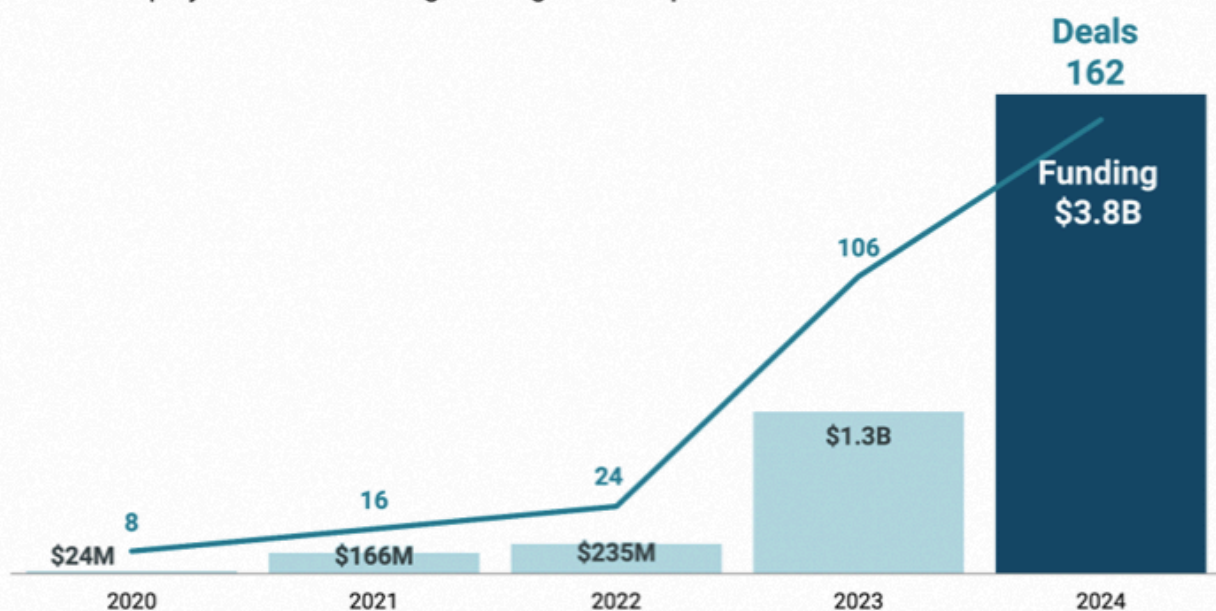
- **Transformational Impact:** Like electricity and cloud computing, agentic AI marks a major shift, revolutionizing automation and creating vast business value.

2.2 Investment boom:

- AI agents' mentions on corporate earnings calls grew **4x** QoQ in **Q4 2024**. And they are on pace to double again this quarter.
- Funding to startups in the field of AI Agents nearly **3x** in **2024**.
- About **15k** funded startups in the U.S. are working on AI Agents.

AI agents have entered the conversation

Disclosed equity deals and funding to AI agent startups



Source: CB Insights financing data (as of 2/27/2025). For the purposes of a funding analysis, excludes LLM developers and companies where agents are not their primary product offering (e.g., Replit, Intercom, etc.). Includes agentic applications and agent-specific infrastructure.

CBINSIGHTS

2.3 Enormous market potential

- By 2028, Gartner predicts that **33% of enterprise software applications** will include agentic AI, up from less than 1% in 2024, with at least 15% of day-to-day work decisions being made autonomously through AI agents. In fact, the AI agents market is expected to see significant growth, rising from **\$5.1 billion in 2024 to \$47.1 billion by 2030**.
- According to CB Insights, **63%** of organizations at Q4 2024 are "placing a lot of importance" on AI agents in the next 12 months, and all are at least experimenting with them.

3. About Manus: breakthrough of hype?

- **LLMs (Claude, DS-R1, Qwen) + Tools, Functions, and APIs:** Gather information and present it in a user-centric



完整测评：硅谷AI导航

way.

- **Applications:** Research, daily life, data analysis, education, and entertainment; current use cases face challenges in multi-source integration and accuracy.
- **Indication:** AI Agents offer a relatively complete 2C user experience; focus on commercialization at the application layer; founders must move fast to gain a first-mover advantage and iterate quickly on data and tools.

4. Technological readiness for boost of AI Agent

Reasoning

- **Oct 2022:** Princeton & Google Brain introduced the ReAct framework.
- **Mar 2023:** ChatGPT-4 launched with significantly improved reasoning and response quality.
- **Mar 2023:** ChatGPT Plugin enabled external tool usage, expanding GPT's capabilities.
- **2024:** OpenAI o1 & o3 tackled complex problems with near-human precision; o3 improved efficiency by 20% and set new ARC benchmarks.
- **Jan 2025:** DeepSeek R1 is A breakthrough in AI reasoning, offering high performance at lower costs, leveraging

Memory (Context)

- **May 2023:** Claude model introduced a 100K context window, over 100× larger than GPT's.
- **Nov 2023:** Claude 2.1 expanded to a 200K context window, significantly enhancing reasoning.
- **Feb 2024:** Google released Gemini 1.5 with a million-token context window.

Tooling

- **Dec 2023:** Simular AI Agent Demo – An AI agent showcasing autonomy and tool interaction.
- **May 2024:** OpenAI's GPT-4o – Autonomous AI capable of executing tasks and integrating with external tools via the Assistants API.
- **Oct 2024:** Claude Computer Use – AI updated to interact with computers and perform tasks requiring direct computer interaction.
- **Jan 2025:** OpenAI's Operator – AI emulating a research analyst to gather and analyze data using various tools.

transparency, innovative training, and open-source collaboration to challenge AI giants.

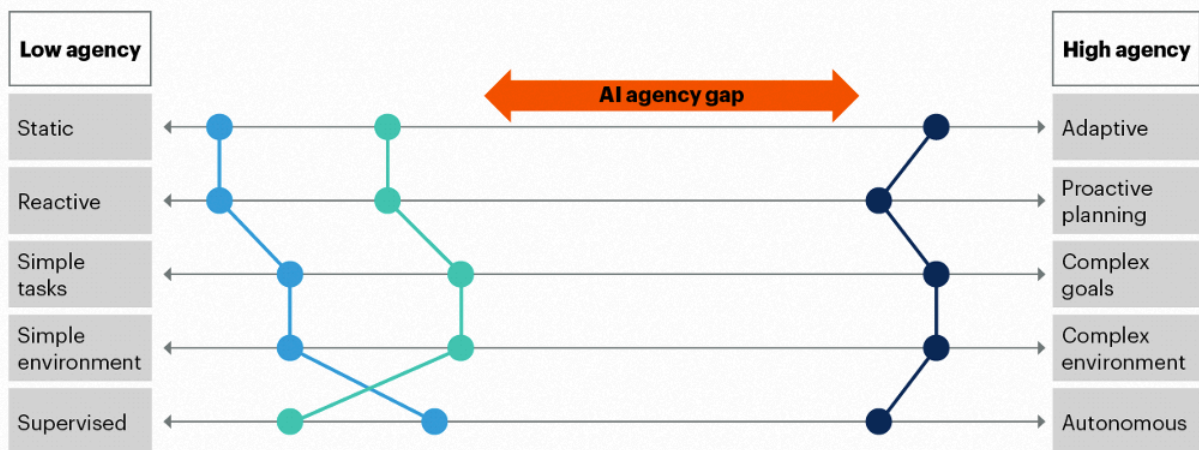
Conclusion: the current development of technology can support the commercial application of a large number of mature scenarios, and can also support the application of embryonic C- end products. However, if AI is required to undertake more complex work, there is still a long space for development in long-term memory, multimodal, API integration and multi-Agent collaboration.

5. Future trends

5.1.1 Improvement of independent agency capabilities: AI Agency is a spectrum, running from current systems with limited agency that only perform specific tasks under narrowly defined conditions to future systems with full agency that **learn from their environment, make decisions and perform tasks independently**. Current LLM-based systems, like copilots and virtual assistants, are more adaptable and can tackle more complex tasks than previous systems. However, they remain unreliable and require **human supervision**; as such, they represent early examples of AI agents rather than the full-fledged agents promised by future AI models.

Mind the AI Agency Gap

● Human agency ● Deterministic chatbots ● LLM-based assistants



Source: Gartner
806843 C

5.1.2 Multiagent systems (MAS):

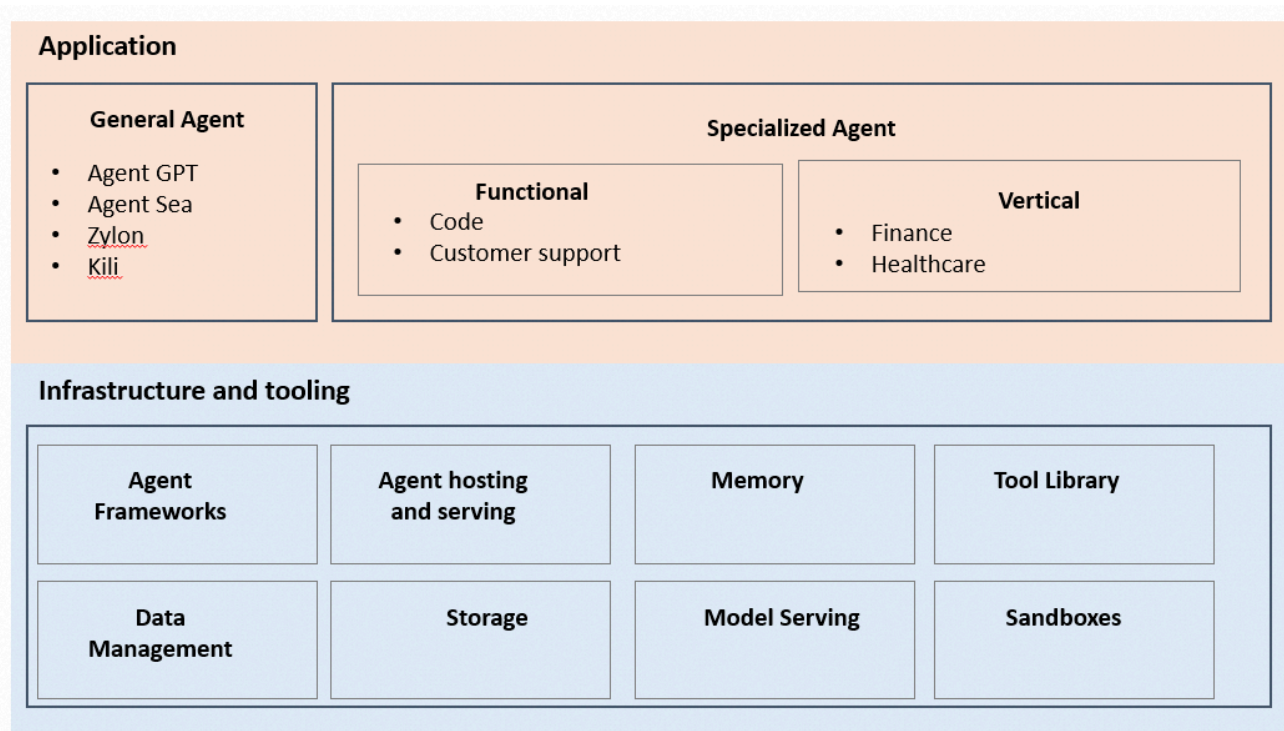
Gartner definition:

MAS is a multiagent system(MAS) is a type of AI system composed of multiple, independent, interacting agents, each capable of perceiving their environment and taking actions. the composition of MAS may vary:

- **By architecture:** MAS architectures exist in a spectrum, from predefined sets of agents with centralized control and deterministic behavior and outcomes, to highly dynamic, constantly evolving collectives of agents with distributed control, adaptive behavior and emerging outcomes at the system level.
- **By agent type/role:** MASs may also vary by type of agent; sometimes all agents have the same role, while in other cases agents play different roles, such as coordinator, guard, communicator, sensor, solver or actuator.
- **By technique:** Agents may also vary in terms of the AI technique that enables each of them, depending on their task in the collective.

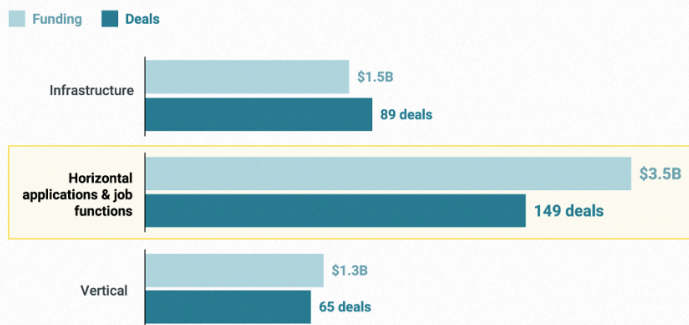
Bottleneck: A distributed agent network and interconnection protocol integrating cloud and edge: **A globally unified AI OS is needed.** A Dev Agent startup, founded by a former Android founder, stated: "*We need an AI moment similar to the Android revolution.*"

6. Landcapse of AI Agent



Horizontal AI agent applications lead venture activity

Disclosed equity deals and funding to AI agent startups since 2020



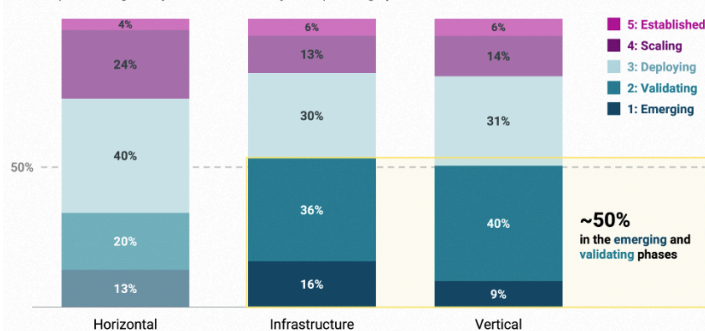
Source: CB Insights. Data as of 2/27/2025. Note: For the purposes of a funding analysis, excludes LLM developers and companies where agents are not their primary product offering (e.g., Replit, Intercom, etc.).

CBINSIGHTS

As horizontal AI agents mature, what's next?

Vertical and infrastructure categories are on the cusp of market maturation

Share of private AI agents by Commercial Maturity score per category



Source: CB Insights Commercial Maturity scores (as of 2/27/2025). Includes established companies that have launched agent offerings. Commercial Maturity measures a private company's ability to compete for customers or serve as a partner.

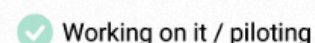
CBINSIGHTS
























6.1 Big tech dominate general AI Agent

- **Cost:** AI agents from big tech and leading LLM developers like OpenAI and Anthropic will pressure private players in 2025 due to their cost advantages.
- **Distribution:** OpenAI's ChatGPT, with 400M weekly users, has a significant distribution edge as it rolls out agents like Operator in January 2025.
- **Risk:** 40% of corporate strategy teams (CB Insights) prioritize risk management in genAI vendor selection. AI agents require access to sensitive data and operate autonomously, raising concerns over privacy, liability, and security. Risk-averse enterprises will likely favor established cloud providers they already trust.

The AI agent arms race

Big tech and LLM developers expand AI agent offerings



Company	Valuation	AI agent development tooling	AI agent offering	Details
	\$3.67T			Angling toward personal agents with Apple Intelligence and Siri
	\$3.40T			Launched a developer platform for building and deploying AI agents called Nvidia AI Blueprints (January 2025)
	\$3.05T			Offers Azure AI Agent Service developer platform and Copilot Studio to create agents; offers pre-built agents in Dynamics 365
	\$2.42T			Offers Amazon Bedrock Agents for developers to build and orchestrate agents; offers genAI-powered assistant Amazon Q
	\$2.25T			Offers Vertex AI Agent Builder and Google Agentspace; debuted Gemini-based agents like Project Mariner and Deep Research
	\$1.82T			Offers Meta AI, its intelligent assistant, with reports suggesting it's exploring more complex web-based task agents
	\$157B			Offers Assistants API; introduced Operator, a web browsing agent (January 2025)
ANTHROPIC	\$60B			Introduced tool use for Claude (May 2024) and computer use for developers (October 2024)

Source: CB Insights and company releases. Data as of 2/18/2025. Includes US-based big tech companies based on market cap and leading LLM developers based on valuation.



6.2 Specialized agents promising field in capital markets:

- **Programming:** The most mature and widely adopted category, with early successful deployments. For example, Cognition AI's **Devin** can replace junior developers, while Replit's AI Agent generates full web applications based on natural language instructions.
- **Customer Support & Sales:** Companies like **Decagon**, serving Duolingo, Notion, and Eventbrite, enable AI agents to automate up to **70%** of support tickets. In sales, **Clay** leverages AI agents to scrape and enrich target customer data, facilitating personalized mass outreach.
- **Business Research:** **Hebbia** is used by **one-third of the top 50 global asset management firms**, as well as consulting firms and law offices, to streamline research.
- **Academic Research:** **Elicit** excels in generating sections of academic papers and processing vast datasets.
- **Healthcare:** **Epic**, the largest medical information company, employs AI agents to automate administrative tasks like patient scheduling, reducing manual workload.

6.3 Private AI Agent go specialized

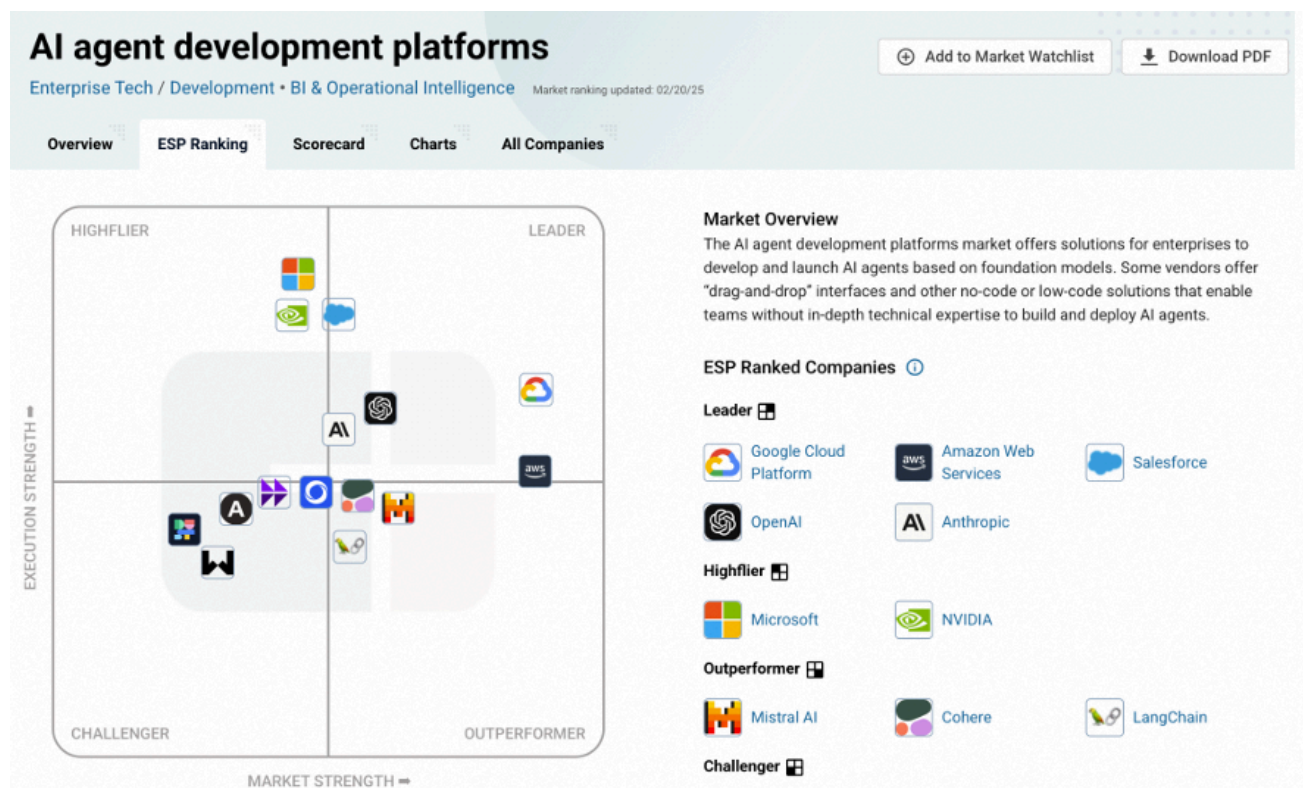
Define a good use case for AI Agent:

- **High-repetition tasks:** Ideal for customer support, sales, and similar roles.
- **Modular workflows:** Best for self-contained software development tasks requiring minimal background knowledge.
- **Information retrieval & processing:** Effective for research tasks without complex tool dependencies.
- **Augmentation over replacement:** AI should assist and integrate gradually rather than fully replace human roles.
- **Autonomy & adaptability:** Suitable for dynamic environments, complex decision-making, and continuous learning.

6.4 Infrastructure, framework and tooling

6.4.1 Critical competition field for Alibaba cloud: AI agent development platforms

Full-Stack Solutions: These platforms provide end-to-end support for enterprises developing and deploying AI agents. Some offer no-code or low-code tools with drag-and-drop interfaces, enabling teams without deep AI expertise to build and launch agents—democratizing AI development.



• AWS:

- **Team structure:** AWS formed a new Agentic AI group, Led by **Swami Sivasubramanian**, this group focuses on developing AI agents to automate tasks for users. Sivasubramanian reports directly to AWS CEO Matt Garman.
- **Products:**

- **Amazon Q Developer Agents:** AWS has introduced new agents in Amazon Q to reduce workloads. These agents leverage advanced models like Claude 3.7 Sonnet for more interactive developer experiences.
- **Alexa+:** An upgraded version of Alexa that incorporates agentic AI capabilities, allowing it to perform tasks autonomously, such as booking rides or browsing websites.

■ **BedRock:**

Function Name	Brief Description	Release Date
Amazon Bedrock Agents	Executes multistep tasks using foundation models and enterprise systems.	Jul-22
Multi-Agent Orchestrator Framework	Manages multiple AI agents for complex conversational scenarios, integrating with AWS Lambda and other platforms.	Dec-24
Amazon Bedrock Multi-Agent Collaboration	Enables multiple agents to collaborate under a supervisor agent for complex tasks.	Mar-25
Inline Agent Support	Dynamically creates supervisor agents at runtime for flexible agent management.	Mar-25
Payload Referencing	Allows supervisor agents to reference external data without embedding it in requests.	Mar-25
CloudFormation and CDK Support	Deploys agent networks as code for scalability and reusability.	Mar-25
Enhanced Traceability and Debugging	Provides structured logs and Amazon CloudWatch integration for better monitoring.	Mar-25

● **Microsoft:**

- **CoreAI Team:** Microsoft has formed a new engineering team called CoreAI – Platform and Tools to accelerate AI development (Jan 2025). This team combines resources from the developer division, AI platform, and key teams from the Office of the CTO to build end-to-end AI stacks for both first-party and third-party customers.
- AI Agent products:

■ to B products:

Product Name	Brief Description	Release Date
GitHub Copilot	AI-powered coding assistant for developers, providing code completion, chat, and pull request summaries.	2021 (Initial Release), Ongoing Updates
GitHub Copilot X	Enhanced version of GitHub Copilot with advanced AI capabilities for the entire development lifecycle.	2023 (Announced)
Microsoft Security Copilot	AI agents for security tasks, including phishing triage, alert management, and vulnerability remediation.	2023 (Initial Release), 2025 (Expanded with AI Agents)
Microsoft 365 Copilot	Integrates AI agents into Microsoft 365 applications to enhance productivity and workflow automation.	2023 (Announced), 2025 (Expanded with Adobe Collaboration)
Adobe Express Agent for Microsoft 365 Copilot	Enables creation of visual assets directly within Microsoft applications using AI.	2025 (Private Preview)
Adobe Marketing Agent	AI agent for marketing tasks, integrated with Microsoft 365 Copilot.	2025 (Private Preview)

■ Semantic Kernal/ Azure AI studio

Function Name	Brief Description	Release Date
Agent Chat and Agent Channel	Facilitates agent interactions and collaboration.	August 2024 (Experimental)
Plugin and	Extends agent capabilities with custom	August 2024

Function Calling	plugins and external services.	(Experimental)
Semantic Kernel Agent Framework	Enables creation and management of AI agents for autonomous tasks.	August 2024 (Experimental), March 2025 (Release Candidate)
ChatCompletion Agent	Completes chats using AI models.	August 2024 (Experimental), March 2025 (Release Candidate)
OpenAIAssistant Agent	Utilizes OpenAI Assistant API for agent tasks.	August 2024 (Experimental), March 2025 (Release Candidate)
Multi-Agent Collaboration	Allows multiple agents to collaborate on complex tasks.	February 2025 (Experimental)
Multi-Model Orchestration	Enables agents to use different AI models for diverse tasks.	February 2025 (Experimental)
AzureAI Agent	Integrates Azure AI services into agent workflows.	March 2025 (Release Candidate)
AgentResponse	Handles chat message threads with optional agent threads.	March 2025 (Release Candidate 2)

- **Google**

- **Gemini 2.0:** This AI model is designed to enhance multimodality, native tool use, and advanced reasoning capabilities. It supports the development of AI agents with improved dialogue, memory, and latency
- **Products:**
 - **to B products:**

Product Name	Brief Description	Release Date
Project Mariner	AI agent for human-agent interaction in web browsers.	December 2024 (Announced)
Jules	AI-powered code agent for developers.	December 2024

		(Announced)
Gemini Robotics	AI models designed to control robots.	Mar-25
AI Coach	Tool to improve employee productivity using AI.	Mar-25
Gemma 3	Open AI models for building AI applications.	Mar-25
Agent Trainer	AI-powered simulation and training tool for customer support agents.	Mar-25
Agent Desktop Application	Standalone client for contact centers with AI integration.	Mar-25

■ **Vertex Agents Builder and Google AI Studio:**

Feature/Function	Brief Description	Release Date
Vertex AI Agent Builder No-Code Interface	Allows users to build AI agents without coding, using natural language instructions.	Jan-25
Code-First Orchestration with LangChain and LlamaIndex	Provides advanced tools for developers to build complex AI applications.	Jan-25
Real-Time Data Integration via RAG System	Connects AI agents to trusted data sources like Google Search and enterprise data.	Jan-25
Low-Code to High-Code Development	Offers flexibility for both simple and complex AI application development.	Jan-25
Integration with Dialogflow CX	Enhances natural language understanding for AI agents.	Jan-25
Integration with Vertex AI Search	Enables information retrieval from enterprise data.	Jan-25

6.4.2 Ecosystem example:

- **Development tool:** mainstream developer tools, such like Langchain, a \$30M-funded open-source toolkit for agent developers to integrate LLMs, databases, long-term memory, and multi-step workflows.
 - **Data Curation:** for agents to make informed decisions, they need access to the right context. Examples such as **LlamaIndex**, Enhances AI agents' ability to process and utilize private domain data.
 - **Web search & tool use:** effective agents need to be able to browse the web and leverage other tools and applications. Companies such like browserbase allow developers to automate web interactions using AI and headless browsers.
 - **Evaluation & observability:** many agent applications today operate in constrained environments due to reasoning and reliability challenges. Continuous monitoring and testing are essential for agent reliability, with early-stage startups like Langfuse, Haize Labs, and Coval.
-

7. Customer list



Agentic AI List_20250325.xlsx
617.2KB



Agentic AI funding_20250326.xlsx
196.2KB

Reference and appendix



AI agent_the next big thing about AI.pdf
1.9MB

- What AI Agent can or cannot do:
<https://www.gartner.com/en/podcasts/thinkcast/from-hype-to-reality-what-ai-agents-can-and-cant-do>
- Innovation insights: AI agent (need to be requested separately)
- Business Value for Agentic AI:



Business Value Driver for Agentic AI.pdf

2.9MB

- Emerging Patterns for Building LLM-Based AI Agents (need to be requested separately)
- Capitalize on the AI Agent Opportunity (need to be requested separately)
- Transform Data productivity with AI Agents:



Transform Data Productivity with AI Agents.pdf

3.1MB

- McK POV:



why-agents-are-the-next-frontier-of-generative-ai (1).pdf

396.0KB

- **Comparison of world's leading AI agents:**

	OpenAI's Operator	Anthropic's Claude	Nvidia's Eureka	SAP's Joule	Quidget AI Agent	Perplexity Assistant	Trengo AI Agent
Primary Function	Browser-based automation tasks	Browser and desktop app interaction	Real-time data tasks and research	Enterprise collaboration & SAP integration	Customer support and sales automation	E-commerce purchases, taxi booking, etc.	Unified communication and customer support
Task Automation Capabilities	Fills forms, books flights, orders groceries	Similar to Operator, with desktop app support	Focuses on data integration and execution	Integrates with SAP systems for workflow	Automates customer inquiries and sales	Automates tasks on Android apps	Automates ticket routing & customer interactions

				automatio n			
Integration Options	API-focused, but operates without APIs	Operates without APIs	Research tools	SAP ecosystem	Over 6,000 apps	Integrates with Spotify, Uber	Integrates with CRM tools
Deployment Model	Cloud-based, in ChatGPT Pro	Broader beta availability	On-premise	Hybrid	Multi-channel	Mobile app	Cloud-based
Cost Structure	Usage-based, part of ChatGPT Pro (\$200/month)	Free beta access	Custom pricing	SAP license-based	Starting at \$16/month	Not specified	Starting at £82/month
Key Features	Uses CUA model for GUI interaction	Extends functionality to desktop applications	Emphasizes real-time task execution	Designed for enterprise collaboration	No-code setup, global language support	Multimodal processing for smartphone	Unified inbox for all communication channels
Performance Benchmarks	87% on WebVoyager, 38.1% on OSWorld25	56% on WebVoyager, 22% on OSWorld5	Not specified	Not specified	Not specified	Not specified	Not specified