TRƯỜNG ĐẠI HỌC
SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH
HCMC University of Technology and Education
Faculty of Information Technology

Chapter 4
AI & Intelligence Systems
(Fundamental Concepts)

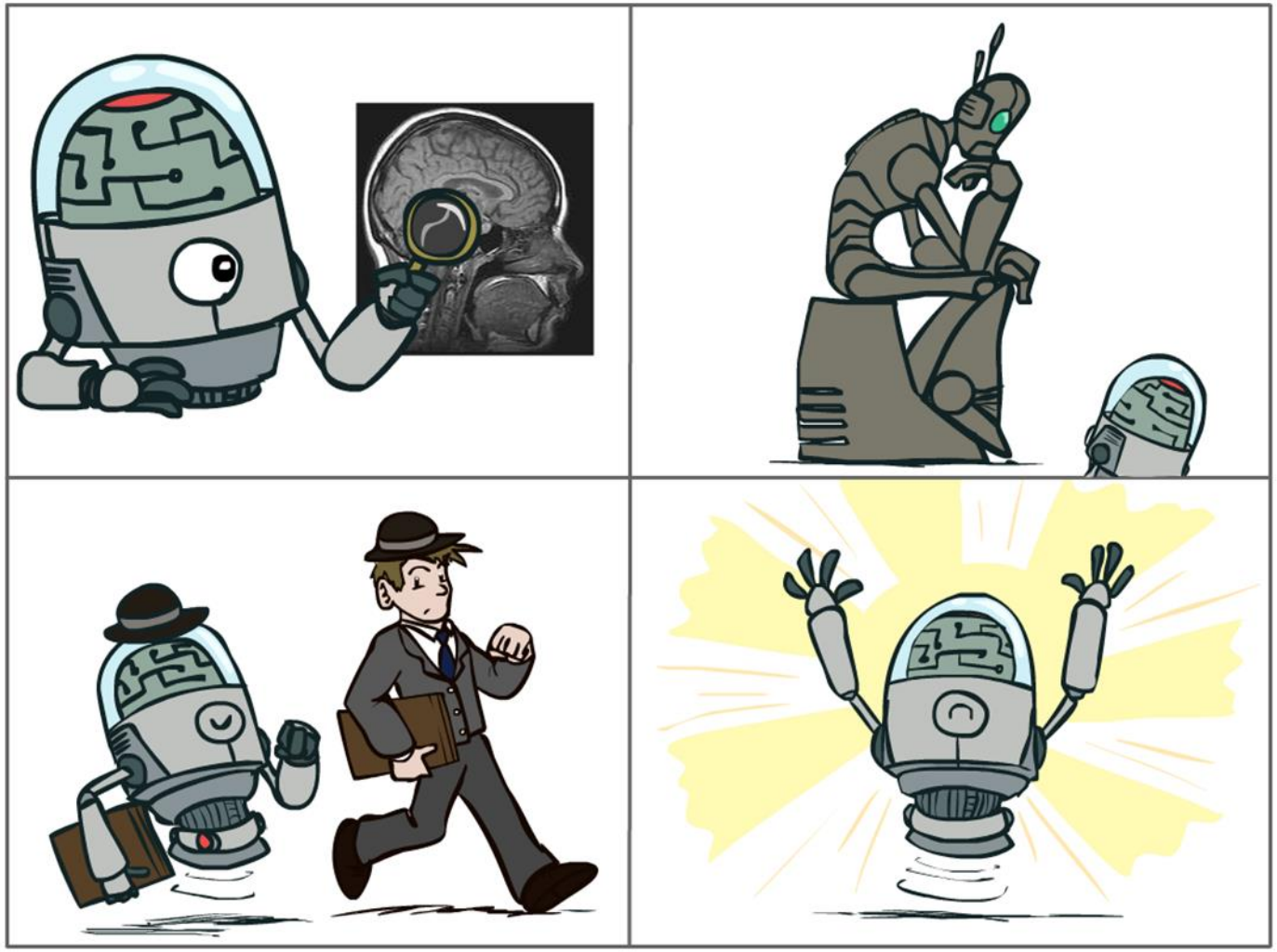
PGS.TS. Hoàng Văn Dũng

Email: dunghv@hcmute.edu.vn

What is AI?

The science of making machines that:

Think like people



Think rationally

Act like people

Act rationally

What is AI?

"Artificial intelligence (AI) refers to systems that display intelligent behaviour by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals.

AI-based systems can be purely software-based, acting in the virtual world (*e.g. voice assistants, image analysis software, search engines, speech and face recognition systems*) or AI can be embedded in hardware devices (*e.g. advanced robots, autonomous cars, drones or Internet of Things applications*)."

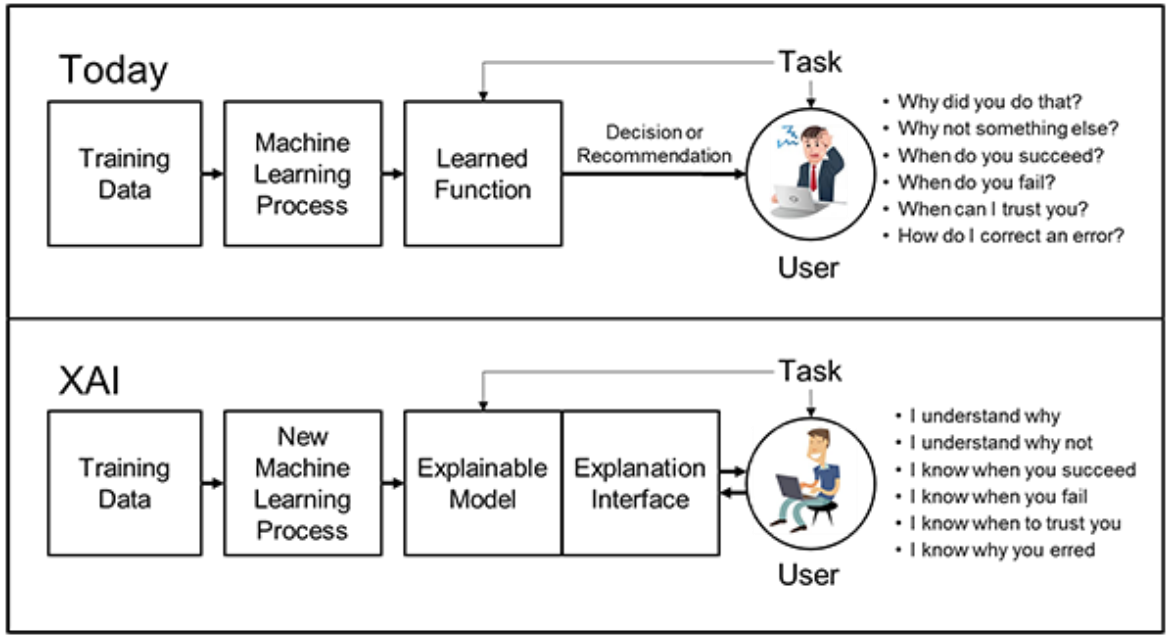
European Commission: https://digital-strategy.ec.europa.eu, Document made public on 8 April 2019.

What is AI?

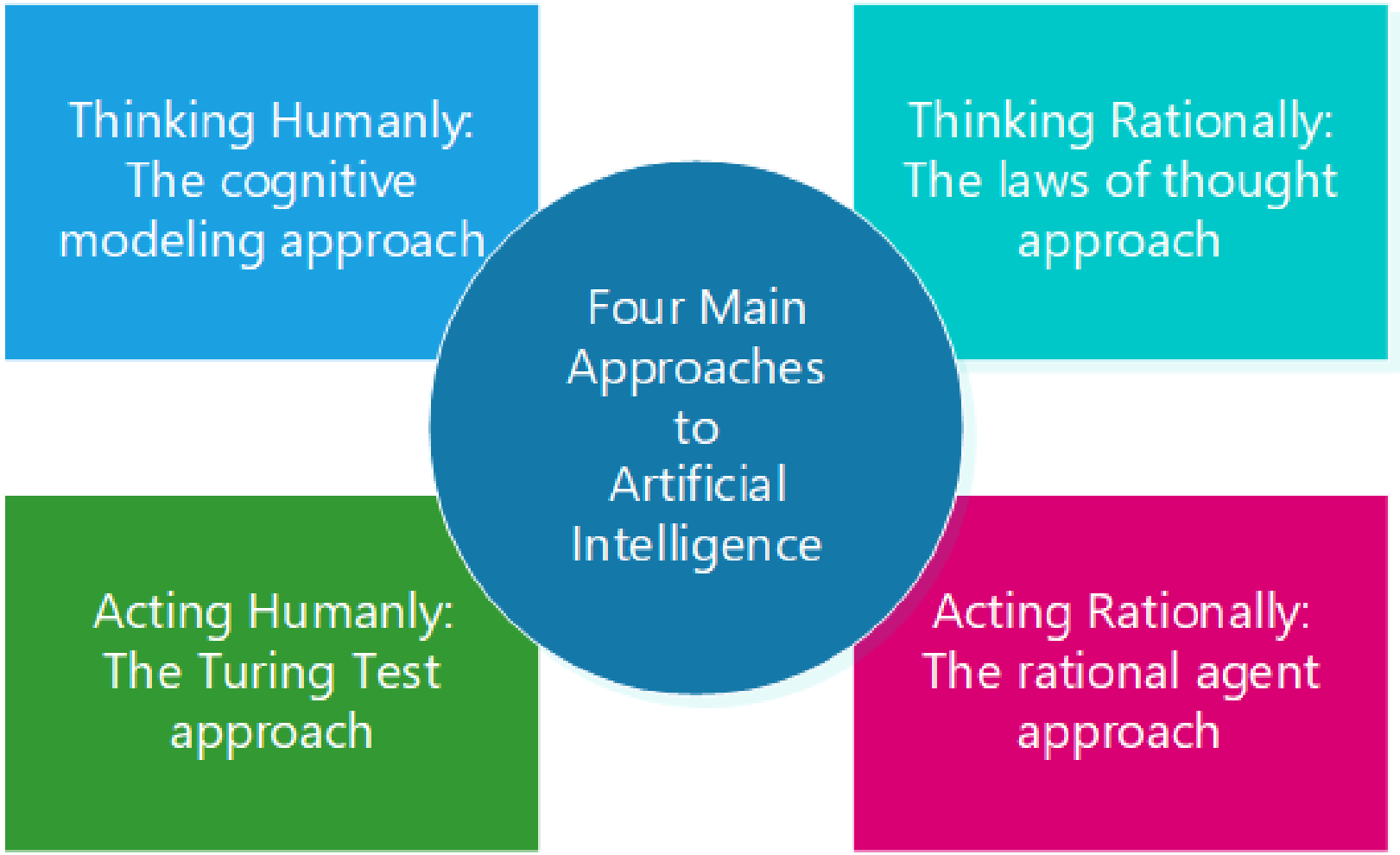
E**x**plainable **AI** (XAI)

*"New machine-learning systems will have the ability to **explain their rationale**, characterize their strengths and weaknesses, and **convey an understanding of how they will behave in the future**. […] These models will be combined with state-of-the-art human-computer interface techniques capable of translating models into understandable and useful **explanation dialogues for the end user**."*

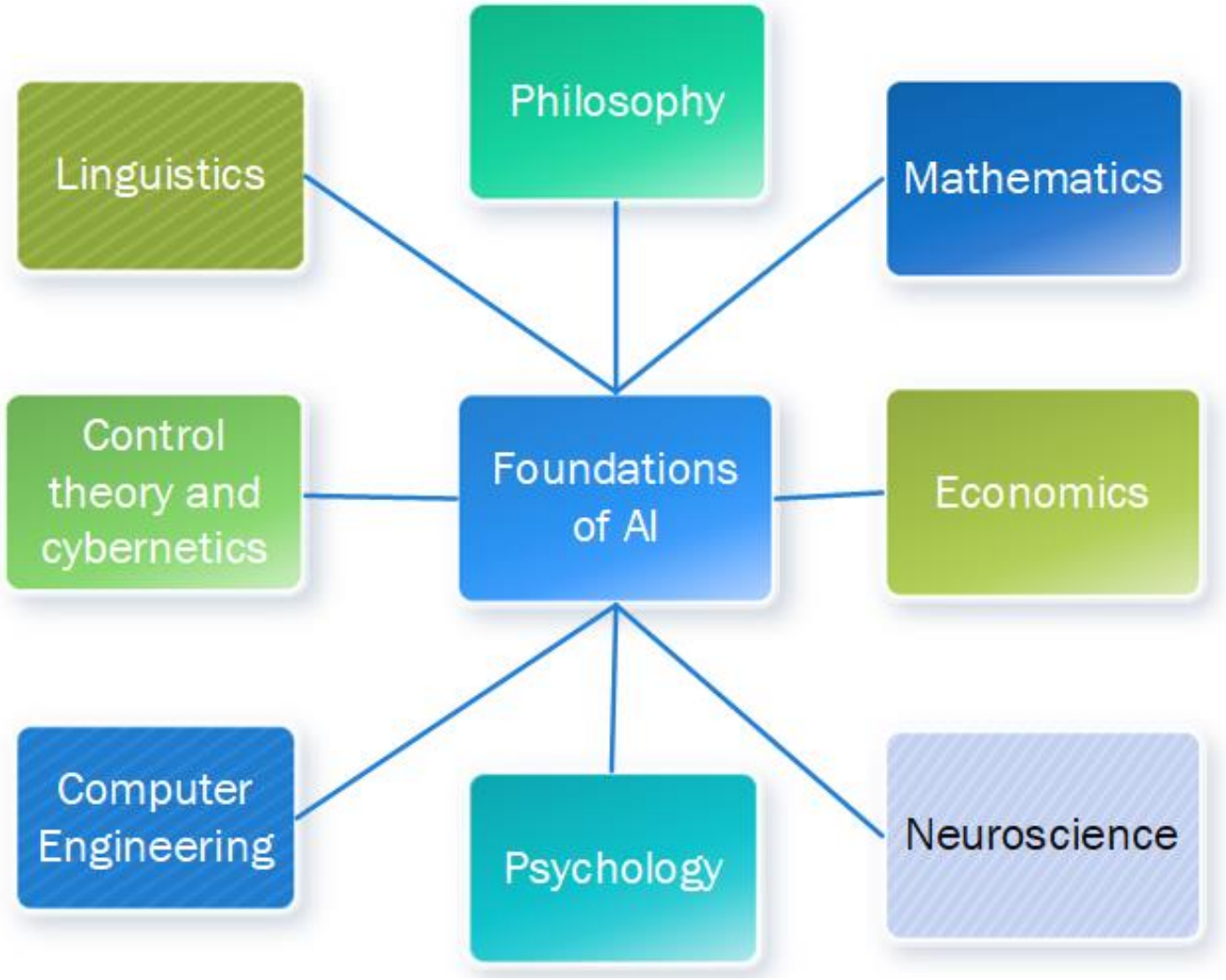
Source: DARPA, www.darpa.mil



Approaches to AI



Foundations of AI



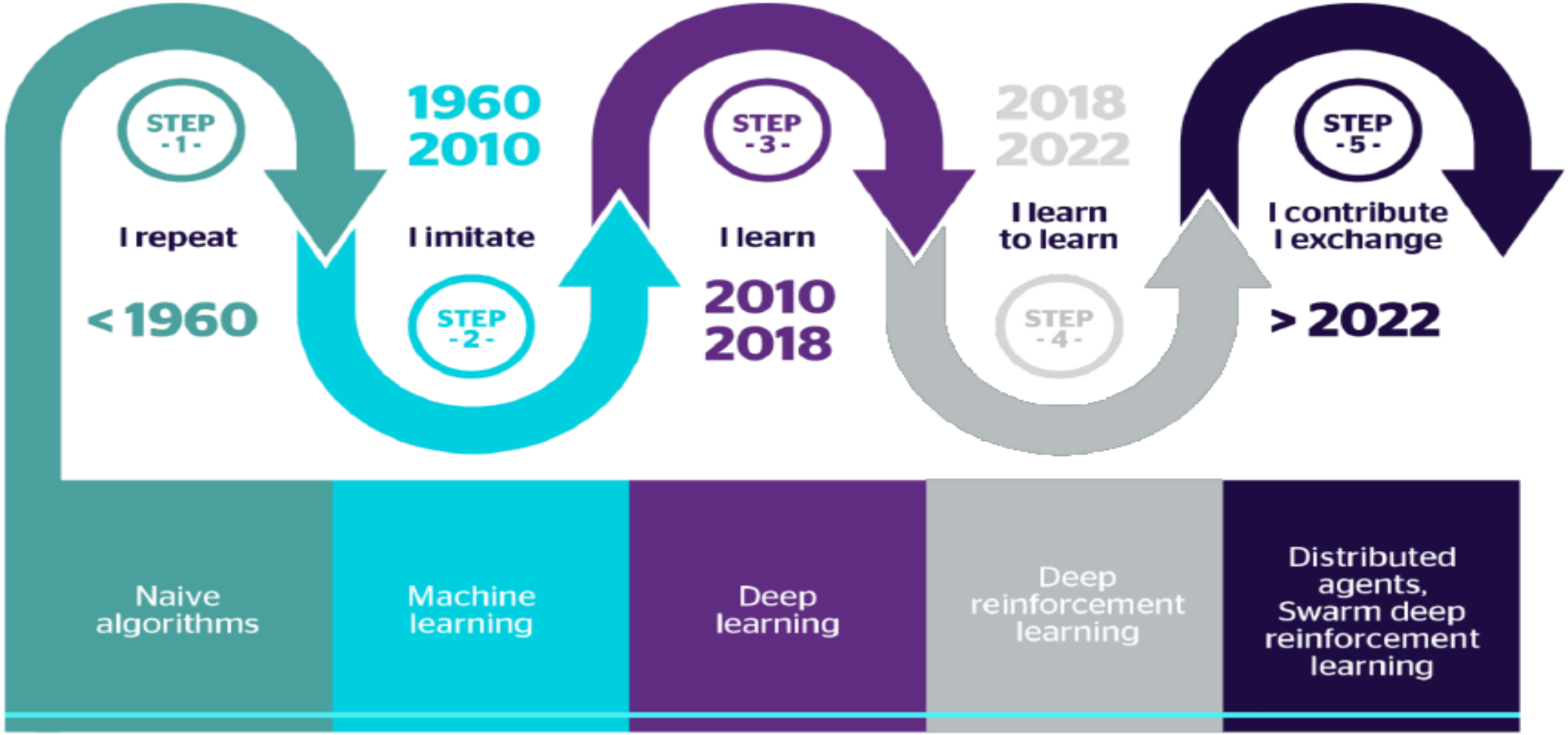
To learn more: Section 1.2 of (Russell, 2016)

Foundations of AI

• Philosophy	Logic, methods of reasoning, mind as physical system foundations of learning, language, rationality
• Mathematics	Formal representation and proof algorithms, computation, (un)decidability, (in)tractability, probability
• Economics	utility, decision theory
• Neuroscience	physical substrate for mental activity
• Psychology	phenomena of perception and motor control, experimental techniques
• Computer engineering	building fast computers
• Control theory	design systems that maximize an objective function over time
• Linguistics	knowledge representation, grammar

To learn more: Section 1.2 of (Russell, 2016)

A brief history of AI



https://atos.net/

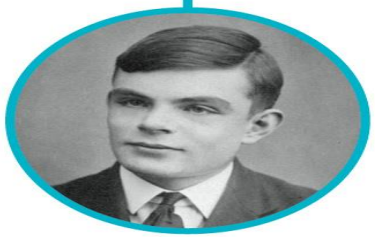
A Brief History of AI

1642 — First mechanical calculating machine built by French mathematician and inventor Blaise Pascal.

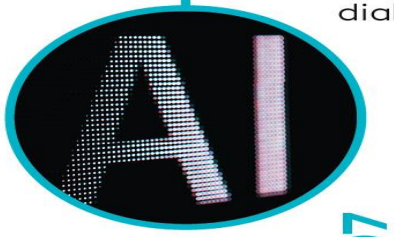
1837 — First design for a programmable machine, by Charles Babbage and Ada Lovelace.

1943 — Foundations of neural networks established by Warren McCulloch and Walter Pitts, drawing parallels between the brain and computing machines.

1950 — Alan Turing introduces a test—the Turing test—as a way of testing a machine's intelligence.



1955 — 'Artificial intelligence' is coined during a conference devoted to the topic.



1965 — ELIZA, a natural language program, is created. ELIZA handles dialogue on any topic; similar in concept to today's chatbots.

1980s — Edward Feigenbaum creates expert systems which emulate decisions of human experts.

1997 — Computer program Deep Blue beats world chess champion Garry Kasparov.

2002 — iRobot launches Roomba, an autonomous vacuum cleaner that avoids obstacles.

2009 — Google builds the first self-driving car to handle urban conditions.

2011 — IBM's Watson defeats champions of US game show Jeopardy!

2011-2014 — Personal assistants like Siri, Google Now, Cortana use speech recognition to answer questions and perform simple tasks.

2014 — Ian Goodfellow comes up with Generative Adversarial Networks (GAN).

2016 — AlphaGo beats professional Go player Lee Sedol 4-1.

2018 — Most universities have courses in Artificial Intelligence.

https://datalchemy.ai

AI achievements

- Automatic planning and scheduling: NASA has designed an automatic planning program (called Remote Agent) to control the scheduling of spacecraft activities
- Play chess: Deep Blue (IBM computer system) defeated World chess master Garry Kasparov in 1997
- Automatic control: A minivan is controlled automatically by the ALVINN system (from CMU) during 98% of the time traveling from Pittsburgh to San Diego (~2850 miles)
- Robotics: Today, many medical surgeries use robotic assistance in microsurgery operations.
- Diagnosis: Medical diagnosis programs based on probability analysis can already perform at the same level as specialists in some fields of medicine.
- Military logistics planning: During the 1991 Gulf War, U.S. military forces deployed a logistics planning and scheduling program to move 50,000 vehicles. , goods and soldiers , and soldiers
- Understand language and solve problems: The PROVERB computer program can solve crossword puzzles better than many people can

Debates about AI

The capabilities of AI?

•Play a game of table tennis? Driving on autopilot along a winding mountain road?

• Buy a week's worth of groceries online for a grocery store?

•Discovered and proved a new mathematical theory? Can you converse with one person for 1 hour? Automatically perform a complex surgery?

•Direct (instant) translation between two languages for a conversation?

• Can computers think (like humans)?

• Employment and unemployment issues

• People will have too much free time (compared to too little, like today)

• People feel a loss of their dominant (highest) intelligence

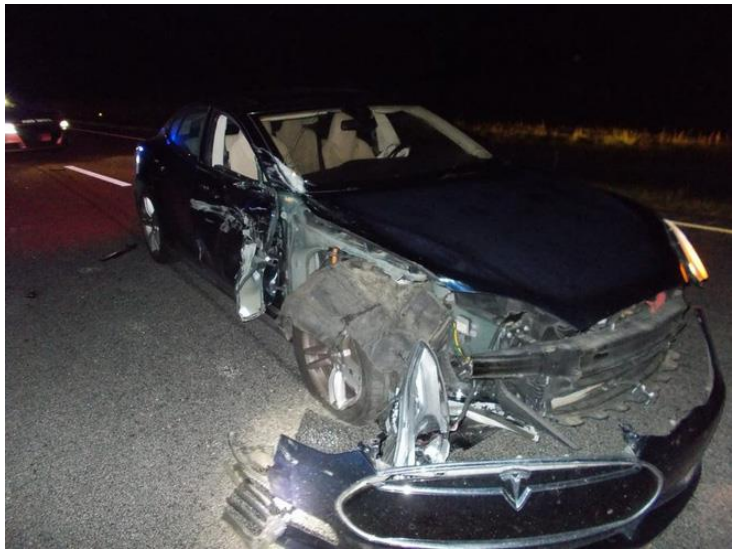
• Because computers do (and interfere) with many people's daily tasks, they will feel their privacy rights are violated.

• Using multiple AI systems can reduce (lose) accountability in tasks

•The (perfect) success of AI means the end of humanity

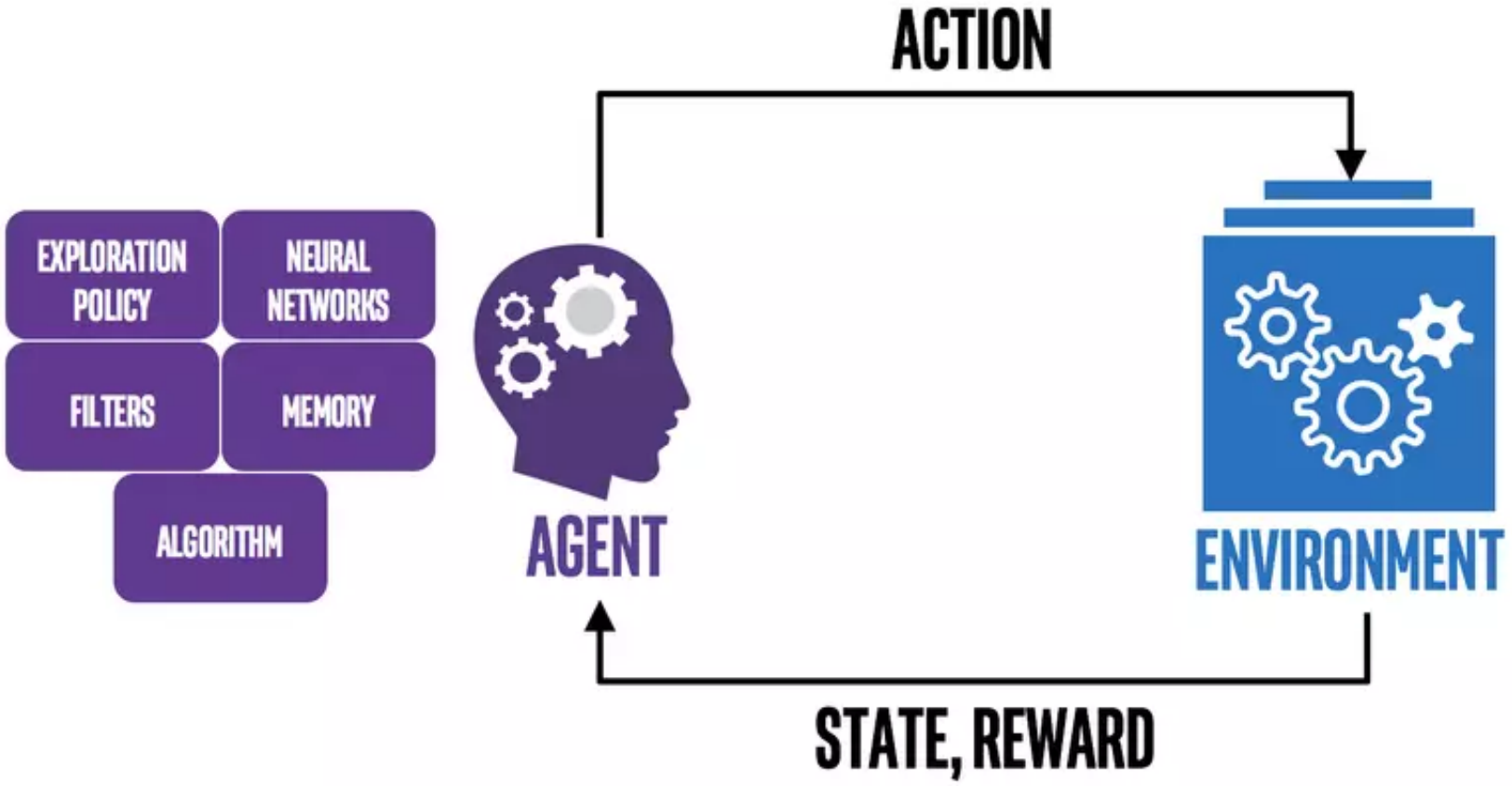
Quick assignment

- Given and analyzed 5 products of AI.
- Given and analyzed 5 issues which (AI can do it, can't do it, which ones are better than humans,...)



Basic concepts

- Agent – Environment

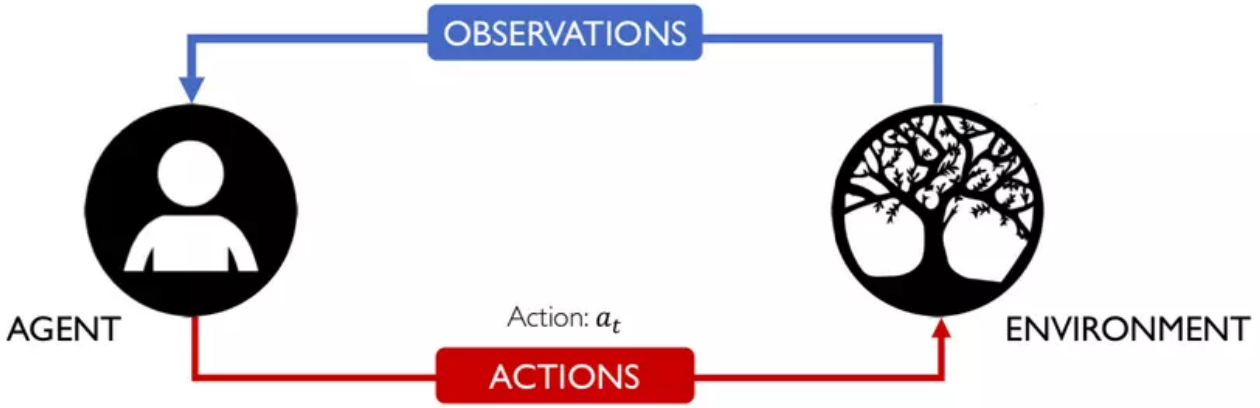


Basic concepts

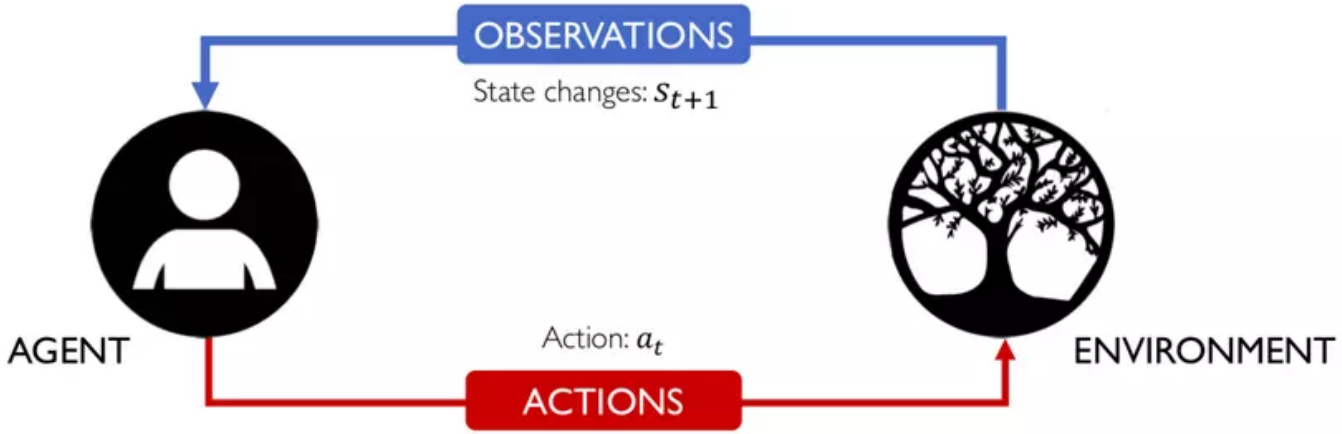
Environment: the world in which the agent exists and operates.

Action: a move the agent can make in the environment.

Basic concepts



Observations: of the environment after taking actions.



State: a situation which the agent perceives.

Basic concepts

Reward: feedback that measures the success or failure of the agent's action.

Total Reward

$$R_t = \sum_{i=t}^{\infty} r_i$$

Basic concepts

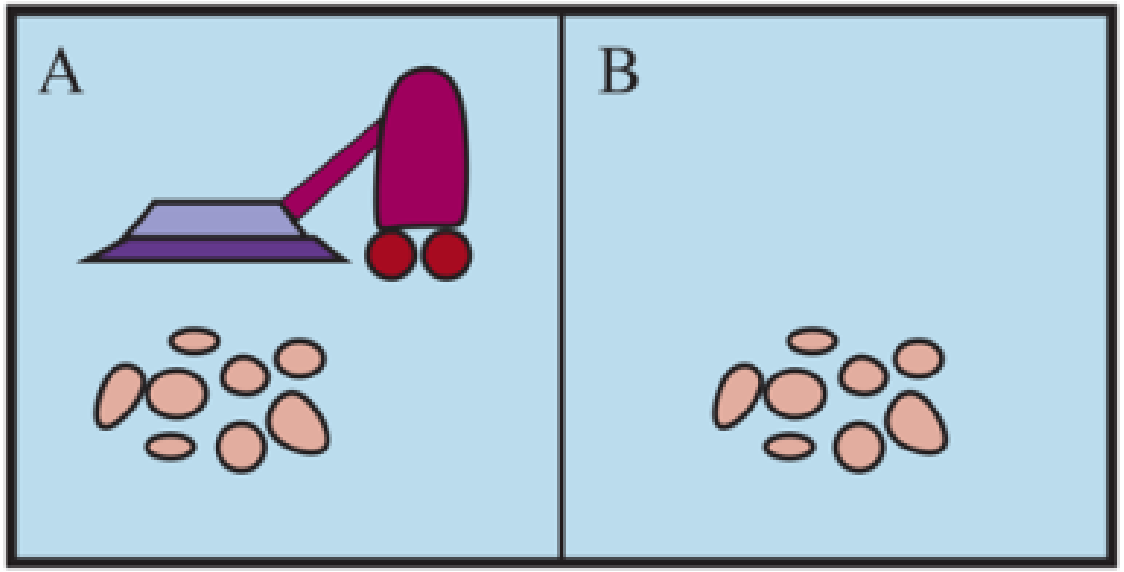
Discounted Total Reward

$$R_t = \sum_{i=t}^{\infty} \gamma^i r_i = \gamma^t r_t + \gamma^{t+1} r_{t+1} \ldots + \gamma^{t+n} r_{t+n} + \cdots$$

γ: discount factor

Basic concepts

Example



Percepts: location (A or B), state (clean or dirty)

Actions: *Left*, *Right*, *Suck*, *NoOp*

A vacuum-cleaner world with just two locations. Each location can be clean or dirty, and the agent can move left or right and can clean the square that it occupies.
Different versions of the vacuum world allow for different rules about what the agent can perceive, whether its actions always succeed, and so on.

Source: (Russell, 2016)

Basic concepts

- Example

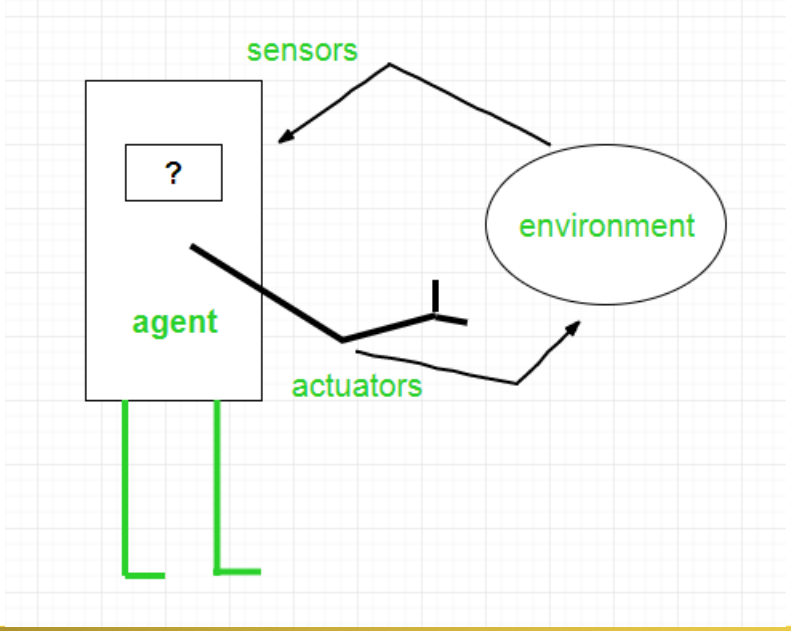
Percept sequence	Action
[A, Clean]	*Right*
[A, Dirty]	*Suck*
[B, Clean]	*Left*
[B, Dirty]	*Suck*
[A, Clean], [A, Clean]	*Right*
[A, Clean], [A, Dirty]	*Suck*
⋮	⋮
[A, Clean], [A, Clean], [A, Clean]	*Right*
[A, Clean], [A, Clean], [A, Dirty]	*Suck*
⋮	⋮

```
function Reflex-Vacuum-Agent( [location,status]) returns an action
    if status = Dirty then return Suck
    else if location = A then return Right
    else if location = B then return Left
```

Basic concepts

Rational agent

- Rational agent is one that does the right thing.
- Obviously, doing right thing is better than doing the wrong thing
- What does it mean to do the right thing?



Basic concepts

Rational agent

Performance measure embodies the criterion for success of an agent's behavior.

E.g., performance measure of a vacuum-cleaner agent:

- amount of dirt cleaned up
- amount of time taken
- amount of electricity consumed
- amount of noise generated

Design reward systems

- Design reward systems

Example: design reward system for chess playing agent.

- System 1: *+1 for a win, 0 for a draw, -1 for a lose*
- System 2: *+1 for capturing an opponent's piece, -1 for being captured a piece*

Exercise: design reward system for maze escape agent.

PEAS description of agents

4 factors should be considered when design an automated agent:

- **P**erformance measure
- **E**nvironment
- **A**ctuators
- **S**ensors

Example: PEAS - automated taxi driver

Performance measure: Safe, fast, legal, comfortable trip, maximize profits, …
Environment: Roads, other traffic, pedestrians, weather, …
Actuators: Steering wheel, accelerator, brake, signal, horn, …
Sensors: Cameras, sonar, speedometer, GPS, odometer, engine sensors, keyboard, …

Example: PEAS - Medical diagnosis system

Performance measure: Healthy patient, minimize costs, lawsuits, …
Environment: Patient, hospital, staff
Actuators: Screen display (questions, tests, diagnoses, treatments, referrals)
Sensors: Keyboard (entry of symptoms, findings, patient's answers)

PEAS description of agents

- Example

Example: PEAS - automated taxi driver

Agent Type	Performance Measure	Environment	Actuators	Sensors
Taxi driver	Safe, fast, legal, comfortable trip, maximize profits, minimize impact on other road users	Roads, other traffic, police, pedestrians, customers, weather	Steering, accelerator, brake, signal, horn, display, speech	Cameras, radar, speedometer, GPS, engine sensors, accelerometer, microphones, touchscreen

PEAS description of the task environment for an automated taxi driver.

Example: PEAS - Medical diagnosis system

Agent Type	Performance Measure	Environment	Actuators	Sensors
Medical diagnosis system	Healthy patient, reduced costs	Patient, hospital, staff	Display of questions, tests, diagnoses, treatments	Touchscreen/voice entry of symptoms and findings
Satellite image analysis system	Correct categorization of objects, terrain	Orbiting satellite, downlink, weather	Display of scene categorization	High-resolution digital camera
Part-picking robot	Percentage of parts in correct bins	Conveyor belt with parts; bins	Jointed arm and hand	Camera, tactile and joint angle sensors
Refinery controller	Purity, yield, safety	Refinery, raw materials, operators	Valves, pumps, heaters, stirrers, displays	Temperature, pressure, flow, chemical sensors
Interactive English tutor	Student's score on test	Set of students, testing agency	Display of exercises, feedback, speech	Keyboard entry, voice

Environment types

- Fully observable (vs. partially observable): agent's sensors give its access to the complete state of the environment at each point in time.
- Deterministic (vs. stochastic): next state of the environment is completely determined by the current state and the action executed by agent.
- Episodic (vs. sequential): agent's experience is divided into atomic "episodes" (each episode consists of agent perceiving and then performing single action)
- Static (vs. dynamic): environment is unchanged while agent is deliberating.
- Discrete (vs. continuous): limited number of distinct, clearly defined percepts and actions.
- Single agent (vs. multiagent): agent operating by itself in environment.

Agent types

Five basic agent types:

- Simple reflex agents
- Model-based reflex agents
- Goal-based agents
- Utility-based agents
- Learning Agent

Thanks for your attention!

Q&A