McKinsey & Company

McKinsey Explainers

What is an AI agent?

AI agents are the tools we use to interact with AI. They can automate and perform complex tasks, such as natural language processing, that would normally require humans.

An AI agent is a software component that has the agency to act on behalf of a user or a system to perform tasks. Users can organize agents into systems that can orchestrate complex workflows, coordinate activities among multiple agents, apply logic to thorny problems, and evaluate answers to user queries.

If you've ever interacted with a customer service chatbot or asked a gen AI model to write you a sonnet, then you're likely already familiar with a rudimentary version of AI agents. And if you've noticed improvements in gen AI's performance since it went mainstream with ChatGPT, you're not wrong. While versions of AI agents have existed for years, the natural-language-processing capabilities of today's gen AI models have unleashed a host of new possibilities, which are enabling systems of agents to plan, collaborate, and complete tasks—and even learn to improve their own performance. As agents become more accurate, companies can increasingly use them to automate organizational processes and help make employees' day-to-day work more efficient.

"The development of gen AI has been extremely fast," says McKinsey Senior Partner Lari Hämäläinen. "Today, the joint human-plus-machine outcome can generate great quality and great productivity." Recent developments in short- and long-term memory structures have enabled these agents to better personalize interactions with both external and internal users, which means the agents are quickly getting better at whatever they are asked to do.

Looking ahead, they're about to get even better; put simply, AI agents are moving from thought to action. In the past 18 months, Google, Microsoft, OpenAI, and others have invested in software libraries and frameworks to support agentic functionality. And with applications such as Microsoft Copilot, Amazon Q, and Google's upcoming Project Astra, which are powered by large language models (LLMs), agents are making a shift from knowledge-based tools to ones that are more action based. In the near future, agents could become as commonplace as mobile applications are today.

Learn more about QuantumBlack, AI by McKinsey.

What are some different types of AI agents?

AI agents can be organized by their capability, role, skill, and the outcome they're trained to deliver. Here is a nonexhaustive list of some of the agents being created today:

— *Individual augmentation ("copilot" agents).* These agents serve as copilots for individual users, with the intention of augmenting that person's productivity and capabilities. Copilot agents, such as Microsoft's 365 Copilot and OpenAI's ChatGPT, can assist with drafting content, writing code, or retrieving knowledge. In some cases, copilot agents can act as "smart" assistants that are tailored to a user's specific workflow. Of course, the impact of this kind of agent is dependent upon each individual worker's own motivation and investment.

— *Workflow automation platforms.* This type of agent focuses on automating single or multistep tasks or smaller workflows, serving as an AI-powered process orchestrator and executor for existing workflows. Examples of this type of agent include Microsoft's Copilot Studio and Salesforce's Agentforce (currently in development). Because these agents are, for the most part, applied to existing processes, success will rely on heavy efforts in implementation, change management, and agent management.

— *Gen AI–native agents for domain solutions.* These agents are purpose-built solutions for specific business domains or functions. Examples include AI-driven customer service systems or AI-enabled software development pipelines. Gen AI–native agents reimagine a particular domain with AI at the core of the solution, rather than traditional AI agents, which layer AI onto existing roles or workflows.

— *AI-native enterprises and operating models.* These agents are woven throughout the enterprise operating model, instead of being applied to individual workflows or functions. In these cases, a company undergoes an end-to-end AI-first redesign where the interaction layer, processes, organizational structures, and even the business model are reimagined. Organizations experienced changes of this magnitude when they first underwent digital transformations, and the same is likely to happen with AI.

— *AI virtual workers.* AI virtual workers are agents that function as employees or team members and represent the most potentially disruptive category of agents. These virtual workers could enable companies to sidestep full organizational transformation by allowing AI to operate within the company's current model, which might help capture value more quickly.

These AI agents are not mutually exclusive. Many organizations will pursue a mix—for instance, rolling out personal AI copilots while automating select workflows and piloting a few virtual workers.

How do AI agents work?

AI agents can support highly complex and ambiguous use cases across industries and business functions. They can use tools designed for humans, like a web browser, as well as tools designed for computers, such as an API. The ability to do both gives AI agents the flexibility to operate across technology architectures, inside and outside of organizations, without requiring significant modifications to those architectures.

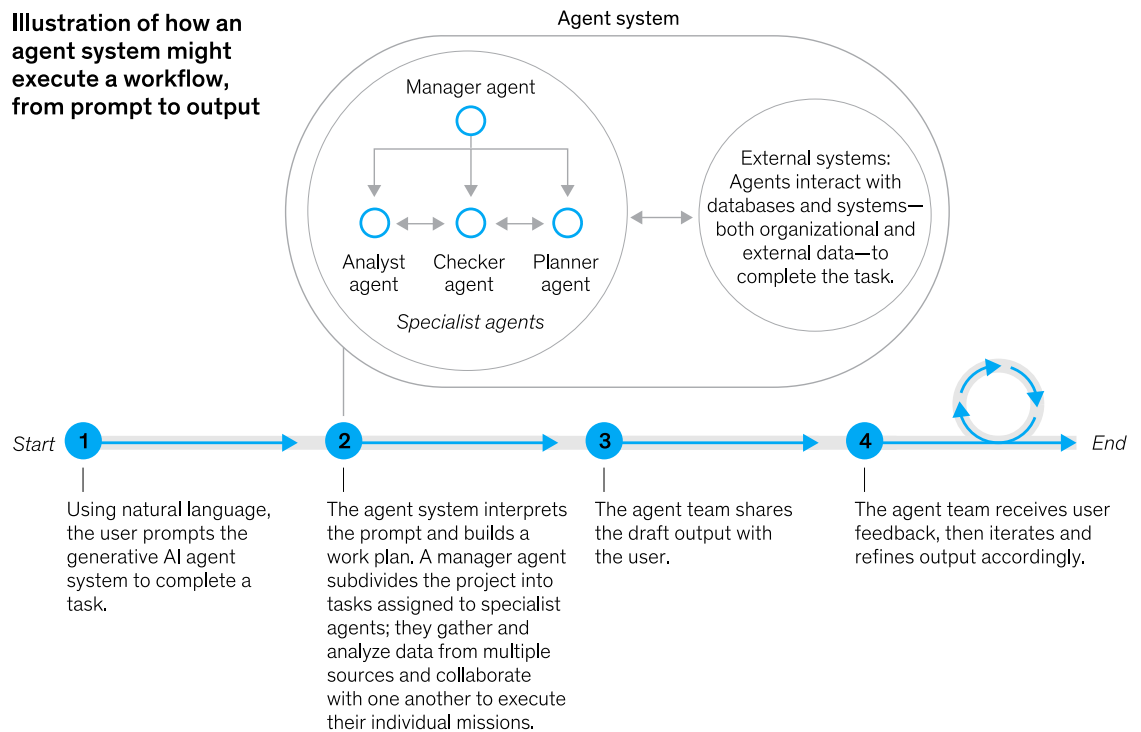The process by which an AI agent works generally follows four steps (Exhibit 1):

— *A user gives the agent system a task.* AI agents work autonomously to plan and derive how to achieve the task.

— *Agent system plans, allocates, and executes work.* An AI agent system breaks down a workflow into tasks and subtasks, which a manager agent assigns to other, specialized subagents. These specialized agents draw on prior experiences and learned domain expertise, coordinate with one another, and use both organizational and external data to execute assignments.

— *Agent system may iteratively improve output.* The agent system may request additional user input to ensure accuracy and relevance. Once the final output is delivered, the agent system may request feedback from the user.

— *Agent executes action.* The agent executes any necessary actions to fully complete the task.

Exhibit 1

Agents enabled by generative AI soon could function as hyperefficient virtual coworkers.

Illustration of how an agent system might execute a workflow, from prompt to output

Agent system

Manager agent

Analyst agent   Checker agent   Planner agent

Specialist agents

External systems: Agents interact with databases and systems—both organizational and external data—to complete the task.

Start 1 ———————— 2 ———————— 3 ———————— 4 ———— End

1. Using natural language, the user prompts the generative AI agent system to complete a task.

2. The agent system interprets the prompt and builds a work plan. A manager agent subdivides the project into tasks assigned to specialist agents; they gather and analyze data from multiple sources and collaborate with one another to execute their individual missions.

3. The agent team shares the draft output with the user.

4. The agent team receives user feedback, then iterates and refines output accordingly.

McKinsey & Company

Any deployment of AI agents should include a range of controls. Constructive-feedback loops, for example, allow agents to review and refine their work. AI agents can also be programmed to teach themselves to solve problems or to escalate them to a human manager. Agents also can work better together: A critic specialist agent can review a plan created by a creator agent and ask for iterations, which can result in better outputs. Some AI agents can even ask managers direct questions. Organizations can also develop specialized agents to automatically test and correct the output of other agents based on ethical and bias concerns.

Learn more about QuantumBlack, AI by McKinsey.

How are AI agents related to LLMs?

AI agents work with different AI models to accomplish tasks. When used to communicate with humans, AI agents work with LLMs that are equipped with natural-language-processing capabilities. Take the example of a self-driving car, which runs on a series of agents that work with various AI models. The AI agent in charge of understanding where a user wants to go might use an LLM. But the agent in charge of making sure it's safe for the car to turn left would use a highly specialized design model, not an LLM, for that specific type of decision.

How might AI agents affect business growth?

McKinsey estimates that in the long term, enterprise use cases of gen AI could create up to $4.4 trillion of value annually. However, organizations won't be able to turn this potential into business growth and higher productivity unless they can quickly implement AI to reimagine and transform how work is done. AI agents can help mine that mountain of value faster, better, and cheaper than other, older technologies.

But gen AI's value goes beyond the automation of common work tasks. McKinsey forecasts that organizations could deploy AI agents to help reimagine processes and modernize their IT infrastructures. This could include everything from switching to easier-to-use programming languages and transitioning to modern frameworks that provide more functionality, to restructuring systems so they're more modular and migrating applications to run on cheaper cloud-computing environments. Tech leaders can use multiple specialized AI agents, each with a distinct role and expertise, to collaborate on complex tasks and iterate in real time with feedback from humans. The real value, though, will come from orchestrating agents to complete discrete tasks as well as entire software development processes.

Some industries are already deploying AI agents regularly. Customer service bots, for instance, have become de rigueur on many customer-facing websites. According to McKinsey research on gen AI's economic potential, organizations using gen AI–enabled customer service agents increased issue resolution by 14 percent per hour and reduced time spent handling issues by 9 percent. "Over time," says McKinsey Senior Partner Jorge Amar, "I expect to see gen AI agents improve customer satisfaction and generate revenue. They will be critical in selling new services

or addressing broader needs. This will open up a wider array of customer experience options for companies, such as offering more high-touch interactions with human agents as a premium service."

More generally, gen AI–enabled agents can potentially ease the automation of complex use cases in three important ways:

— *Agents can more easily manage less predictable situations.* Rule-based systems tend to break down when they face situations the rules' designers didn't anticipate. Gen AI agent systems, by contrast, can handle a wide variety of scenarios for a given use case. Because they are built using foundation models, which have been trained on large, unstructured data sets, gen AI agents can adapt in real time to different scenarios and perform specialized tasks.

— *Agent systems can be directed with natural language.* Natural language processing can allow users to encode complex workflows more quickly and easily than ever before. Gen AI agents coded with natural language processing can enable a wider range of workers to envision and use AI tools to achieve outcomes without needing to specifically program or orchestrate a larger IT project.

— *Agents can work with existing software tools and platforms.* Gen AI agents run on foundation models, which allow the agents to use tools and communicate across a broader digital ecosystem. Without foundation models, these capabilities would require extensive manual efforts to integrate systems or to collate the outputs from different systems. Agents can significantly reduce the amount of work and rework required to integrate across systems.

What's more, according to McKinsey Partner Aaron Bawcom, "Agents' capabilities can compound in reaction to their environments when they work together. They can develop unexpected behaviors and skills that are not explicitly programmed, equaling greater than the sum of their parts. This is what's known as emergent AI."

Learn more about QuantumBlack, AI by McKinsey.

What's a real example of a company working with AI agents?

Lenovo, the Chinese technology company, has deployed AI agents in two main areas of its business: software engineering and customer support. Software engineers at the company are already seeing up to 15 percent improvements, says Arthur Hu, chief technology officer of Lenovo's Solutions and Services Group. The company's COO and head of strategy, Linda Yao, adds that the customer service side has seen double-digit productivity gains in call handling time.

To date, Lenovo has refined gen AI agents to act as virtual assistants. In the future, Yao envisions AI agents acting as deputies to humans, deployed to complete tasks independently.

What are some other enterprise use cases for AI agents?

Here are three hypothetical use cases that offer a glimpse of what could be possible with AI agents in the future:

— *Loan underwriting.* An AI agent system—which comprises multiple specialized agents—could handle a wide range of credit risk scenarios. In the past, this has typically been a time-consuming and highly collaborative process that involves compiling, analyzing, and reviewing various pieces of information related to the borrower, loan type, and other variables.

— *Code documentation and modernization.* Legacy software applications and systems at large enterprises often pose security risks and can slow the pace of business innovation. AI agents could help streamline the process of modernizing these systems. For example, a company could deploy one specialized agent as a legacy software expert to analyze old code, while another—a quality assurance agent—could critique the documentation and iteratively refine the AI's output to ensure accuracy and adherence to organizational standards.

— *Online-marketing campaigns.* Designing, launching, and running an online-marketing campaign can be a complex, multilayered process involving many different people. A single AI agent system could develop, test, and iterate campaign ideas based on input from a company's marketing professionals. Then, a digital-marketing agent could gather insights to craft strategies and copy while design agents could build tailored content.

Learn more about QuantumBlack, AI by McKinsey.

What hurdles are organizations facing in their adoption of AI agents?

Building trust is a big hurdle in adopting AI agent technology, says McKinsey Partner Nicolai von Bismarck: "We find that customers across all age groups—even Gen Z—still prefer live phone conversations for customer help and support." One bank, he continues, addressed this by creating an architecture that checks for errors or "hallucinations" from the AI agent before an answer is shared with a customer, thereby reducing the number of incorrect responses and building trust.

"The companies that derive the most value from AI," says McKinsey Partner Roger Roberts, "will be those that create trust with their customers, employees, and stakeholders. People must trust AI enough to hand over tasks. Companies' ethical decisions must be rooted in the values unique to each organization and the values of a society that places humans at the center of the AI ecosystem."

Another challenge, according to McKinsey Senior Partner Amar, will be change management as organizations scale AI agents. "It's much broader than simply rolling out a new set of tools," he says. "Companies are going to need to rewire how functions work so they can get the full value from gen AI agents."

Part of this rewiring includes the adoption of new technologies, as well as tweaks to existing ones so that they work better with tools based on machine learning and gen AI. Organizations should also adjust their operating models to support small teams that are working iteratively on new services. What's more, leaders should create incentives that help workers learn how to use—and trust—the new tools.

Finally, "Data protection is a major concern for leaders when it comes to deploying AI agents," says McKinsey Partner Bawcom. Companies that are pursuing an AI agent program should carefully implement the proper controls for security, operations, and data. There are many evolving approaches and solutions—both off the shelf and bespoke.

How might AI agents change an organization's tech architecture?

The proliferation of AI agents will likely change the way organizations run their technology programs. McKinsey anticipates that IT architectures will shift away from a traditional application-focused pattern to a new, multiagent model. In this type of architecture, tech leaders oversee up to thousands of agents that communicate with one another, as well as with humans and outside programs, to achieve a common goal.

Here are three ways that tech leaders could deploy agents within their current environments (Exhibit 2):
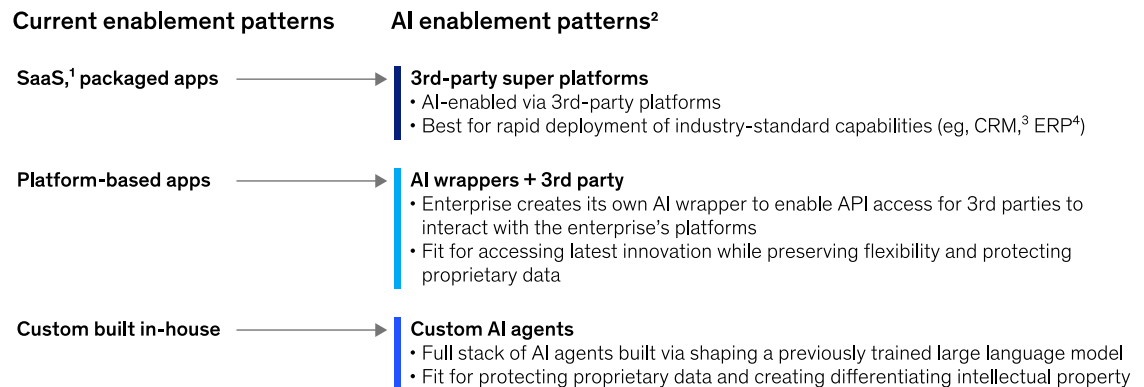
— *Super platforms.* These are the next generation of third-party business applications, such as collaboration tools or customer relationship management (CRM) programs, that include built-in gen AI agents. A user could deploy a CRM super platform, for example, to run a sales report while automatically communicating with the company's analytics tools. In the past, this would have required use-case-specific programming.

— *AI wrappers.* These tools allow enterprise services to communicate and collaborate with third-party services via APIs without exposing a company's proprietary data. A bank, for example, could build a gen A–enabled wrapper around an internally developed, AI-driven credit risk model. The AI wrapper could then initiate actions, such as asking a vendor's LLM to generate risk factors based on customer data and credit scores, without revealing the data to the vendor or even to the LLM.

— *Custom AI agents.* Employees can develop custom gen AI–enabled agents by fine-tuning a

pretrained LLM or accessing a company's proprietary data by using retrieval-augmented generation (RAG). For example, in a customer service setting, an employee could feed an existing model with customer data, call center transcripts, and company policies to create a gen AI agent to assist call center staff in responding to customer questions.

Organizations will need to adjust their IT architectures to better serve these new agent-focused workflows and allow for continuous iteration and improvement.

Exhibit 2

Architectures may evolve with three AI enablement patterns.

Current enablement patterns	AI enablement patterns[2]
SaaS,[1] packaged apps	**3rd-party super platforms** • AI-enabled via 3rd-party platforms • Best for rapid deployment of industry-standard capabilities (eg, CRM,[3] ERP[4])
Platform-based apps	**AI wrappers + 3rd party** • Enterprise creates its own AI wrapper to enable API access for 3rd parties to interact with the enterprise's platforms • Fit for accessing latest innovation while preserving flexibility and protecting proprietary data
Custom built in-house	**Custom AI agents** • Full stack of AI agents built via shaping a previously trained large language model • Fit for protecting proprietary data and creating differentiating intellectual property

[1]Software as a service.
[2]For more on strategic gen AI considerations, see "Technology's generational moment with generative AI: A CIO and CTO guide," McKinsey, July 11, 2023.
[3]Customer relationship management.
[4]Enterprise resource planning.

McKinsey & Company

Learn more about QuantumBlack, AI by McKinsey.

What steps can organizations take to implement AI agents?

Leaders can focus on three important areas:

— *Closely review any tech proposal that has a long timeline and requires many people.* This means reviewing all large proposals to determine how gen AI could reduce costs and

shorten timelines. Leaders should be particularly skeptical of proposals that purport to incorporate gen AI capabilities, which may be limited or ancillary in terms of the value potential.

— *Focus on the biggest problems.* Small-scale initiatives generally lead to small-scale outcomes. So companies would do well to identify the largest and most complex tech problems—the ones that are very expensive, with multiyear timelines, and are responsible for serious technical debt—and focus their use of gen AI on solving them.

— *Get ahead of the talent, technology, and operating-model implications.* As the multiagent approach scales, leaders will need to understand and plan for the business implications. These include rethinking their talent strategy and reskilling programs, adapting their operating models accordingly, and reevaluating their operating and capital expenditures.

The evolution of AI agents is only just beginning, with many learnings and discoveries still to come. Much of the work in the field of AI agents is moving out of the lab to full scale. AI agents offer new capabilities to help reimagine the ways that people work across organizations of all industries and sizes.

Learn more about QuantumBlack, AI by McKinsey. And check out AI agent–related job opportunities if you're interested in working with McKinsey.

Articles referenced:

— "Why investing in AI trust pays off," December 18, 2024, Roger Roberts

— "Extracting value from AI in banking: Rewiring the enterprise," December 9, 2024, Carlo Giovine, Larry Lerner, Renny Thomas, Shwaitang Singh, Sudhakar Kakulavarapu, and Violet Chung, with Yuvika Motwani

— "AI for IT modernization: Faster, cheaper, better," December 2, 2024, Aaron Bawcom and Matt Fitzpatrick, with Chi Wai Cheung, Dan Collins, and Dante Gabrielli

— "Enterprise technology's next chapter: Four gen AI shifts that will reshape business technology," December 2, 2024, James Kaplan, Mark Gu, and Megha Sinha

— "Charting a path to the data- and AI-driven enterprise of 2030," September 5, 2024, Asin Tavakoli, Holger Harreis, Kayvaun Rowshankish, and Michael Bogobowicz

— "Why agents are the next frontier of generative AI," July 24, 2024, Lareina Yee, Michael Chui, and Roger Roberts, with Stephen Xu

— "The promise and the reality of gen AI agents in the enterprise," May 17, 2024, Jorge Amar, Lari Hämäläinen, and Nicolai von Bismarck

— "The economic potential of generative AI: The next productivity frontier," June 14, 2023, Alex Singla, Alex Sukharevsky, Eric Hazan, Kate Smaje, Lareina Yee, Michael Chui, Rodney Zemmel, and Roger Roberts

Get to know and directly engage with senior McKinsey experts on AI agents

Aaron Bawcom and **Nicolai von Bismarck** are partners in McKinsey's Boston office; **Asin Tavakoli** is a partner in the Düsseldorf office, where **Holger Harreis** is a senior partner; **Carlo Giovine** is a partner in the London office; **James Kaplan** is a partner in the New York office, where **Kayvaun Rowshankish** is a senior partner; **Jorge Amar** is a senior partner in the Miami office; **Lareina Yee** is a senior partner in the Bay Area office, where **Michael Chui** is a senior fellow and **Roger Roberts** is a partner; **Lari Hämäläinen** is a senior partner in the Seattle office; **Larry Lerner** is a partner in the Washington, DC, office; **Renny Thomas** is a senior partner in the Mumbai office; and **Violet Chung** is a senior partner in the Hong Kong office.