Unit 2 - Intelligent Agents

An **Intelligent Agent (IA)** is an autonomous entity — either a **software program** or a **robotic system** — that **perceives its environment** through **sensors**, **processes the information** using reasoning and decision-making mechanisms, and **takes actions** through **actuators** in order to achieve specific goals or optimize performance.

In simpler terms:

An intelligent agent is like a "digital or mechanical assistant" that can **observe**, **think**, and **act** to solve problems, adapt to changes, and work toward objectives — often without human intervention.

Key Characteristics

1. **Autonomy** – Can operate without continuous human control.
2. **Perception** – Uses sensors (physical or virtual) to gather information about the environment.
3. **Reasoning** – Processes information to make informed decisions.
4. **Adaptability** – Can modify its actions based on changes in the environment.
5. **Goal-Driven Behavior** – Works toward achieving specific objectives.
6. **Learning Capability** *(in advanced agents)* – Improves performance over time.

Examples of Intelligent Agents

1. **Google Search Engine (Software Agent)**
    - **Sensors:** Web crawlers collecting data from the internet.
    - **Reasoning:** Ranking algorithms deciding which results are most relevant.
    - **Actuators:** Displaying the search results to the user.
2. **Self-Driving Car (Robotic Agent)**
    - **Sensors:** Cameras, LiDAR, GPS.
    - **Reasoning:** AI algorithms decide when to accelerate, brake, or turn.

    - **Actuators:** Steering, throttle, and brake control systems.
3. **Smart Thermostat (Simple AI Agent)**
    - **Sensors:** Temperature and motion detectors.
    - **Reasoning:** Decides whether to turn heating or cooling on.
    - **Actuators:** HVAC control systems.
4. **Siri or Alexa (Conversational Agent)**
    - **Sensors:** Microphone input for voice commands.
    - **Reasoning:** Natural Language Processing (NLP) to understand requests.
    - **Actuators:** Voice responses, triggering apps, controlling smart devices.

Agents and Environments

An **agent** is anything that can **perceive its environment** using sensors and **act upon it** using actuators, with the aim of achieving a goal.
The **environment** is the external context in which the agent functions.

Definition:
An **agent** = **Perception → Decision-making → Action**
An **environment** = The surroundings where the agent operates.

Examples:

- **Human Agent:** Eyes, ears (sensors), brain (decision-making), hands, legs, mouth (actuators).
- **Robotic Agent:** Camera, ultrasonic sensor (sensors), CPU/microcontroller (decision-making), motors (actuators).

Rationality

Rationality means the agent chooses the **best possible action** to maximize performance, given:

1. The percept sequence so far.
2. The agent's built-in knowledge.
3. Available actions and expected outcomes.

Important:

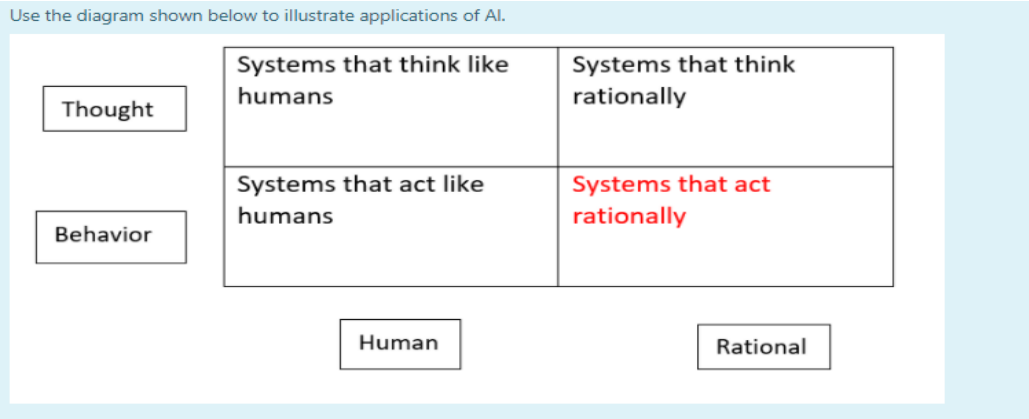
- Rationality ≠ Perfection
- A rational agent acts **optimally with the information it has** at the time.

Example:
A **self-driving car** in traffic chooses the safest and fastest route **based on current traffic updates**, not on unknown future conditions.

Four Approaches to AI

Artificial Intelligence can be classified into **four main approaches**, depending on whether we focus on **thinking or acting**, and whether we aim to **mimic humans** or **act rationally**.

This classification is represented in a **2×2 matrix**:



1. Systems that Think Like Humans

Goal: Make machines think like humans by simulating human thought processes.
Basis: Cognitive Science & Psychology.
Focus: Understanding how the human brain processes information, then replicating this in a machine.

Example:

- Cognitive modeling (e.g., ACT-R cognitive architecture).
- AI in brain-computer interfaces that imitate human memory recall.

Analogy: Teaching a computer to solve math problems the way you would do them in your head, step by step.

2. Systems that Think Rationally

Goal: Make machines think logically and make correct inferences.
Basis: Laws of Logic (Aristotelian logic).
Focus: Representing knowledge in logical form and deriving conclusions through reasoning.

Example:

- Expert systems (like MYCIN in medicine).
- Prolog-based AI systems for theorem proving.

Analogy: Like a mathematician who never makes calculation mistakes — always following logic perfectly.

3. Systems that Act Like Humans

Goal: Make machines behave like humans, regardless of whether they "think" the same way.
Basis: Turing Test (Alan Turing).
Focus: Human-like interaction, communication, and behavior.

Example:

- Chatbots like ELIZA or Siri that interact naturally.
- Humanoid robots like Sophia that can talk, make facial expressions, and respond like humans.

Analogy: An actor who perfectly imitates another person’s behavior, so you can’t tell the difference.

4. Systems that Act Rationally

Goal: Make machines act to achieve the best possible outcome (or the best expected outcome).
Basis: Rational Agent Theory.
Focus: Acting in a way that maximizes performance based on available information.

Example:

- Self-driving cars deciding the safest and fastest route.
- AI in finance making optimal investment decisions.

Analogy: A chess-playing AI like AlphaZero, which always makes the best move given the situation.

PEAS Framework

The **PEAS** model helps define the **task environment** of an intelligent agent.

P – Performance Measure

- Defines success criteria for the agent.
- *Performance measure is the unit to define the success of an agent. Performance varies with agents based on their different precepts.*
- Example: For a vacuum cleaner – amount of dirt cleaned, time taken, energy efficiency.

E – Environment

- The setting in which the agent operates.
- *Environment is the surrounding of an agent at every instant. It keeps changing with time if the agent is set in motion.*
- Example: Different floor types, room sizes, obstacles.

A – Actuators

- The mechanisms through which the agent takes actions.
- *An actuator is a part of the agent that delivers the output of action to the environment*
- Example: Wheels, brushes, suction mechanism.

S – Sensors

- Devices that detect environmental conditions.
- *Sensors are the receptive parts of an agent that takes in the input for the agent.*
- Example: Dirt sensors, bump sensors, cameras.

Example: Self-Driving Taxi (PEAS Table)

Component	Description
P	Safety, legality, comfort, speed, fuel efficiency
E	Roads, traffic, pedestrians, weather
A	Steering wheel, brakes, accelerator
S	GPS, cameras, lidar, radar

Types of Environments in Artificial Intelligence

In AI, an **environment** refers to everything that lies outside the agent and with which the agent interacts.
The environment provides **inputs (percepts)** to the agent, and the agent responds with **outputs (actions)** using its sensors and actuators.

The type of environment greatly influences **how an AI agent is designed** and **how it functions**.
Let's go through each type in detail:

1. Fully Observable vs. Partially Observable

- **Fully Observable Environment:**
  The agent's sensors can access **complete and accurate information** about the environment at any given time.
  There is no hidden or unknown state.
  Example: Playing chess — all pieces and their positions are visible to both players at all times.
- **Partially Observable Environment:**
  The agent only has **limited or incomplete information** about the environment.
  Some parts of the state are hidden or uncertain.
  Example: Driving a car in fog — the driver (agent) can only see part of the road and must make decisions with incomplete visibility.

2. Deterministic vs. Stochastic

- **Deterministic Environment:**
  The outcome of an action is **completely predictable** and determined by the current state.
  Example: Solving a math equation — the result is fixed if the same steps are followed.
- **Stochastic Environment:**
  The outcome of an action is **not fully predictable** because there's randomness or uncertainty involved.

Example: Playing poker — even if you play perfectly, the next card you draw is uncertain.

3. Episodic vs. Sequential

- **Episodic Environment:**
  The agent's actions are divided into **independent episodes**.
  Each episode is separate, and the agent does not need to remember past actions.
  Example: Image classification — each image is processed independently without referring to past images.
- **Sequential Environment:**
  Current actions affect **future states and decisions**.
  The agent must consider the history of past actions.
  Example: Playing a video game — each move affects the future outcome.

4. Static vs. Dynamic

- **Static Environment:**
  The environment does not change while the agent is deciding what to do.
  Example: A crossword puzzle — the puzzle remains the same while you are solving it.
- **Dynamic Environment:**
  The environment **changes over time**, possibly while the agent is thinking.
  Example: Stock market — prices change continuously, even while an investor is deciding to buy or sell.

5. Discrete vs. Continuous

- **Discrete Environment:**
  Has a **finite set** of states, actions, and percepts.
  Example: Chess — limited number of moves at any turn.
- **Continuous Environment:**
  Has an **infinite number** of possible states, actions, and percepts.
  Example: Driving a car — steering angle, speed, and road conditions can vary continuously.

6. Single-Agent vs. Multi-Agent

- **Single-Agent Environment:**
  Only one agent is operating in the environment.
  Example: Solving a Sudoku puzzle — you are the only decision-maker.

- **Multi-Agent Environment:**
  Multiple agents are involved, which may cooperate or compete.
  Example: A football match — each player (agent) acts, and decisions depend on others' actions.

Types of Agents in Artificial Intelligence

In AI, **agents** differ in complexity and decision-making capability. The classification depends on **how they sense, think, and act**.

1. Simple Reflex Agents

Definition:

- Act **only** based on the **current percept** (what they sense right now).
- Do **not** store any past information.
- Follow **condition–action rules** ("IF condition THEN action").

How They Work:

- Sense → Match condition → Execute action.
- No thinking about past or future.

Example:

- **Thermostat:**
  - IF temperature < 20°C → Turn heater ON.
  - IF temperature ≥ 20°C → Turn heater OFF.
- **Traffic light controller:**
  - Changes lights based on a timer, without considering real-time traffic.

Analogy:
Like a **light switch sensor** that turns on when it detects motion, without remembering if it was on or off before.

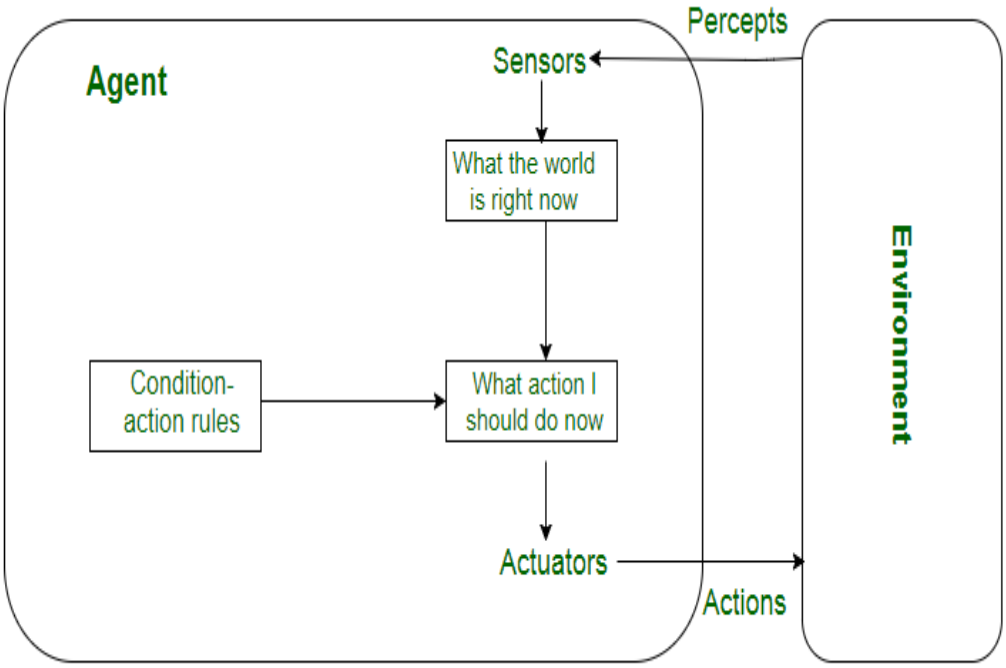
Advantages:

- Fast decision-making.
- Simple design.

Limitations:

- Cannot deal with partial observability.

- Fails in changing environments.



Simple Reflex Agent

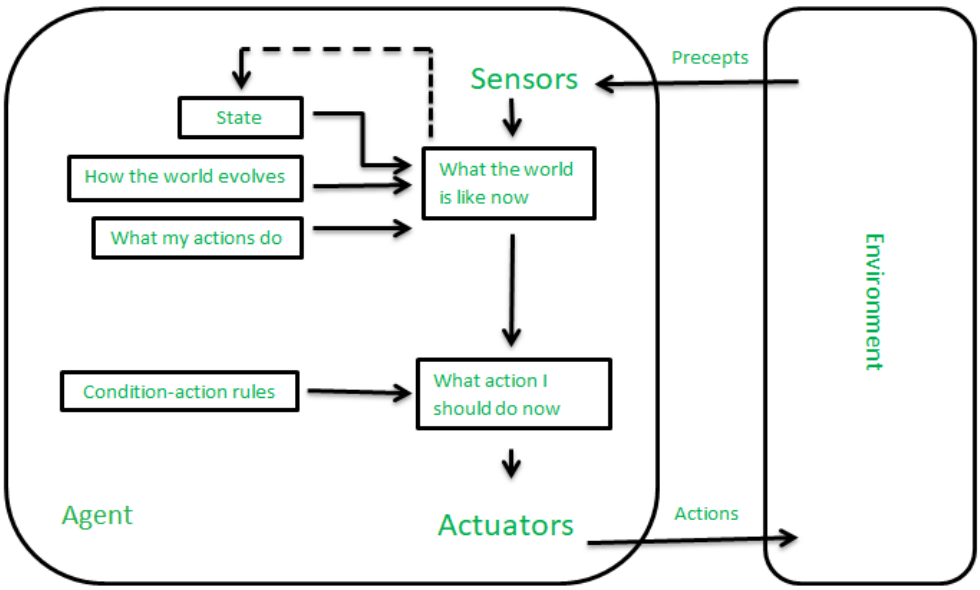
2. Model-Based Reflex Agents

Definition:

- Have an **internal model** of the environment to remember past states.
- Can handle **partially observable environments**.

How They Work:

- Store past percepts → Update model → Decide action.
- Model helps fill in missing information.



Example:

- **Robotic Vacuum Cleaner (Roomba):**
  - Remembers where it has already cleaned.
  - Avoids going over the same spot repeatedly.
- **Navigation Robot:**
  - Uses stored map to move around obstacles.

Analogy:
Like a **delivery person** who remembers which houses they have already delivered to.

Advantages:

- Works even if some info is missing.
- Can adapt to more complex environments.

Limitations:

- Requires storage and processing power.
- The model can become outdated if environment changes rapidly.

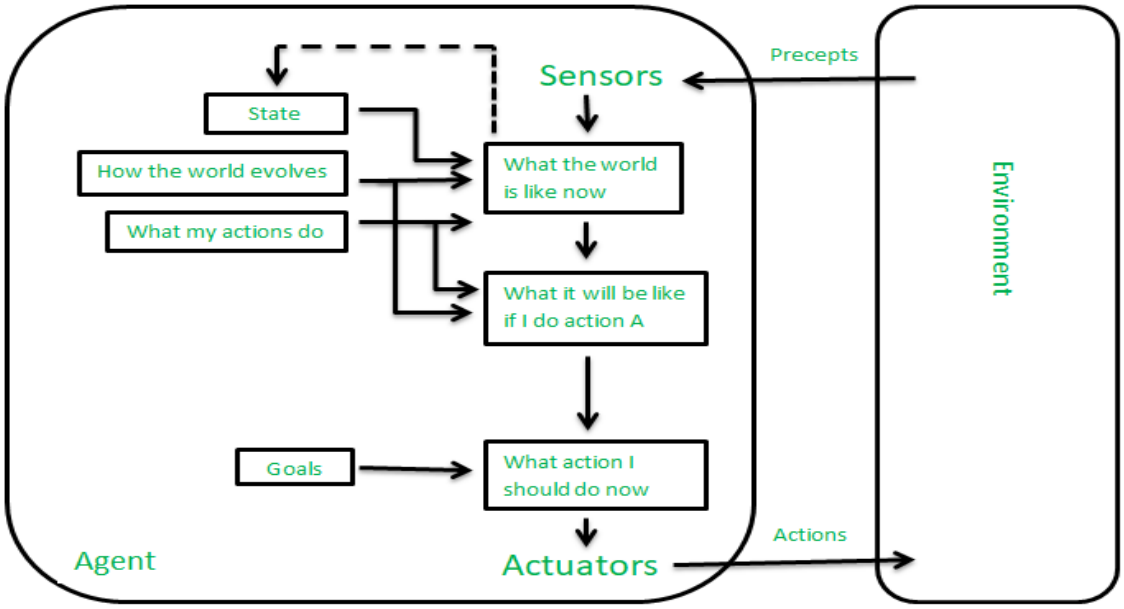
3. Goal-Based Agents

Definition:

- Choose actions to **achieve a specific goal**.
- Use **planning and search** to decide actions.

How They Work:

- Set goal → Compare possible actions → Choose best path toward goal.



Example:

- **GPS Navigation System:**
  - Goal: Reach destination in shortest time.
  - Calculates and suggests best route.
- **Chess-playing program:**
  - Goal: Win the game → Select moves to checkmate opponent.

Analogy:
Like a **student** deciding which subjects to study first to pass an exam — focuses on the end goal.

Advantages:

- Flexible for multiple goals.
- Can plan ahead.

Limitations:

- Planning may be slow in large environments.
- Needs a clear goal to work.

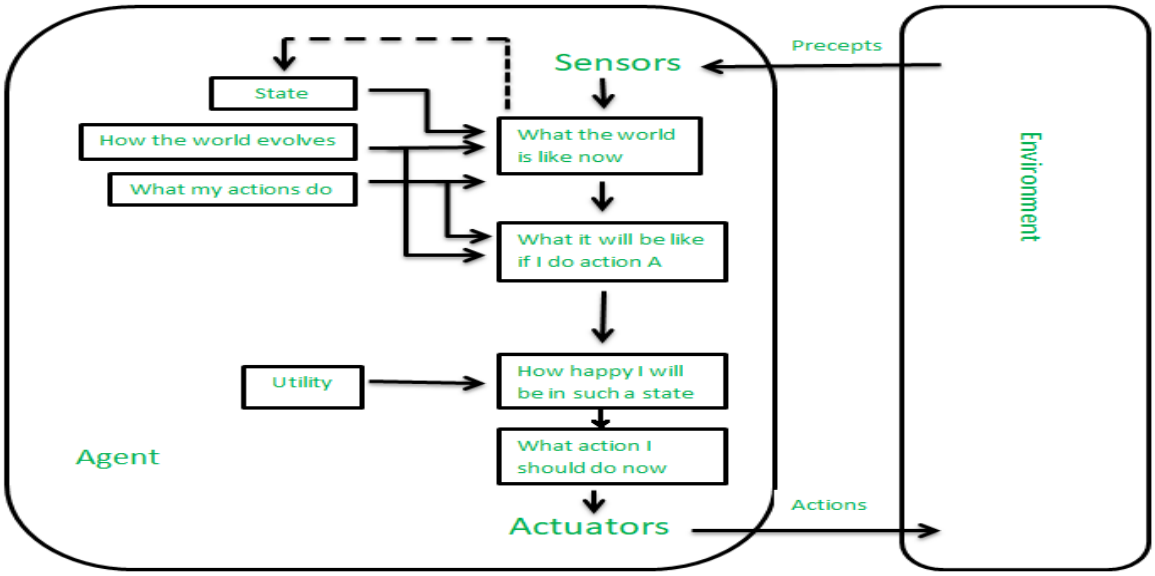
4. Utility-Based Agents

Definition:

- Make decisions based on **maximizing utility (happiness, satisfaction, efficiency)**.
- Consider trade-offs between goals.

How They Work:

- Assign value (utility) to outcomes → Choose action with highest utility.
- Works well for environments with **conflicting goals**.



Example:

- **Self-Driving Car:**
  - Balances speed and safety.
  - May choose a slightly longer route if it's safer.
- **Netflix Recommendation System:**
  - Shows movies based on highest probability of user liking them.

Analogy:
Like **choosing a holiday destination** — you weigh costs, weather, and enjoyment to pick the best option.

Advantages:

- Handles uncertainty well.

- Can prioritize better outcomes.

Limitations:

- Requires defining an accurate utility function.
- More complex to design.

5. Learning Agents

Definition:

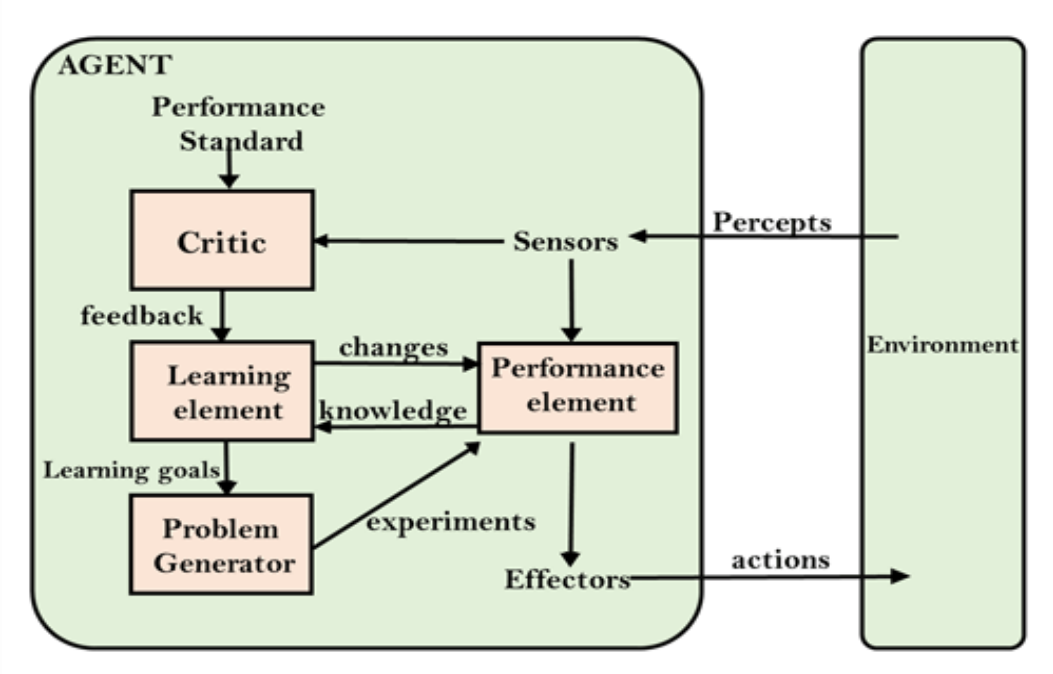
- Improve performance over time based on **experience**.
- Can adapt to new environments and situations.

How They Work:

- Learn from past actions and outcomes → Adjust future behavior.

Components of a Learning Agent:

1. **Learning Element:** Improves the agent over time.
2. **Performance Element:** Chooses actions.
3. **Critic:** Gives feedback.
4. **Problem Generator:** Suggests exploratory actions.



Example:

- **AlphaGo:** Learned to play Go by playing millions of games.

- **Spam Filter:** Learns from flagged emails to improve filtering.

Analogy:
Like a **chef** who keeps improving recipes based on customer feedback.

Advantages:

- Can handle dynamic, unknown environments.
- Becomes better with time.

Limitations:

- Requires training data and time.
- May learn bad behavior if given poor feedback.