What is agentic AI?

What is agentic AI?

Agentic AI is an artificial intelligence system that can accomplish a specific goal with limited supervision. It consists of AI agents—machine learning models that mimic human decision-making to solve problems in real time. In a multiagent system, each agent performs a specific subtask required to reach the goal and their efforts are coordinated through AI orchestration.

Unlike traditional AI models, which operate within predefined constraints and require human intervention, agentic AI exhibits autonomy, goal-driven behavior and adaptability. The term "agentic" refers to these models' agency, or, their capacity to act independently and purposefully.

Agentic AI builds on generative AI (gen AI) techniques by using large language models (LLMs) to function in dynamic environments. While generative models focus on *creating* content based on learned patterns, agentic AI extends this capability by applying generative outputs toward specific goals. A generative AI model like OpenAI's ChatGPT might produce text, images or code, but an agentic AI system can use that generated content to complete complex tasks autonomously by calling external tools. Agents can, for example, not only tell you the best time to climb Mt. Everest given your work schedule, it can also book you a flight and a hotel.

Your subscription will be delivered in English. You will find an unsubscribe link in every newsletter. You can manage your subscriptions or unsubscribe here. Refer to our IBM Privacy Statement for more information.
https://www.ibm.com/us-en/privacy

What are the advantages of agentic AI?

Agentic systems have many advantages over their generative predecessors, which are limited by the information contained in the datasets upon which models are trained.

Autonomous

The most important advancement of agentic systems is that they allow for autonomy to perform tasks without constant human oversight. Agentic systems can maintain long-term goals, manage multistep problem-solving tasks and track progress over time.

Proactive

Agentic systems provide the flexibility of LLMs, which can generate responses or actions based on nuanced, context-dependent understanding, with the structured, deterministic and reliable features of traditional programming. This approach allows agents to "think" and "do" in a more human-like fashion.

LLMs by themselves can't directly interact with external tools or databases or set up systems to monitor and collect data in real time, but agents can. Agents can search the web, call application programming interfaces (APIs) and query databases, then use this information to make decisions and take actions.

Specialized

Agents can specialize in specific tasks. Some agents are simple, performing a single repetitive task reliably. Others can use perception and draw on memory to solve more complex problems. An agentic architecture might consist of a "conductor" model powered by an LLM that oversees tasks and decisions and supervises other, simpler agents. Such architectures are ideal for sequential workflows but are vulnerable to bottlenecks. Other architectures are more horizontal, with agents working in harmony as equals in a decentralized fashion, but this architecture can be slower than a vertical hierarchy. Different AI applications demand different architectures.

Adaptable

Agents can learn from their experiences, take in feedback and adjust their behavior. With the right guardrails, agentic systems can improve continuously. Multiagent systems possess the scalability to eventually handle broadly scoped initiatives.

Intuitive

Because agentic systems are powered by LLMs, users can engage with them with natural language prompts. This means that entire software interfaces—think of the many tabs, dropdowns, charts, sliders, pop-ups and other UI elements involved in the SaaS platform of one's choice—can be replaced by simple language or voice commands. Theoretically, any software user experience can now be reduced to "talking" with an agent, who can fetch the information one needs and take action based on that information. This productivity benefit can barely be overstated, when one considers the time it takes for workers to learn and master new interfaces and tools.

How agentic AI works

Agentic AI tools can take many forms and different frameworks are better suited to different problems, but here are the general steps that agentic systems take to perform their operations.

Perception

Agentic AI begins by collecting data from its environment through sensors, APIs, databases or user interactions. This step ensures that the system has up-to-date information to analyze and act upon.

Reasoning

Once the data is collected, the AI processes it to extract meaningful insights. Using natural language processing (NLP), computer vision or other AI capabilities, it interprets user queries, detects patterns and understands the broader context. This ability helps the AI determine what actions to take based on the situation.

Goal setting

The AI sets objectives based on predefined goals or user inputs. It then develops a strategy to achieve these goals, often by using decision trees, reinforcement learning or other planning algorithms.

Decision-making

AI evaluates multiple possible actions and chooses the optimal one based on factors such as efficiency, accuracy and predicted outcomes. It might use probabilistic models, utility functions or machine learning-based reasoning to determine the best course of action.

Execution

After selecting an action, the AI executes it, either by interacting with external systems (APIs, data, robots) or providing responses to users.

Learning and adaptation

After executing an action, the AI evaluates the outcome, gathering feedback to improve future decisions. Through reinforcement learning or self-supervised learning, the AI refines its strategies over time, making it more effective in handling similar tasks in the future.

Orchestration

AI orchestration is the coordination and management of systems and agents. Orchestration platforms automate AI workflows, track progress toward task completion, manage resource usage, monitor data flow and memory and handle failure events. With the right architecture, dozens, hundreds or even thousands of agents could theoretically work together in harmonious productivity.

Examples of agentic AI

Agentic AI solutions can be deployed across virtually any AI use case in any real-world ecosystem. Agents can integrate within complex workflows to perform business processes autonomously.

- An AI-powered trading bot can analyze live stock prices and economic indicators to perform predictive analytics and execute trades.
- In autonomous vehicles, real-time data sources such as GPS and sensor data can improve navigation and safety.
- In healthcare, agents can monitor patient data, adjust treatment recommendations based on new test results and provide real-time feedback to clinicians through chatbots.
- In cybersecurity, agents can continuously monitor network traffic, system logs and user behavior for anomalies that might indicate vulnerabilities to malware, phishing attacks or unauthorized access attempts.
- AI can streamline supply chain management through process automation and optimization, autonomously placing orders with suppliers or adjusting production schedules to maintain optimal inventory levels.

Challenges for agentic AI systems

Agentic AI systems have massive potential for the enterprise. Their autonomy is their primary benefit, but this autonomous nature can bring serious consequences if agentic systems go "off the rails." The usual AI risks apply, but can be magnified in agentic systems.

Many agentic AI systems use reinforcement learning, which involves maximizing a reward function. If the reward system is poorly designed, the AI might exploit loopholes to achieve "high scores" in unintended ways.

Consider a few examples:

- An agent tasked with maximizing social media engagement that prioritizes sensational or misleading content, inadvertently spreading misinformation
- A warehouse robot optimizing for speed that damages products to move faster.
- A financial trading AI meant to maximize profits that engages in risky or unethical trading practices, triggering market instability.
- A content moderation AI designed to reduce harmful speech overcensors legitimate discussions.

Some agentic AI systems can become self-reinforcing, escalating behaviors in an unintended direction. This issue happens when the AI optimizes too aggressively for a particular metric without safeguards. And because agentic systems are often composed of multiple autonomous agents working together, there are opportunities for failure. Traffic jams, bottlenecks, resource conflicts—all of these errors have the potential to cascade. It's important for models to have clearly-defined goals that can be measured, with feedback loops in place so models can move ever closer to the organization's intention over time.