**AGENTIC AI**

What is an AI agent?

AI agents are software systems that use AI to pursue goals and complete tasks on behalf of users. They show reasoning, planning, and memory and have a level of autonomy to make decisions, learn, and adapt.

Their capabilities are made possible in large part by the multimodal capacity of generative AI and AI foundation models. AI agents can process multimodal information like text, voice, video, audio, code, and more simultaneously; can converse, reason, learn, and make decisions. They can learn over time and facilitate transactions and business processes. Agents can work with other agents to coordinate and perform more complex workflows.

Key features of an AI agent

As explained above, while the key features of an AI agent are reasoning and acting (as described in [ReAct Framework](https://arxiv.org/pdf/2210.03629" \t "_blank)) more features have evolved over time.

* **Reasoning:** This core cognitive process involves using logic and available information to draw conclusions, make inferences, and solve problems. AI agents with strong reasoning capabilities can analyze data, identify patterns, and make informed decisions based on evidence and context.
* **Acting**: The ability to take action or perform tasks based on decisions, plans, or external input is crucial for AI agents to interact with their environment and achieve goals. This can include physical actions in the case of embodied AI, or digital actions like sending messages, updating data, or triggering other processes.
* **Observing**: Gathering information about the environment or situation through perception or sensing is essential for AI agents to understand their context and make informed decisions. This can involve various forms of perception, such as computer vision, natural language processing, or sensor data analysis.
* **Planning**: Developing a strategic plan to achieve goals is a key aspect of intelligent behavior. AI agents with planning capabilities can identify the necessary steps, evaluate potential actions, and choose the best course of action based on available information and desired outcomes. This often involves anticipating future states and considering potential obstacles.
* **Collaborating**: Working effectively with others, whether humans or other AI agents, to achieve a common goal is increasingly important in complex and dynamic environments. Collaboration requires communication, coordination, and the ability to understand and respect the perspectives of others.
* **Self-refining**: The capacity for self-improvement and adaptation is a hallmark of advanced AI systems. AI agents with self-refining capabilities can learn from experience, adjust their behavior based on feedback, and continuously enhance their performance and capabilities over time. This can involve machine learning techniques, optimization algorithms, or other forms of self-modification.

How do AI agents work?

Every agent defines its role, personality, and communication style, including specific instructions and descriptions of available tools.

* **Persona**: A well defined persona allows an agent to maintain a consistent character and behave in a manner appropriate to its assigned role, evolving as the agent gains experience and interacts with its environment.
* **Memory**: The agent is equipped in general with short term, long term, consensus, and episodic memory. Short term memory for immediate interactions, long-term memory for historical data and conversations, episodic memory for past interactions, and consensus memory for shared information among agents. The agent can maintain context, learn from experiences, and improve performance by recalling past interactions and adapting to new situations.
* **Tools**: Tools are functions or external resources that an agent can utilize to interact with its environment and enhance its capabilities. They allow agents to perform complex tasks by accessing information, manipulating data, or controlling external systems, and can be categorized based on their user interface, including physical, graphical, and program-based interfaces. Tool learning involves teaching agents how to effectively use these tools by understanding their functionalities and the context in which they should be applied.
* **Model**: Large language models (LLMs) serve as the foundation for building AI agents, providing them with the ability to understand, reason, and act. LLMs act as the "brain" of an agent, enabling them to process and generate language, while other components facilitate reason and action.

What are the types of agents in AI?

AI agents can be categorized in various ways based on their capabilities, roles, and environments. Here are some key categories of agents:

There are different definitions of agent types and agent categories.

Based on interaction

One way to categorize agents is by how they interact with users. Some agents engage in direct conversation, while others operate in the background, performing tasks without direct user input:

* **Interactive partners** (also known as, surface agents) – Assisting us with tasks like customer service, healthcare, education, and scientific discovery, providing personalized and intelligent support. Conversational agents include Q&A, chit chat, and world knowledge interactions with humans. They are generally user query triggered and fulfill user queries or transactions.
* **Autonomous background processes** (also known as, background agents) – Working behind the scenes to automate routine tasks, analyze data for insights, optimize processes for efficiency, and proactively identify and address potential issues. They include workflow agents. They have limited or no human interaction and are generally driven by events and fulfill queued tasks or chains of tasks.

Based on number of agents

* **Single agent**: Operate independently to achieve a specific goal. They utilize external tools and resources to accomplish tasks, enhancing their functional capabilities in diverse environments. They are best suited for well defined tasks that do not require collaboration with other AI agents. Can only handle one foundation model for its processing.
* **Multi-agent**: Multiple AI agents that collaborate or compete to achieve a common objective or individual goals. These systems leverage the diverse capabilities and roles of individual agents to tackle complex tasks. Multi-agent systems can simulate human behaviors, such as interpersonal communication, in interactive scenarios. Each agent can have different foundation models that best fit their needs.

**Benefits of using AI agents**

AI agents can enhance the capabilities of language models by providing autonomy, task automation, and the ability to interact with the real world through tools and embodiment.

**What is memory for AI Agent?**

AI agent memory refers to an [artificial intelligence](https://www.ibm.com/think/topics/artificial-intelligence) (AI) system’s ability to store and recall past experiences to improve decision-making, perception and overall performance.  
  
Unlike traditional AI models that process each task independently, AI agents with memory can retain context, recognize patterns over time and adapt based on past interactions. This capability is essential for goal-oriented AI applications, where feedback loops, knowledge bases and adaptive learning are required.

Memory is a system that remembers something about previous interactions. [AI agents](https://www.ibm.com/think/topics/ai-agents) do not necessarily need memory systems. Simple reflex agents, for example, perceive real-time information about their environment and act on it or pass that information along.

A basic thermostat does not need to remember what the temperature was yesterday. But a more advanced “smart” thermostat with memory can go beyond simple on or off temperature regulation by learning patterns, adapting to user behavior and optimizing energy efficiency. Instead of reacting only to the current temperature, it can store and analyze past data to make more intelligent decisions.

[Large language models](https://www.ibm.com/think/topics/large-language-models) (LLMs) cannot, by themselves, remember things. The memory component must be added. However, one of the biggest challenges in AI memory design is optimizing retrieval efficiency, as storing excessive data can lead to slower response times.  
  
Optimized memory management helps ensure that AI systems store only the most relevant information while maintaining low-[latency](https://www.ibm.com/think/topics/latency) processing for real-time applications.

Types of agentic memory

Researchers categorize agentic memory in much the same way that psychologists categorize human memory. The influential [Cognitive Architectures for Language Agents (CoALA) paper](https://arxiv.org/abs/2309.02427)1 from a team at Princeton University describes different types of memory as:

Short-term memory

Short-term memory (STM) enables an AI agent to remember recent inputs for immediate decision-making. This type of memory is useful in conversational AI, where maintaining context across multiple exchanges is required.  
  
For example, a [chatbot](https://www.ibm.com/think/topics/chatbots) that remembers previous messages within a session can provide coherent responses instead of treating each user input in isolation, improving [user experience](https://www.ibm.com/think/topics/user-experience). For example, OpenAI’s ChatGPT retains chat history within a single session, helping to ensure smoother and more context-aware conversations.

STM is typically implemented using a rolling buffer or a [context window](https://www.ibm.com/think/topics/context-window), which holds a limited amount of recent data before being overwritten. While this approach improves continuity in short interactions, it does not retain information beyond the session, making it unsuitable for long-term personalization or learning.

Long-term memory

Long-term memory (LTM) allows AI agents to store and recall information across different sessions, making them more personalized and intelligent over time.  
  
Unlike short-term memory, LTM is designed for permanent storage, often implemented using databases, [knowledge graphs](https://www.ibm.com/think/topics/knowledge-graph) or [vector embeddings](https://www.ibm.com/think/topics/vector-embedding). This type of memory is crucial for AI applications that require historical knowledge, such as personalized assistants and recommendation systems.  
  
For example, an AI-powered customer support agent can remember previous interactions with a user and tailor responses accordingly, improving the overall customer experience.

One of the most effective techniques for implementing LTM is [retrieval augmented generation](https://www.ibm.com/think/topics/retrieval-augmented-generation) (RAG), where the agent fetches relevant information from a stored knowledge base to enhance its responses.

Episodic memory

Episodic memory allows AI agents to recall specific past experiences, similar to how humans remember individual events. This type of memory is useful for case-based reasoning, where an AI learns from past events to make better decisions in the future.  
  
Episodic memory is often implemented by logging key events, actions and their outcomes in a structured format that the agent can access when making decisions.  
  
For example, an AI-powered financial advisor might remember a user's past investment choices and use that history to provide better recommendations. This memory type is also essential in robotics and autonomous systems, where an agent must recall past actions to navigate efficiently.

Semantic memory

Semantic memory is responsible for storing structured factual knowledge that an AI agent can retrieve and use for reasoning. Unlike episodic memory, which deals with specific events, semantic memory contains generalized information such as facts, definitions and rules.  
  
AI agents typically implement semantic memory using knowledge bases, symbolic AI or [vector embeddings](https://www.ibm.com/think/topics/vector-embedding), allowing them to process and retrieve relevant information efficiently. This type of memory is used in real-world applications that require domain expertise, such as legal AI assistants, medical diagnostic tools and enterprise knowledge management systems.  
  
For example, an AI legal assistant can use its knowledge base to retrieve case precedents and provide accurate legal advice.

Procedural memory

Procedural memory in AI agents refers to the ability to store and recall skills, rules and learned behaviors that enable an agent to perform tasks automatically without explicit reasoning each time.  
  
It is inspired by human procedural memory, which allows people to perform actions such as riding a bike or typing without consciously thinking about each step. In AI, procedural memory helps agents improve efficiency by automating complex sequences of actions based on prior experiences.

AI agents learn sequences of actions through training, often using reinforcement learning to optimize performance over time. By storing task-related procedures, AI agents can reduce computation time and respond faster to specific tasks without reprocessing data from scratch.

**Frameworks & tools for agentic AI memory**

Developers implement memory using external storage, specialized architectures and feedback mechanisms. Since AI agents vary in complexity—ranging from simple reflex agents to advanced learning agents—memory implementation depends on the [agent’s architecture](https://www.ibm.com/think/topics/agentic-architecture), use case and required adaptability.

LangChain

One key [agent framework](https://www.ibm.com/think/insights/top-ai-agent-frameworks) for building memory-enabled AI agents is [LangChain](https://www.ibm.com/think/topics/langchain" \t "_self), which facilitates the integration of memory, [APIs](https://www.ibm.com/think/topics/api) and reasoning [workflows](https://www.ibm.com/think/topics/agentic-workflows). By combining LangChain with [vector databases](https://www.ibm.com/think/topics/vector-database), AI agents can efficiently store and retrieve large volumes of past interactions, enabling more coherent responses over time.

LangGraph

[LangGraph](https://www.ibm.com/think/topics/langgraph) allows developers to construct hierarchical memory graphs for AI agents, improving their ability to track dependencies and learn over time.  
  
By integrating vector databases, agentic systems can efficiently store embeddings of previous interactions, enabling contextual recall. This is useful for AI-driven docs generation, where an agent must remember user preferences and past modifications.

Other open source offerings

The rise of [open source](https://www.ibm.com/think/topics/open-source) frameworks has accelerated the development of memory-enhanced AI agents. Platforms such as GitHub host numerous repositories that provide tools and templates for integrating memory into [AI workflows](https://www.ibm.com/think/topics/ai-workflow).  
  
Additionally, [Hugging Face](https://huggingface.co/) offers pretrained models that can be fine-tuned with memory components to improve AI recall capabilities. Python, a dominant language in AI development, provides libraries for handling [orchestration](https://www.ibm.com/think/topics/ai-agent-orchestration), memory storage and retrieval mechanisms, making it a go-to choice for implementing AI memory systems.

**Agentic AI**

**Agentic AI** refers to artificial intelligence systems designed to operate autonomously, making decisions and taking actions to achieve predefined goals without constant human oversight. Unlike traditional AI, which relies on predefined rules or human intervention, agentic AI exhibits **autonomy**, **goal-driven behavior**, and **adaptability**. It acts as an independent agent, interacting with its environment, learning from experiences, and adapting its strategies to optimize outcomes.

Agentic AI systems are built on advanced machine learning models, such as **large language models (LLMs)**, and leverage techniques like **reinforcement learning** and **decision-making algorithms**. These systems can perceive their environment, reason through complex scenarios, and execute actions while continuously learning and improving.

**Key Characteristics of Agentic AI**

1. **Autonomy**: Agentic AI operates independently, making decisions and taking actions without human intervention. It can handle dynamic and unpredictable environments.
2. **Proactivity**: It anticipates needs, identifies patterns, and takes initiative to address potential issues before they arise.
3. **Adaptability**: Agentic AI learns from its environment and adjusts its behavior based on real-time feedback, improving over time.
4. **Collaboration**: It can work alongside humans or other AI agents, coordinating efforts to achieve shared goals.
5. **Specialization**: Agentic AI often consists of multiple specialized agents, each focusing on a specific task, which collaborate to solve complex problems.

**Applications of Agentic AI**

Agentic AI has diverse applications across industries:

* **Autonomous Vehicles**: It powers self-driving cars by making real-time decisions based on traffic, road conditions, and environmental factors.
* **Healthcare**: It assists in patient diagnosis, treatment planning, and monitoring by analyzing medical data and providing actionable insights.
* **Finance**: Agentic AI is used for algorithmic trading, fraud detection, and automating contract analysis.
* **Customer Service**: It automates responses to customer inquiries, searches for solutions, and escalates unresolved issues to human agents.
* **Supply Chain Management**: It optimizes logistics by predicting demand, rerouting shipments, and managing inventory[**2**](https://www.bing.com/ck/a?!&&p=c30a59e78a40dc255fe6a82d1a0e82bde51cbb3ce86e42026a11c58a3c3fd7afJmltdHM9MTc2MTUyMzIwMA&ptn=3&ver=2&hsh=4&fclid=329d357b-8a28-62a8-364a-23128b36631d&u=a1aHR0cHM6Ly93d3cuaWJtLmNvbS90aGluay90b3BpY3MvYWdlbnRpYy1haQ&ntb=1)[**3**](https://www.bing.com/ck/a?!&&p=6272498985507e5160eccad0cbe2c771f7203f4352c9f1e13f67c6a552bc9fbfJmltdHM9MTc2MTUyMzIwMA&ptn=3&ver=2&hsh=4&fclid=329d357b-8a28-62a8-364a-23128b36631d&u=a1aHR0cHM6Ly9hd3MuYW1hem9uLmNvbS93aGF0LWlzL2FnZW50aWMtYWkv&ntb=1).

**Differences Between Agentic AI and Traditional AI**

Agentic AI differs significantly from traditional AI in its capabilities:

* **Autonomy**: Traditional AI requires human intervention, while agentic AI operates independently.
* **Decision-Making**: Traditional AI follows predefined rules, whereas agentic AI adapts and makes complex decisions in real-time.
* **Environment Interaction**: Traditional AI is limited to static tasks, while agentic AI actively interacts with and responds to its environment[**1**](https://www.bing.com/ck/a?!&&p=e8cb42df87730045acb18d3b1b709ddeed7e7f62d37f9299cb9ca9a5caa536faJmltdHM9MTc2MTUyMzIwMA&ptn=3&ver=2&hsh=4&fclid=329d357b-8a28-62a8-364a-23128b36631d&u=a1aHR0cHM6Ly93d3cuZ2Vla3Nmb3JnZWVrcy5vcmcvYXJ0aWZpY2lhbC1pbnRlbGxpZ2VuY2Uvd2hhdC1pcy1hZ2VudGljLWFpLw&ntb=1)[**2**](https://www.bing.com/ck/a?!&&p=c30a59e78a40dc255fe6a82d1a0e82bde51cbb3ce86e42026a11c58a3c3fd7afJmltdHM9MTc2MTUyMzIwMA&ptn=3&ver=2&hsh=4&fclid=329d357b-8a28-62a8-364a-23128b36631d&u=a1aHR0cHM6Ly93d3cuaWJtLmNvbS90aGluay90b3BpY3MvYWdlbnRpYy1haQ&ntb=1).

**Challenges and Ethical Considerations**

While agentic AI offers immense potential, it also presents challenges:

* **Accountability**: Determining responsibility for decisions made by autonomous systems can be complex.
* **Bias and Fairness**: Ensuring unbiased decision-making is critical, as agentic AI can inherit biases from training data.
* **Safety**: Autonomous systems must be designed to make safe and reliable decisions, especially in high-stakes environments like healthcare or finance.
* **Transparency**: Building trust requires clear explanations of how decisions are made[**2**](https://www.bing.com/ck/a?!&&p=c30a59e78a40dc255fe6a82d1a0e82bde51cbb3ce86e42026a11c58a3c3fd7afJmltdHM9MTc2MTUyMzIwMA&ptn=3&ver=2&hsh=4&fclid=329d357b-8a28-62a8-364a-23128b36631d&u=a1aHR0cHM6Ly93d3cuaWJtLmNvbS90aGluay90b3BpY3MvYWdlbnRpYy1haQ&ntb=1)[**3**](https://www.bing.com/ck/a?!&&p=6272498985507e5160eccad0cbe2c771f7203f4352c9f1e13f67c6a552bc9fbfJmltdHM9MTc2MTUyMzIwMA&ptn=3&ver=2&hsh=4&fclid=329d357b-8a28-62a8-364a-23128b36631d&u=a1aHR0cHM6Ly9hd3MuYW1hem9uLmNvbS93aGF0LWlzL2FnZW50aWMtYWkv&ntb=1).

Agentic AI represents a transformative leap in AI technology, enabling systems to act with purpose and independence. However, its development must balance innovation with ethical responsibility to ensure it benefits society as a whole.