Week-2

Rational Agents

Definition

A **rational agent** selects actions that **maximize its performance measure**, based on:

1. **Performance measure** — what success looks like
2. **Prior knowledge of the environment**
3. **Actions it can perform**
4. **Sequence of percepts** it has received

Rationality vs Omniscience

- Rationality = best decision *with current knowledge*
- Omniscience = perfect information about all outcomes (*not realistic*)
- Rational agents **maximize expected performance**, not guarantee perfect outcomes

  Real-life example:

  Stock-trading AI uses past data and probabilities — it's *rational* but not *omniscient*. It cannot know future market crashes perfectly.

Learning and Autonomy

- A **learning agent** updates behaviour based on consequences
- A **hard-coded agent** may be rational but lacks autonomy
- More learning → more autonomy

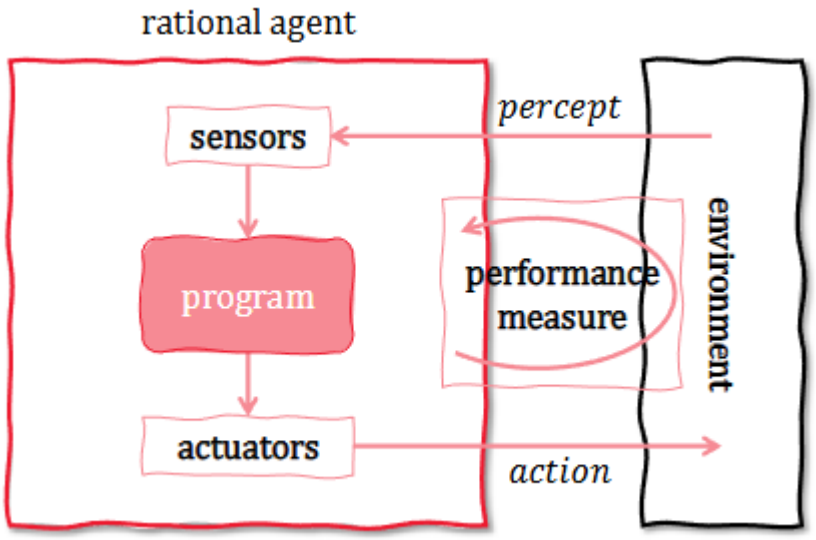
  Real-life example:

  Recommendation systems on Netflix learn your tastes over time versus a static TV schedule.

Agent Structure

Basic Elements

- **Environment** — where the agent operates
- **Sensors** — how the agent perceives the environment
- **Actuators** — how the agent acts



Agent Program

A function mapping **percept** → **action**

Example pseudocode:

```
def program(percept):
    state = interpret(percept)
    ifstate == case_0:
        return action_0
    ifstate == case_1:
        return action_1
```

Simple Reflex Agents

Definition → Agents that choose actions **only based on current percepts**, using **conditional rules**.

No memory of history.

Example:

A thermostat:

- Percept: current temp
- Rule: if temp < set point → heat ON

Current news example:

Basic factory robots still use reflex rules to avoid collisions — they don't plan ahead; they simply react.

Task Environments and PEAS

"Task environment" = the *problem* an agent is designed to solve.

It captures:

- **Performance measure**
- **Environment**
- **Sensors**
- **Actuators**

PEAS Framework (Exam)

PEAS stands for:

- **P**erformance measure → (How success is judged)
    - The performance measure defines the success criteria of an agent.
        - Performance measures should be *clearly defined*, *measurable*, and *relevant* to the task at hand.
- **E**nvironment → (Where the agent operates)
    - The environment is the *external* context in which an agent is operating.
        - The environment should include *all factors* that influence the agent's *perception* and *actions*.
- **A**ctuators → (How the agent acts)

    - Actuators are the mechanisms through which an agent interacts with its environment.
        - The choice of actuators is determined by the actions that are *required* and the agents *capabilities*.
- **S**ensors → (How the agent perceives)
    - Sensors are the means by which an agent is able to perceive its environment.
        - Choosing sensors depends on the information that is required by the agent to both *execute its actions* and evaluate its *performance*

Example: Self-driving car

- **Performance:** safety, speed, comfort
- **Environment:** roads, traffic
- **Actuators:** steering, braking, acceleration
- **Sensors:** cameras, lidar, GPS

Current news tie-in:

Debates over autonomous vehicle safety evaluations (e.g., Tesla Full Self-Driving) hinge on defining performance measures like *reducing fatalities* vs *maintaining traffic flow*.

Observability

Type	Meaning	Example
Fully observable	Agent can access the complete state of the environment through sensors	Chess
Partially observable	Agent has incomplete or noisy information about the environment	Poker, self-driving car in fog
Unobservable	Agent has no access to the state of the environment	Theoretical agent with no sensors

Number of agents

Type	Meaning	Example
Single-agent	Only one agent operates in the environment	Crossword puzzle

Type	Meaning	Example
Multi-agent	Multiple agents interact in the same environment	Football match, road traffic

Deterministic vs stochastic

Type	Meaning	Example
Deterministic	Next state is fully determined by current state and action	Sudoku
Stochastic	Randomness or uncertainty affects the next state	Stock market, robot on slippery floor

Episodic vs sequential

Type	Meaning	Example
Episodic	Each decision is independent of previous ones	Image classification
Sequential	Current actions affect future states and decisions	Chess, driving

Static vs dynamic

Type	Meaning	Example
Static	Environment does not change while the agent is deciding	Crossword puzzle
Dynamic	Environment can change while the agent is deciding	Driving in traffic

Discrete vs continuous

Type	Meaning	Example
Discrete	States, time, or actions are in distinct steps	Chess
Continuous	States, time, or actions vary smoothly over a range	Steering a car, robot arm movement

Summary

Slide Topic	Core Idea (Exam)	Real-Life / Context
Rational choice	Depends on 4 factors	AI assistants pick responses based on past data & rules
Rational agent	Maximizes performance measure	Self-checkout barcode scanners
Rational vs omniscient	Best with current evidence	Weather forecasting — probabilistic, not perfect
Learning vs autonomy	Learn → more adaptive	ChatGPT refining replies over interactions
Agent architecture	Percept → Action pipeline	Smart thermostat
Simple reflex agent	Immediate reaction	Automatic door sensors
Task environment / PEAS	Defines agent problem	Online shopping recommender specs