Types of AI agents

Types of AI agents

Artificial intelligence (AI) has transformed the way machines interact with the world, enabling them to perceive, reason and act intelligently. At the core of many AI systems are intelligent agents, autonomous entities that make decisions and perform tasks based on their environment.

These agents can range from simple rule-based systems to advanced learning systems powered by large language models (LLMs) that adapt and improve over time.

AI agents are classified based on their level of intelligence, decision-making processes and how they interact with their surroundings to reach wanted outcomes. Some agents operate purely on predefined rules, while others use learning algorithms to refine their behavior.

There are 5 main types of AI agents: simple reflex agents, model-based reflex agents, goal-based agents, utility-based agents and learning agents. Each type has distinct strengths and applications, ranging from basic automated systems to highly adaptable AI models.

All 5 types can be deployed together as part of a multi-agent system, with each agent specializing in handling the part of the task for which they are best suited.

Think Newsletter

Join over 100,000 subscribers who read the latest news in tech

Stay up to date on the most important—and intriguing—industry trends on AI, automation, data and beyond with the Think newsletter. See the IBM Privacy Statement.
We use your email to validate you are who you say you are, to create your IBMid, and to contact you for account related matters.
Business email

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.

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

Simple reflex agents

A simple reflex agent is the most basic type of AI agent, designed to operate based on direct responses to environmental conditions. These agents follow predefined rules, known as condition-action rules, to make decisions without considering past experiences or future consequences.

Reflex agents apply current perceptions of the environment through sensors and take action based on a fixed set of rules.

For example, a thermostat is a simple reflex agent that turns on the heater if the temperature drops below a certain threshold and turns it off when the wanted temperature is reached. Similarly, an automatic traffic light system changes signals based on traffic sensor inputs, without remembering past states.

Simple reflex agents are effective in structured and predictable environments where the rules are well-defined. However, they struggle in dynamic or complex scenarios that require memory, learning or long-term planning.

Because they do not store past information, they can repeatedly make the same mistakes if the predefined rules are insufficient for handling new situations.

Model-based reflex agents

A model-based reflex agent is a more advanced version of the simple reflex agent. While it still relies on condition-action rules to make decisions, it also incorporates an internal model of the world. This model helps the agent track the current state of the environment and understand how past interactions might have impacted it, allowing it to make more informed decisions.

Unlike simple reflex agents, which respond solely to current sensory input, model-based reflex agents use their internal model to reason about the environment's dynamics and make decisions accordingly.

For instance, a robot navigating a room might not just react to obstacles in its immediate path but also consider its previous movements and the locations of obstacles that it has already passed.

This ability to track past states enables model-based reflex agents to function more effectively in partially observable environments. They can handle situations where the context needs to be remembered and used for future decisions, making them more adaptable than simpler agents.

However, while model-based agents improve flexibility, they still lack the advanced reasoning or learning capabilities required for truly complex problems in dynamic environments.

Goal-based agents

A goal-based reflex agent extends the capabilities of a simple reflex agent by incorporating a proactive, goal-oriented approach to problem-solving.

Unlike reflex agents that react to environmental stimuli with predefined rules, goal-based agents consider their ultimate objectives and use planning and reasoning to choose actions that move them closer to achieving their goals.

These agents operate by setting a specific goal, which guides their actions. They evaluate different possible actions and select the one most likely to help them reach that goal.

For instance, a robot designed to navigate a building might have a goal of reaching a specific room. Rather than reacting to immediate obstacles only, it plans a path that minimizes detours and avoids known obstacles, based on a logical assessment of available choices.

The goal-based agent's ability to reason allows it to act with greater foresight compared to simpler reflex agents. It considers future states and their potential impact on reaching the goal.

However, goal-based agents can still be relatively limited in complexity compared to more advanced types, as they often rely on preprogrammed strategies or decision trees for evaluating goals.

Goal-based reflex agents are widely used in robotics, autonomous vehicles and complex simulation systems where reaching a clear objective is crucial, but real-time adaptation and decision-making are also necessary.

Utility-based agents

A utility-based reflex agent goes beyond simple goal achievement by using a utility function to evaluate and select actions that maximize overall benefit.

While goal-based agents choose actions based on whether they fulfill a specific objective, utility-based agents consider a range of possible outcomes and assign a utility value to each, helping them determine the most optimal course of action. This allows for more nuanced decision-making, particularly in situations where multiple goals or tradeoffs are involved.

For example, a self-driving car might face a decision to choose between speed, fuel efficiency and safety when navigating a route. Instead of just aiming to reach the destination, it evaluates each option based on utility functions, such as minimizing travel time, maximizing fuel economy or ensuring passenger safety. The agent selects the action with the highest overall utility score.

An e-commerce company might employ a utility-based agent to optimize pricing and recommend products. The agent evaluates various options, such as sales history, customer preferences and inventory levels to make informed decisions on how to price items dynamically.

Utility-based reflex agents are effective in dynamic and complex environments, where simple binary goal-based decisions might not be sufficient. They help balance competing objectives and adapt to changing conditions, ensuring more intelligent, flexible behavior.

However, creating accurate and reliable utility functions can be challenging, as it requires careful consideration of multiple factors and their impact on decision outcomes.

Learning agents

A learning agent improves its performance over time by adapting to new experiences and data. Unlike other AI agents, which rely on predefined rules or models, learning agents continuously update their behavior based on feedback from the environment. This allows them to enhance their decision-making abilities and perform better in dynamic and uncertain situations.

Learning agents typically consist of 4 main components:

1. **Performance element**: Makes decisions based on a knowledge base.

2. **Learning element**: Adjusts and improves the agent's knowledge based on feedback and experience.

3. **Critic**: Evaluates the agent's actions and provides feedback, often in the form of rewards or penalties.

4. **Problem generator**: Suggests exploratory actions to help the agent discover new strategies and improve its learning.

For example, in reinforcement learning, an agent might explore different strategies, receiving rewards for correct actions and penalties for incorrect ones. Over time, it learns which actions maximize its reward and refine its approach.

Learning agents are highly flexible and capable of handling complex, ever-changing environments. They are useful in applications such as autonomous driving, robotics and virtual assistants that assist human agents in customer support.

The ability to learn from interactions makes learning agents valuable for applications in fields such as persistent chatbots and social media, where natural language processing (NLP) analyzes user behavior to predict and optimize content recommendations.

Multi-agent systems

As AI systems become more intricate, the need for hierarchical agents arises. These agents are designed to break down complex problems into smaller, manageable subtasks, making it easier to handle complex problems in real-world scenarios. Higher-level agents focus on overarching goals, while lower-level agents handle more specific tasks.

An AI orchestration that integrates the different types of AI agents can make for a highly intelligent and adaptive multi-agent system capable of managing complex tasks across multiple domains.

Such a system can operate in real time, responding to dynamic environments while continuously improving its performance based on past experiences.

For example, in a smart factory, a smart management system might involve reflexive autonomous agents handling basic automation by responding to sensor inputs with predefined rules. These agents help ensure that machinery reacts instantly to environmental changes, such as shutting down a conveyor belt if a safety hazard is detected.

Meanwhile, model-based reflex agents maintain an internal model of the world, tracking the internal state of machines and adjusting their operations based on past interactions, such as recognizing maintenance needs before failure occurs.

At a higher level, goal-based agents drive the factory's specific goals, such as optimizing production schedules or reducing waste. These agents evaluate possible actions to determine the most effective way to achieve their objectives.

Utility-based agents further refine this process by considering multiple factors, such as energy consumption, cost efficiency and production speed, selecting actions that maximize expected utility.

Finally, learning agents continuously improve factory operations through reinforcement learning and machine learning (ML) techniques. They analyze data patterns, adapt workflows and suggest innovative strategies to optimize manufacturing efficiency.

By integrating all 5 types of AI agents, this AI-powered orchestration enhances decision-making

processes, streamlines resource allocation and minimizes human intervention, leading to a more intelligent and autonomous industrial system.

As agentic AI continues to evolve, advancements in generative AI (gen AI) will enhance the capabilities of AI agents across various industries. AI systems are becoming increasingly adept at handling complex use cases and improving customer experiences.

Whether in e-commerce, healthcare or robotics, AI agents are optimizing workflows, automating processes and enabling organizations to solve problems faster and more efficiently.