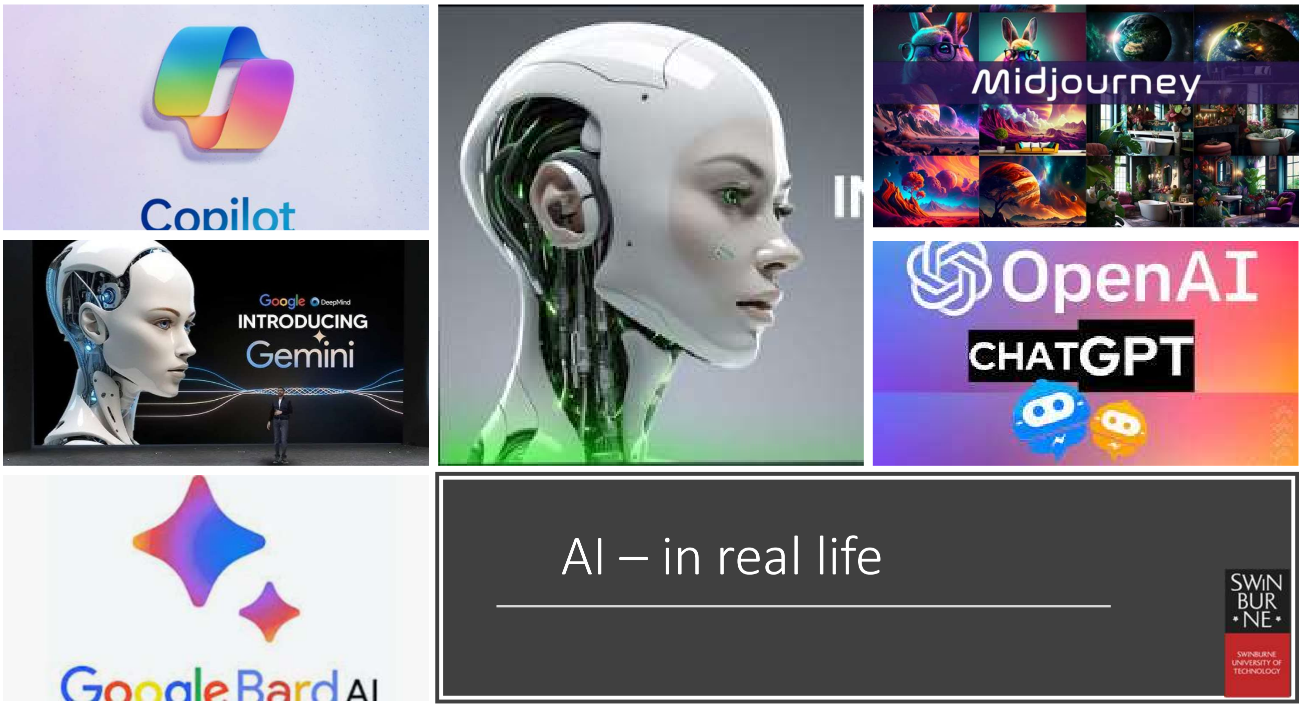
SWINBURNE UNIVERSITY OF TECHNOLOGY

COS30019 – Introduction to AI: Week 1 Introduction

AI – A pop-culture view

AI – in real life

AI – in real life



AI – in real life

What is AI? (Informally)

- Different people can define AI differently
- Most popular definitions:
  - “Artificial intelligence is a constellation of many different technologies working together to enable machines to sense, comprehend, act, and learn with **human-like** levels of intelligence.”
    (Accenture)
  - Artificial intelligence is the **simulation of human intelligence** processes by machines, especially computer systems/AI refers to systems or machines that **mimic human intelligence** to perform tasks and can iteratively improve themselves based on the information they collect.
    (SAS/Oracle)
  - AI refers to systems that acts **rationally** (aka. Intelligent agents): any system that perceives its environment and takes actions that maximize its chance of achieving its goals
    (AI textbooks)

Why study AI (from an academic perspective)

- It provides the core knowledge of computer science
- You'll learn to analyse problems and learn about techniques/algorithms to solve real-world problems
- It paves the way to understanding various sorts of intelligence (in both humans and machines)
- It is also fun (and different to most other subjects)

Why study AI? (from a career perspective)

- The market size in the AI market is projected to reach US$305.90bn in 2024. The market size is expected to show an annual growth rate (CAGR 2024-2030) of 15.83%, resulting in a market volume of US$738.80bn by 2030.

(Statista)

- The total contribution of AI to the global economy is expected to hit $15.7 trillion by 2030.

(PwC Global)

- AI will help boost the GDP of local economies, with China expected to record the greatest gains of 26% by 2030.

(PwC Global)

- The most in-demand AI job of 2023 can pay over US$200,000 and offers remote opportunities

(CNBC, Nov 2023)

- Searches for generative AI jobs on Indeed have increased almost 4,000% in the last year, and openings for generative AI jobs are up 306% over the same period.

(CNBC, Nov 2023)

What is Artificial Intelligence

- Different definitions due to different criteria
  - Two dimensions:
    - Thought processes/reasoning vs. behavior/action
    - Success according to human standards vs. success according to an ideal concept of intelligence: rationality.

Systems that think like humans	Systems that think rationally
Systems that act like humans	Systems that act rationally

  - Each definition falls under a **PARADIGM** in which AI can be built

Systems that act rationally

- Rational behavior: “doing the right thing”
  - The “Right thing” is the course of action that is expected to *maximize goal achievement given the available information*.
- Can include thinking, yet in service of rational action.
  - Action without thinking: e.g. reflexes.

Systems that act rationally (aka. Intelligent agents)

Systems that act rationally

- Two advantages over other approaches:
  - More general than law of thoughts approach
  - More amenable to scientific development (e.g., not constrained by how humans think/act).
- Yet, rationality is only applicable in *ideal* environments.
- Moreover, rationality may not be a good model of reality.

Foundations of AI

- Different fields have contributed to AI in the form of ideas, viewpoints and techniques.
  - *Philosophy*: Logic, reasoning, mind as a physical system, foundations of learning, language and rationality.
  - *Mathematics*: Formal representation and proof algorithms, computation, (un)decidability, (in)tractability, probability.
  - *Psychology*: adaptation, phenomena of perception and motor control.
  - *Economics*: formal theory of rational decisions, game theory.
  - *Linguistics*: knowledge representation, grammar.
  - *Neuroscience*: physical substrate for mental activities.
  - *Control theory*: homeostatic systems, stability, optimal agent design.

History of AI

Early computing

First Golden Age
1956 - 1974

AI Winter

Quiet Progress

Second Golden Age

Dartmouth Conference, 1956

"Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it"

History of AI

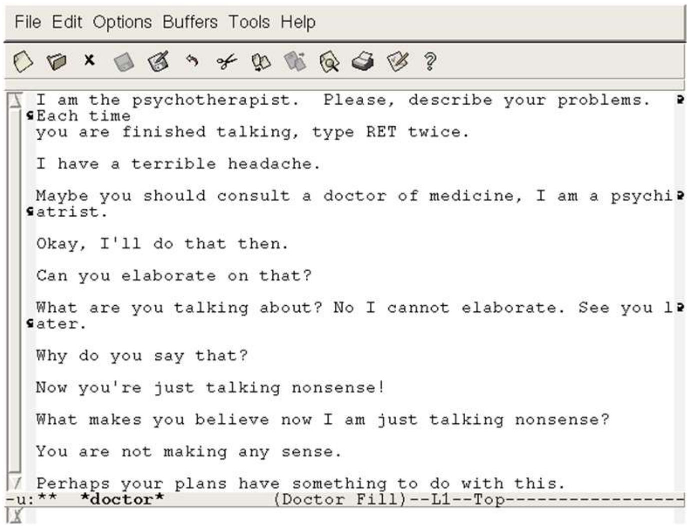
Early computing

First Golden Age
1956 - 1974

AI Winter

Quiet Progress

Second Golden Age



```
I am the psychotherapist.  Please, describe your problems.
Each time
you are finished talking, type RET twice.

I have a terrible headache.

Maybe you should consult a doctor of medicine, I am a psychi
atrist.

Okay, I'll do that then.

Can you elaborate on that?

What are you talking about? No I cannot elaborate. See you l
ater.

Why do you say that?

Now you're just talking nonsense!

What makes you believe now I am just talking nonsense?

You are not making any sense.

Perhaps your plans have something to do with this.
-u:**  *doctor*          (Doctor Fill)--L1--Top----------------
```

ELIZA, 1964

History of AI

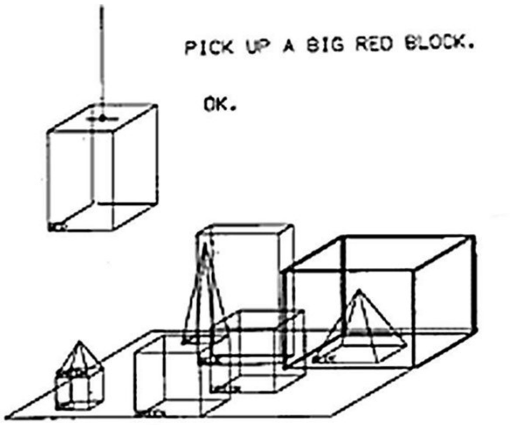
Early computing

First Golden Age
1956 - 1974

AI Winter

Quiet Progress

Second Golden Age



SHRDLU, 1968

History of AI

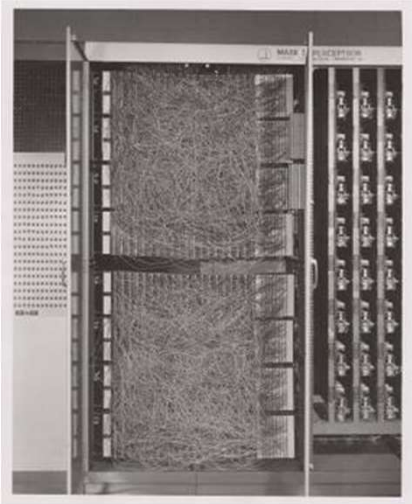
Early computing

First Golden Age
1956 - 1974

AI Winter

Quiet Progress

Second Golden Age



Perceptron, 1957

Attendance word: CHECK SLIDES

History of AI

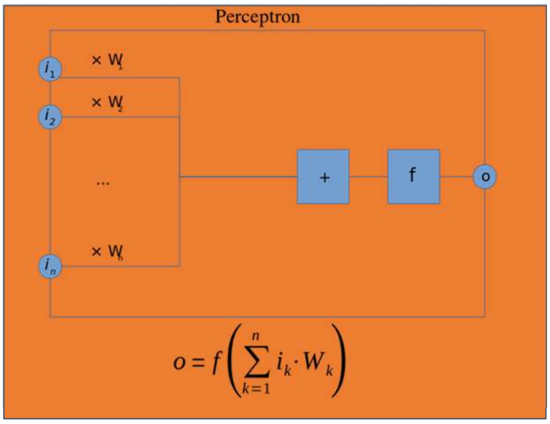
Early computing

First Golden Age
1956 - 1974

AI Winter

Quiet Progress

Second Golden Age



Perceptron, 1957

History of AI

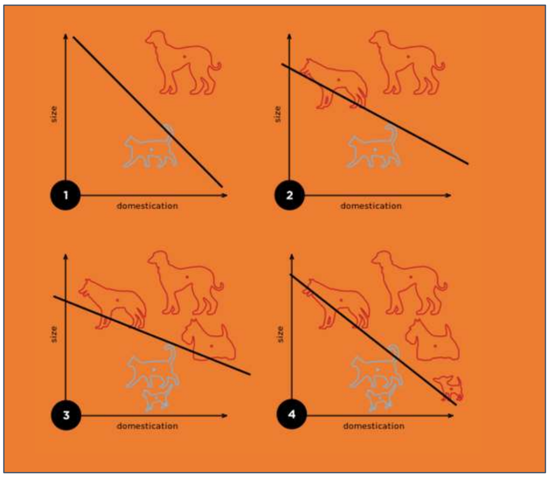
Early computing

First Golden Age
1956 - 1974

AI Winter

Quiet Progress

Second Golden Age



Perceptron, 1957

History of AI

Early computing

First Golden Age
1956 - 1974

AI Winter

Quiet Progress

Second Golden Age

Endless Optimism:

- 1965, H. A. Simon: "machines will be capable, within twenty years, of doing any work a man can do.
- 1970, Marvin Minsky (in Life Magazine): "In from three to eight years we will have a machine with the general intelligence of an average human being."
- 1976, Drew McDermott: "As a field, artificial intelligence has always been on the border of respectability, and therefore on the border of crackpottery. Many critics… have urged that we are over the border."

History of AI

Early computing

First Golden Age

AI Winter
1974 - 1990

Quiet Progress

Second Golden Age

- 1965: Dreyfus, “What computers can’t do”
- 1969: Minsky and Papert publish *Perceptrons*
- 1973: Lighthill Report
- 1974: DARPA cuts AI funding
- 1980: Searle’s Chinese Room Argument, in “Minds, brains, and programs”

History of AI

- Early computing
- First Golden Age
- AI Winter
- Quiet Progress 1990 - 2012
- Second Golden Age

Kasparov vs Deep Blue
May 11, 1997
(the day computers took over the world)

History of AI

- Early computing
- First Golden Age
- AI Winter
- Quiet Progress 1990 - 2012
- Second Golden Age

Watson vs Jeopardy 2011
Smartest Machine on Earth
(NOVA documentary)

History of AI

- Early computing
- First Golden Age
- AI Winter
- Quiet Progress 1990 - 2012
- Second Golden Age

Roomba, 2002, based on Rodney Brooks’ subsumption architecture

History of AI

- Early computing
- First Golden Age
- AI Winter
- Quiet Progress 1990 - 2012
- Second Golden Age

DARPA Grand Challenge, 2005
Great Robot Race (NOVA documentary)

History of AI

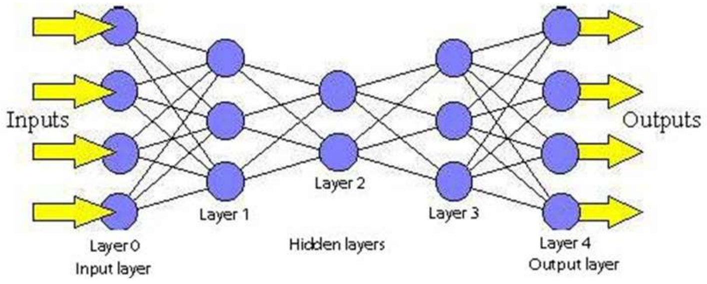
Early computing

First Golden Age

AI Winter

Quiet Progress
1990 - 2012

Second Golden Age



Back-propagation with hidden layers, 1986

History of AI

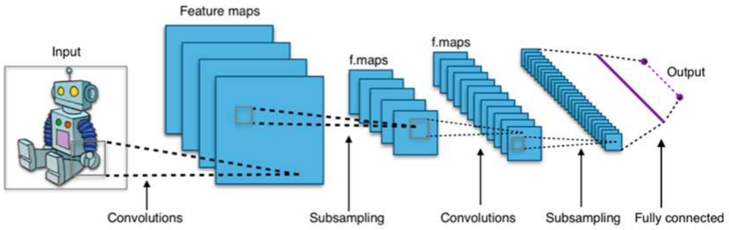
Early computing

First Golden Age

AI Winter

Quiet Progress
1990 - 2012

Second Golden Age



Yann LeCun, 1998
Convolutional Neural Networks

History of AI

Early computing

First Golden Age

AI Winter

Quiet Progress
1990 - 2012

Second Golden Age

Geoffrey Hinton talks @Google, 2007

History of AI

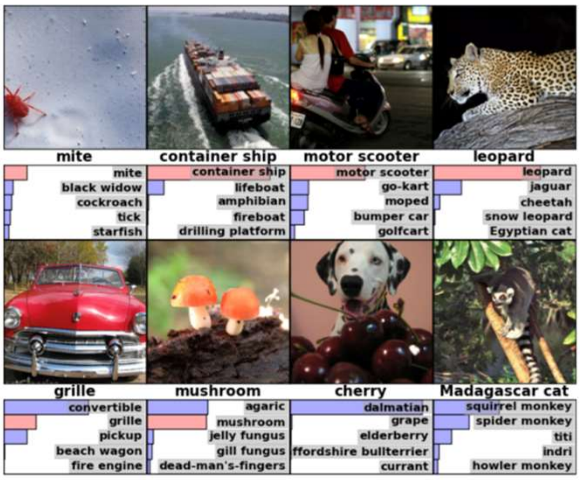
Early computing

First Golden Age

AI Winter

Quiet Progress

Second Golden Age
2012 - Present



Geoffrey Hinton beats ImageNet, 2012

History of AI

Early computing

First Golden Age

AI Winter

Quiet Progress

Second Golden Age
2012 - Present

- 2012: Hinton uses Deep ConvNets to solve ImageNet classification
- 2013: Hinton hired by Google, uses NN on mobile
    - Boston Dynamics acquired by Google
    - Yann LeCun heads FAIR for Facebook
- 2014: Google acquires DeepMind
- 2015: Musk donates $10m to Future of Life to study existential risk for artificial intelligence
    - Autonomous vehicle testing begins
- 2016: AlphaGo beats Lee Sedol
- 2017: AlphaZero beats beats Stockfish
- 2022: OpenAI releases ChatGPT
- … : OpenAI's GPT4o, Google's Bard AI, Google's Gemini, Meta's Llama, SORA, etc.

Different types of current AI systems

- Rule-based systems (e.g., expert systems, Mars Rover AI planner)
- Machine learning (ML)-based systems (e.g., IBM Watson, data analytics systems used by supermarkets, insurance and banking sector, telcos, etc.)
- Deep learning (DL)-based systems (e.g., those used in autonomous cars, Alpha Go, ChatGPT, etc.)

Some other terminologies

- Strong AI vs Weak AI
  - Weak AI:
    - Machines that can be made to act *as if* they were intelligent.
  - **Strong AI:**
    - **Machines that act intelligently with real, conscious minds**.

- Narrow AI vs Artificial General Intelligence (AGI)
  - Narrow AI:
    - Machine that is focused on one narrow (intellectual) task.
  - AGI:
    - Machine with the ability to apply intelligence to any problem, rather than just one specific problem.

Ethics and AI

- Ethical considerations in AI:
  - Bias
  - Privacy
  - job displacement
- Examples of ethical issues in AI
  - facial recognition
  - algorithmic hiring
- Potential solutions to ethical challenges
  - European Commission (EC)'s AI Act (AIA)
  - Responsible AI
  - Transparent & Auditable AI

Future of AI

- In your hand
  - You'll create it!
- Topic for discussion in class (tutorial)

Summary

- Different people think of AI differently.
- Two important questions to ask are:
- Are you concerned with thinking or behavior?
- Do you want to model humans or work from an ideal standard?
- In this course, we adopt the view that intelligence is concerned mainly with rational action.
- Ideally, an intelligent agent takes the best possible action in a situation. We will study the problem of building agents that are intelligent in this sense.

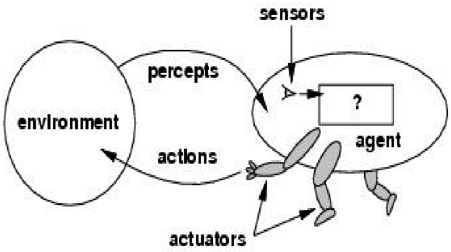
COS30019 – Introduction to AI: Week 1
Intelligent Agents

Outline

- Agents and environments.
  - The vacuum-cleaner world
- The concept of rational behavior.
- Environments.
- Agent structure.

Agents and environments

- Agents include human, robots, softbots, thermostats, etc.
- The *agent function* maps percept sequence to actions



- An agent can perceive its own actions, but not always it effects.

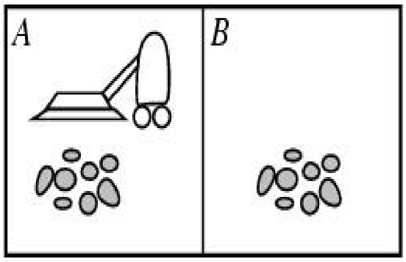
$$f : P^* \rightarrow A$$

To refresh your memory

- This is a mathematical function (square):
  - f: $\Re \rightarrow \Re$
  - $f(x) = x^2$
  - To explain this square function to primary school students, teacher use the following table:

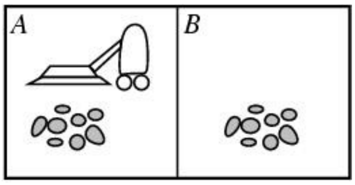
x	x^2
1	1
2	4
3	9
4	16
5	25
6	36
…	…

The vacuum-cleaner world – An example

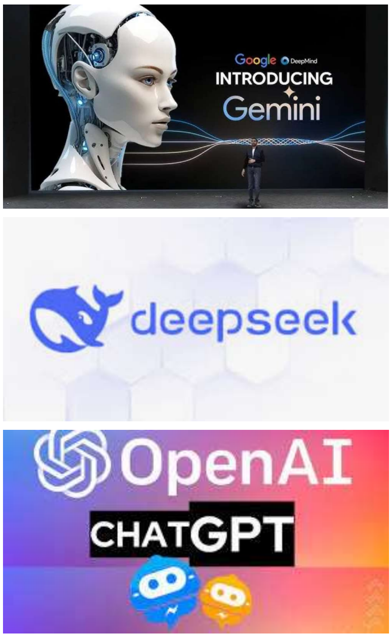


- **Environment:** squares A and B
- **Percepts:** [location and content] e.g. *[A, Dirty]*
- **Actions:** left, right, suck, and no-op

The vacuum-cleaner world – Agent function



Percept sequence	Action
[A, Clean]	Right
[A, Dirty]	Suck
[B, Clean]	Left
[B, Dirty]	Suck
[A, Clean],[A, Clean]	Right
[A, Clean], [A, Dirty]	Suck
…	…

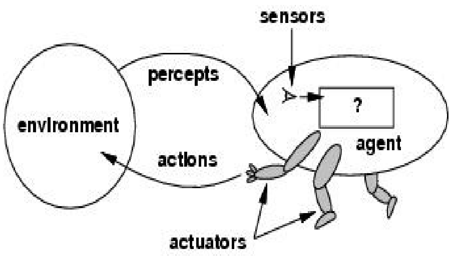


Hang on a second ….

- Why does this sound familiar?
- For Generative AI systems (e.g., ChatGPT, Gemini, DeepSeek, …):
  - The prompts ~ percept sequences
  - AI system outputs ~ Agent's actions
- Agentic AI

Agents and environments

- The *agent function* will internally be represented by the *agent program*.
- The agent program runs on the physical *architecture* to produce *f*.



The concept of rationality

- A **rational agent** is one that does the right thing.
  - Every entry in the table is filled out correctly.
- What is the right thing?
  - Approximation: the most *successful* agent.
  - *Measure of success?*
- Performance measure should be objective
  - E.g. the amount of dirt cleaned within a certain time.
  - E.g. how clean the floor is.
  - …
- ***Performance measure*** *according to what is wanted in the environment instead of how the agents should behave.*

Rationality

- What is rational at a given time depends on four things:
  - Performance measure,
  - Prior environment knowledge,
  - Actions,
  - Percept sequence to date (sensors).
- DEF: *A rational agent chooses whichever action that maximizes the expected value of the performance measure given the percept sequence to date and prior environment knowledge.*

Rationality

- The proposed definition requires:
  - Information gathering/exploration
    - To maximize future rewards
  - Learn from percepts
    - Extending prior knowledge
  - Agent autonomy
    - Compensate for incorrect prior knowledge

Is the vacuum cleaner agent rational?

- Depend!
- For example, it’s rational under the following assumptions:
  - Performance measure: 1 point for each clean square over ‘lifetime’ of 1000 steps
  - ‘geography’ known but dirt distribution, initial position of agent not known
  - Clean squares stay clean, sucking cleans squares
  - Left and Right don’t take agent outside environment
  - Available actions: Left, Right, Suck, NoOp
  - Agent knows where it is and whether that location contains dirt

- Let’s revisit this during the tutorial

Environments

- To design a rational agent, we must specify its **task environment**

- PEAS description of the task environment:
  - Performance
  - Environment
  - Actuators
  - Sensors

Environments

- E.g. Fully automated taxi:
  - PEAS description of the environment:
    - Performance
      - Safety, destination, profits, legality, comfort
    - Environment
      - Streets/freeways, other traffic, pedestrians, weather, …
    - Actuators
      - Steering, accelerating, brake, horn, speaker/display,…
    - Sensors
      - Video, sonar, speedometer, engine sensors, keyboard, GPS, …

Environment types

	Crossword	Backgammon	Chess w/ clock	Taxi
Observable??				
Deterministic??				
Episodic??				
Static??				
Discrete??				
Single-agent??				

The game of backgammon

Environment types

Fully vs. partially observable: an environment is full observable when the sensors can detect all aspects that are *relevant* to the choice of action.

	Crossword	Backgammon	Chess w/ clock	Taxi
Observable??				
Deterministic??				
Episodic??				
Static??				
Discrete??				
Single-agent??				

Environment types

Fully vs. partially observable: an environment is full observable when the sensors can detect all aspects that are *relevant* to the choice of action.

	Crossword	Backgammon	Chess w/ clock	Taxi
Observable??	FULL	FULL	FULL	PARTIAL
Deterministic??				
Episodic??				
Static??				
Discrete??				
Single-agent??				

Environment types

Deterministic vs. stochastic: if the next environment state is completely determined by the current state the executed action then the environment is deterministic.

	Crossword	Backgammon	Chess w/ clock	Taxi
Observable??	FULL	FULL	FULL	PARTIAL
Deterministic??				
Episodic??				
Static??				
Discrete??				
Single-agent??				

Environment types

Deterministic vs. stochastic: if the next environment state is completely determined by the current state the executed action then the environment is deterministic.

	Crossword	Backgammon	Chess w/ clock	Taxi
Observable??	FULL	FULL	FULL	PARTIAL
Deterministic??	YES	NO	YES	NO
Episodic??				
Static??				
Discrete??				
Single-agent??				

Environment types

Episodic vs. sequential: In an episodic environment the agent's experience can be divided into atomic steps where the agents perceives and then performs A single action. The choice of action depends only on the episode itself

	Crossword	Backgammon	Chess w/ clock	Taxi
Observable??	FULL	FULL	FULL	PARTIAL
Deterministic??	YES	NO	YES	NO
Episodic??				
Static??				
Discrete??				
Single-agent??				

Environment types

Episodic vs. sequential: In an episodic environment the agent's experience can be divided into atomic steps where the agents perceives and then performs A single action. The choice of action depends only on the episode itself

	Crossword	Backgammon	Chess w/ clock	Taxi
Observable??	FULL	FULL	FULL	PARTIAL
Deterministic??	YES	NO	YES	NO
Episodic??	NO	NO	NO	NO
Static??				
Discrete??				
Single-agent??				

Environment types

Static vs. dynamic: If the environment can change while the agent is choosing an action, the environment is dynamic. Semi-dynamic if the agent’s performance changes even when the environment remains the same.

	Crossword	Backgammon	Chess w/ clock	Taxi
Observable??	FULL	FULL	FULL	PARTIAL
Deterministic??	YES	NO	YES	NO
Episodic??	NO	NO	NO	NO
Static??				
Discrete??				
Single-agent??				

Environment types

Static vs. dynamic: If the environment can change while the agent is choosing an action, the environment is dynamic. Semi-dynamic if the agent’s performance changes even when the environment remains the same.

	Crossword	Backgammon	Chess w/ clock	Taxi
Observable??	FULL	FULL	FULL	PARTIAL
Deterministic??	YES	NO	YES	NO
Episodic??	NO	NO	NO	NO
Static??	YES	YES	SEMI	NO
Discrete??				
Single-agent??				

Environment types

Discrete vs. continuous: This distinction can be applied to the state of the environment, the way time is handled and to the percepts/actions of the agent.

	Crossword	Backgammon	Chess w/ clock	Taxi
Observable??	FULL	FULL	FULL	PARTIAL
Deterministic??	YES	NO	YES	NO
Episodic??	NO	NO	NO	NO
Static??	YES	YES	SEMI	NO
Discrete??				
Single-agent??				

Environment types

Discrete vs. continuous: This distinction can be applied to the state of the environment, the way time is handled and to the percepts/actions of the agent.

	Crossword	Backgammon	Chess w/ clock	Taxi
Observable??	FULL	FULL	FULL	PARTIAL
Deterministic??	YES	NO	YES	NO
Episodic??	NO	NO	NO	NO
Static??	YES	YES	SEMI	NO
Discrete??	YES	YES	YES	NO
Single-agent??				

Environment types

Single vs. multi-agent: Does the environment contain other agents who are also maximizing some performance measure that depends on the current agent's actions?

	Crossword	Backgammon	Chess w/ clock	Taxi
Observable??	FULL	FULL	FULL	PARTIAL
Deterministic??	YES	NO	YES	NO
Episodic??	NO	NO	NO	NO
Static??	YES	YES	SEMI	NO
Discrete??	YES	YES	YES	NO
Single-agent??				

Environment types

Single vs. multi-agent: Does the environment contain other agents who are also maximizing some performance measure that depends on the current agent's actions?

	Crossword	Backgammon	Chess w/ clock	Taxi
Observable??	FULL	FULL	FULL	PARTIAL
Deterministic??	YES	NO	YES	NO
Episodic??	NO	NO	NO	NO
Static??	YES	YES	SEMI	NO
Discrete??	YES	YES	YES	NO
Single-agent??	YES	NO	NO	NO

Environment types

- The simplest environment is
  - Fully observable, deterministic, episodic, static, discrete and single-agent.
- Most real situations are:
  - Partially observable, stochastic, sequential, dynamic, continuous and multi-agent.

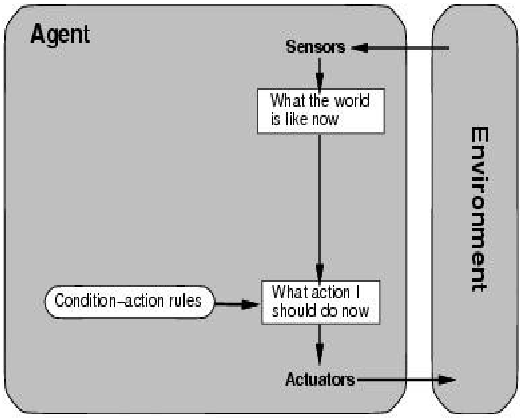
Agent types

- How does the inside of the agent work?
  - Agent = architecture + program
- All agents have the same skeleton:
  - Input = current percepts
  - Output = action
  - Program= manipulates input to produce output
- Note difference with agent function.

Agent types

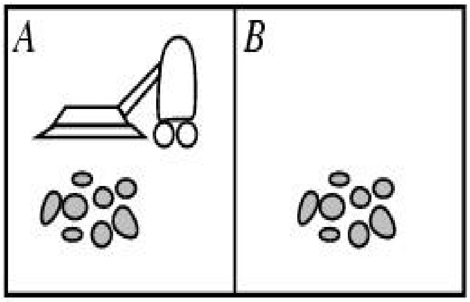
- Four **basic** kind of agent programs will be discussed:
  - Simple reflex agents
  - Model-based reflex agents
  - Goal-based agents
  - Utility-based agents
- All these can be turned into learning agents.
  - And that gives you four additional advanced agent types

Agent types; simple reflex



- Select action on the basis of *only the current* percept.
  - E.g. the vacuum-agent
- Large reduction in possible percept/action situations(next page).
- Implemented through *condition-action rules*
  - If dirty then suck

The vacuum-cleaner world



```
function REFLEX-VACUUM-AGENT ([location, status]) return an action
    if status == Dirty then return Suck
    else if location == A then return Right
    else if location == B then return Left
```

Reduction from 4^T to 4 entries

Agent types; simple reflex

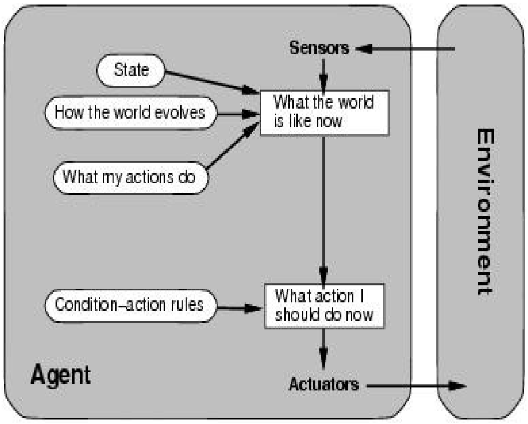
```
function SIMPLE-REFLEX-AGENT(percept) returns an action

    static: rules, a set of condition-action rules

    state ← INTERPRET-INPUT(percept)
    rule ← RULE-MATCH(state, rules)
    action ← RULE-ACTION[rule]
    return action
```

Will only work if the environment is *fully observable* otherwise infinite loops may occur.

Agent types; reflex and state



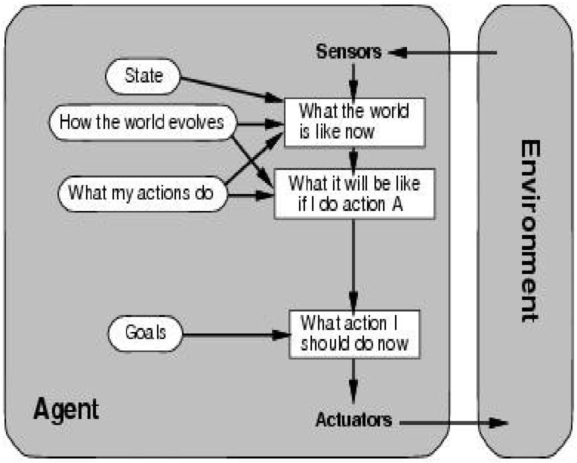
- To tackle *partially observable* environments.
  - Maintain internal state
- Over time update state using world knowledge
  - How does the world change.
  - How do actions affect world.

  ⇒ *Model of World*

Agent types; reflex and state

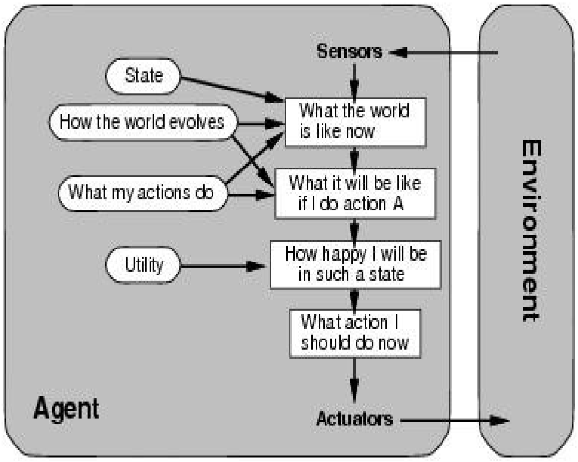
```
function REFLEX-AGENT-WITH-STATE(percept) returns an action
    static: rules, a set of condition-action rules
        state, a description of the current world state
        actions, the most recent actions.
    state ← UPDATE-STATE(state, actions, percept)
    rule ← RULE-MATCH(state, rules)
    action ← RULE-ACTION[rule]
    return action
```

Agent types; goal-based



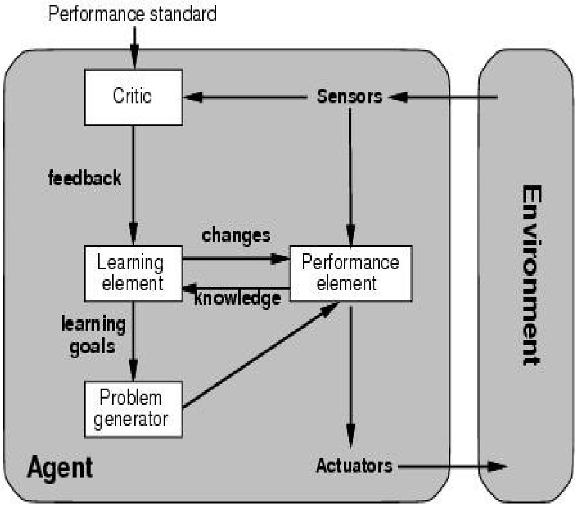
- The agent needs a goal to know which situations are *desirable*.
  - Things become difficult when long sequences of actions are required to find the goal.
- Typically investigated in **search** and **planning** research.
- Major difference: future is taken into account
- Is more flexible since knowledge is represented explicitly and can be manipulated.

Agent types; utility-based



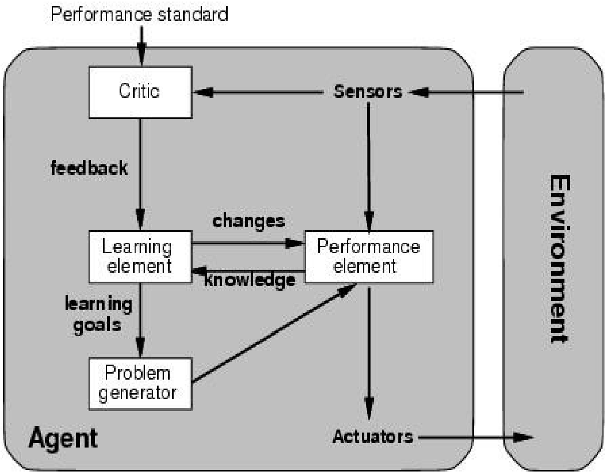
- Certain goals can be reached in different ways.
  - Some are better, have a higher utility.
- Utility function maps a (sequence of) state(s) onto a real number.
- Improves on goals:
  - Selecting between conflicting goals
  - Select appropriately between several goals based on likelihood of success.

Agent types; learning



- All previous agent-programs describe methods for selecting *actions*.
  - Yet it does not explain the origin of these programs.
  - Learning mechanisms can be used to perform this task.
  - Teach them instead of instructing them.
  - Advantage is the robustness of the program toward initially unknown environments.

Agent types; learning



- *Learning element*: introduce improvements in performance element.
  - Critic provides feedback on agents performance based on fixed performance standard.
- *Performance element*: selecting actions based on percepts.
  - Corresponds to the previous agent programs
- *Problem generator*: suggests actions that will lead to new and informative experiences.
  - Exploration vs. exploitation

Summary: Agents

- An **agent** perceives and acts in an environment, has an architecture, and is implemented by an agent program.
- Task environment – **PEAS (P**erformance**, E**nvironment, **A**ctuators, **S**ensors**)**
- An **ideal agent** always chooses the action which maximizes its expected performance, given its percept sequence so far.
- An **autonomous learning agent** uses its own experience rather than built-in knowledge of the environment by the designer.
- An **agent program** maps from percept to action and updates internal state.
  - **Reflex agents** respond immediately to percepts.
  - **Goal-based agents** act in order to achieve their goal(s).
  - **Utility-based agents** maximize their own utility function.
- **Representing knowledge** is important for successful agent design.
- The most challenging environments are not fully observable, nondeterministic, dynamic, and continuous

References/Readings

SWINBURNE UNIVERSITY OF TECHNOLOGY

1. https://deeperinsights.com/ai-blog/the-intelligent-agents-of-tomorrow-a-guide-to-llm-powered-agents#:~:text=LLM%2Dpowered%20agents%20function%20as,until%20the%20problem%20is%20resolved
2. https://bair.berkeley.edu/blog/2024/02/18/compound-ai-systems/