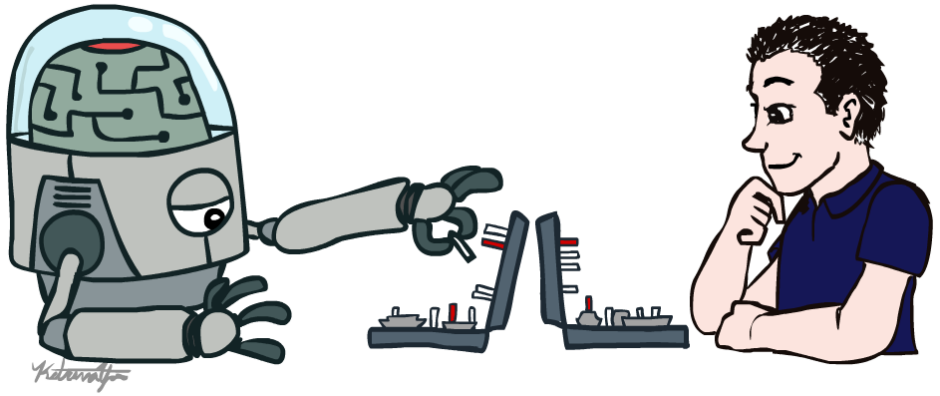
Fall 2024

BECE309L

Artificial Intelligence and Machine Learning

Introduction



Dr. Nitish Katal

Class Management – D1

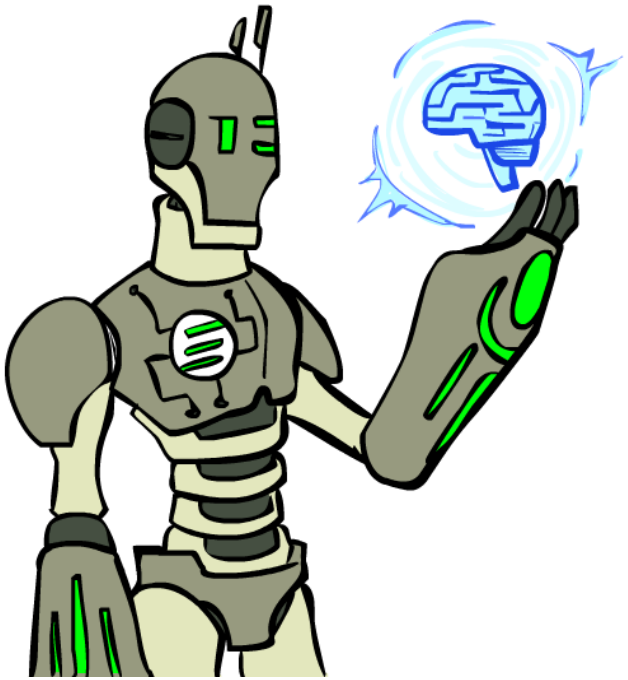
- Google Classroom Code:
  - kexzi2n
- MS Teams Link:
  - 9a32j46

Digital Assignments

- **Project based DAs**
  - Team of **3** students.
  - **3 Reviews (10 marks each)**
    - *Review 1*: Problem definition, literature survey
    - *Review 2*: Methodology & implementation demo
    - *Review 3:* Complete demo. & report submission.
  - Timeline:
    - Team Formation & Tentative Title Submission: **24 July, '24**
    - Submission Deadline: **1 week before/after CATs and before FATs**

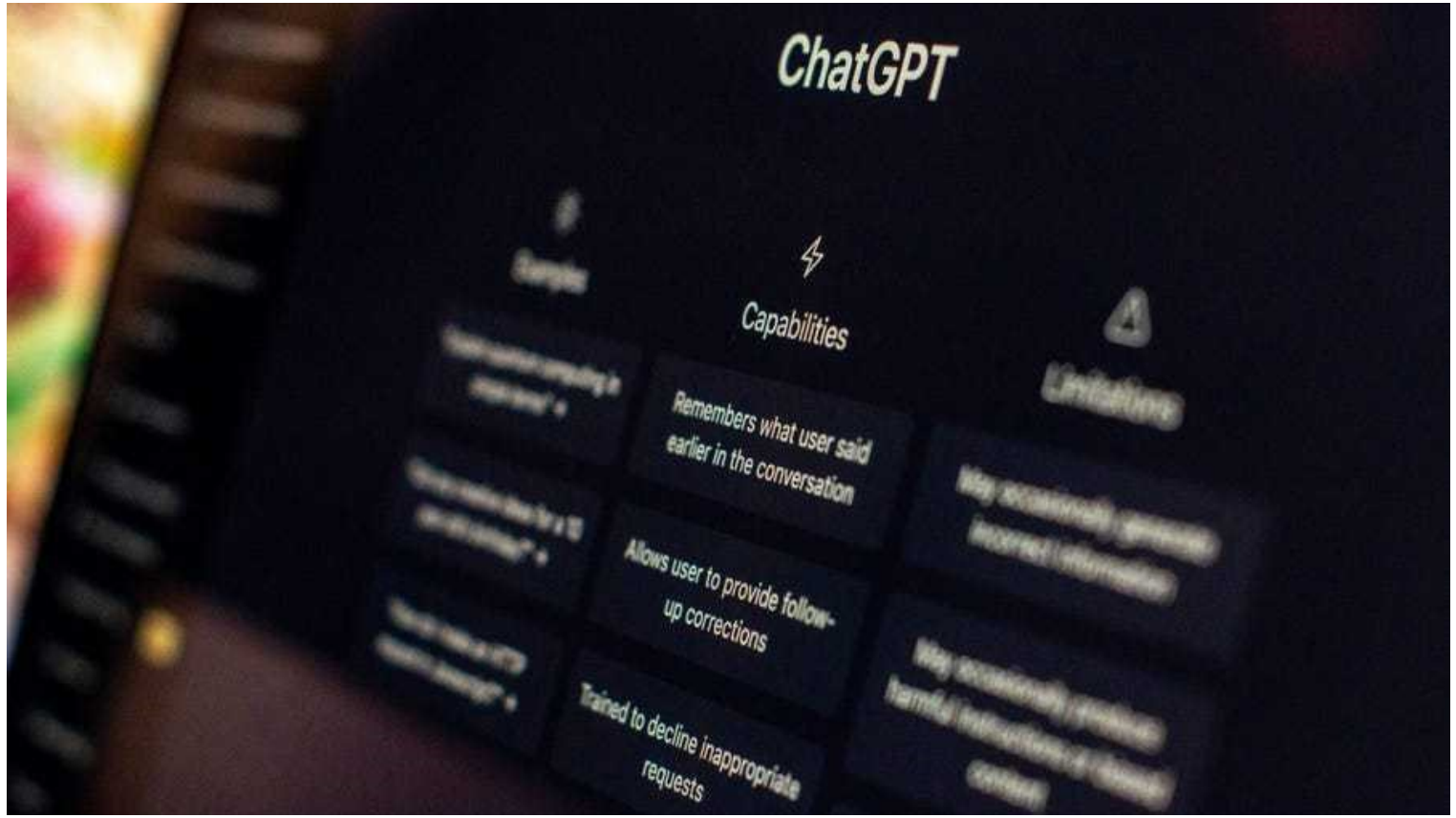
What is AI?

- What is artificial intelligence?
- What can AI do?
- What is this course?



AI is having real-world impact

- Public imagination
  - Text assistants



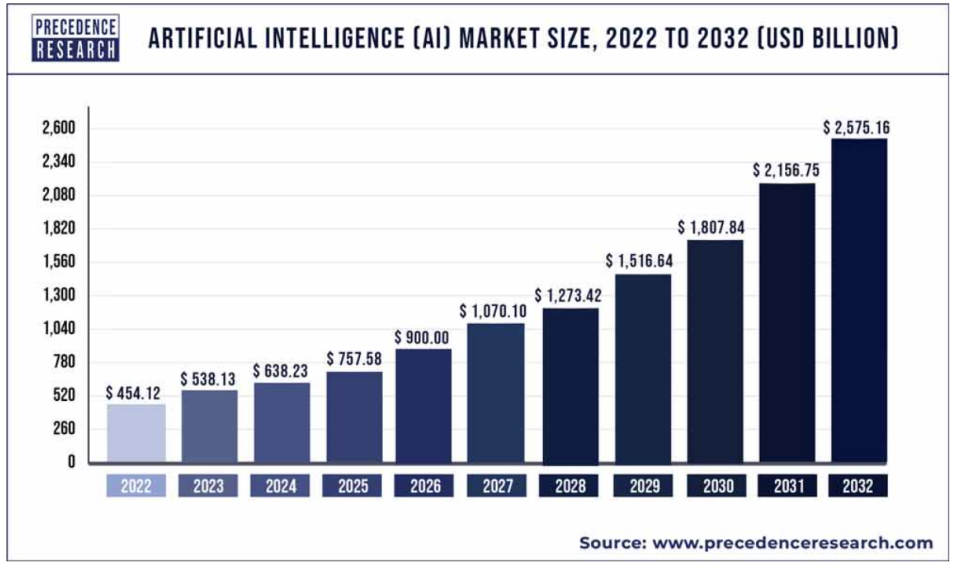
AI is having real-world impact

- Public imagination
  - Text assistants
  - Image generation

AI is having real-world impact

- Public imagination
- Economy
  - 638 billion USD globally
  - By 2032: ~2575 bn $ globally

The global artificial intelligence (AI) market size was valued at USD 454.12 billion in 2022 and is expected to hit around USD 2,575.16 billion by 2032, progressing with a CAGR of 19% from 2023 to 2032. The North America artificial intelligence market was valued at USD 167.30 billion in 2022.



https://www.precedenceresearch.com/artificial-intelligence-market

AI is having real-world impact

- Public imagination
- Economy
- Politics
- Law

Aug. 18, 2023, 12:18 PM; Updated: Aug. 18, 2023, 12:48 PM

AI-Generated Art Lacks Copyright Protection, D.C. Court Says (1)

Riddhi Setty
Reporter

Isaiah Poritz
Legal Reporter

Bloomberg Law, 2023

AI is having real-world impact

- Public imagination
- Economy
- Politics
- Law
- Labor

Finance & economics | Free exchange

New research shows the robots are coming for jobs—but stealthily

Look beneath the aggregate economic numbers, and change is afoot

The Economist, 2021

The Optimist's Guide to Artificial Intelligence and Work

The focus of much discussion is on how it will replace jobs, but nothing is inevitable.

New York Times, 2023

The human labor behind AI chatbots and other smart tools

Data labeling is an important step in developing artificial intelligence but also exposes the people doing the work to harmful content.

MarketWatch, 2023

AI is having real-world impact

- Public imagination
- Economy
- Politics
- Law
- Labor
- Sciences

nature

BIOTECH

AlphaFold Developers Win $3-Million Breakthrough Prize in Life Sciences

DeepMind's system for predicting the 3D structure of proteins is among five recipients of science's most lucrative awards

By Zeeya Merali, Nature magazine on September 22, 2022

Nature, 2022

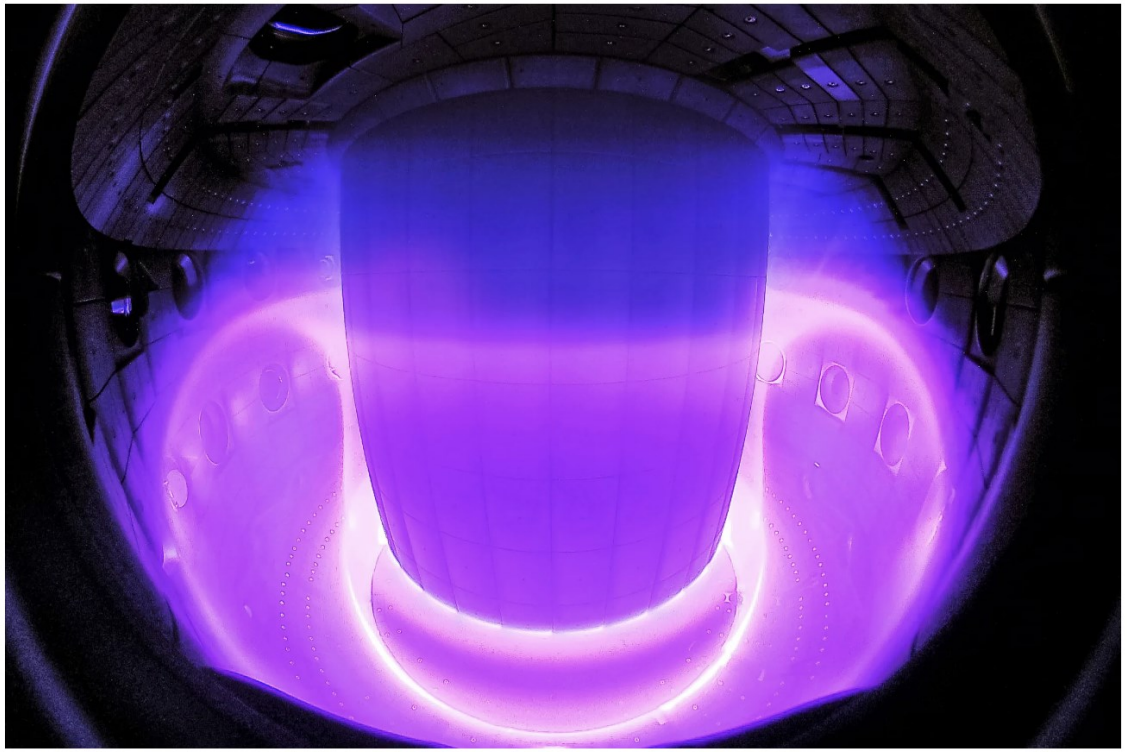
AI is having real-world impact

- Public imagination
- Economy
- Politics
- Law
- Labor
- Sciences

AMIT KATWALA SCIENCE FEB 16, 2022 11:00 AM

DeepMind Has Trained an AI to Control Nuclear Fusion

The Google-backed firm taught a reinforcement learning algorithm to control the fiery plasma inside a tokamak nuclear fusion reactor.



PHOTOGRAPH:CURDIN WÜTHRICH, SPC/EPFL

Wired, 2022

AI is having real-world impact

- Public imagination
- Economy
- Politics
- Law
- Labor
- Sciences
- Education

BREAKING

ChatGPT In Schools: Here's Where It's Banned—And How It Could Potentially Help Students

Arianna Johnson Forbes Staff

I cover the latest trends in science, tech and healthcare.

Follow

2

Jan 18, 2023, 02:31pm EST

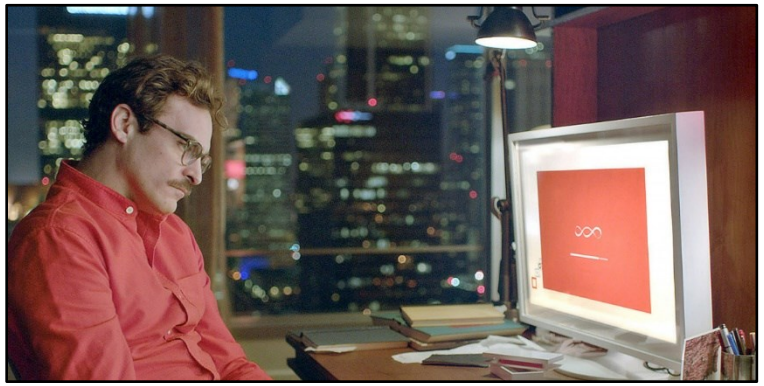
Forbes, 2023

AI is having real-world impact

- Public imagination
- Economy
- Politics
- Law
- Labor
- Sciences
- Education

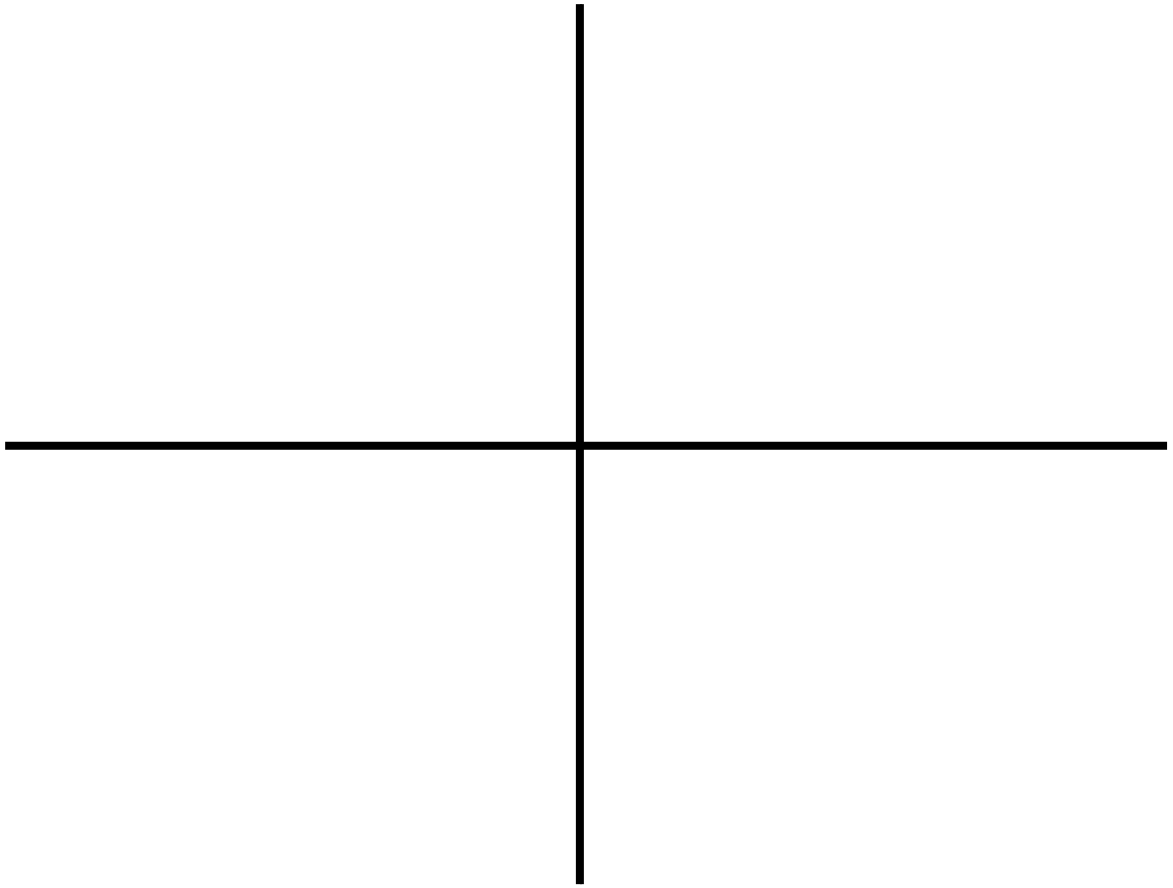
Ok, but what actually is AI???

Science fiction AI?



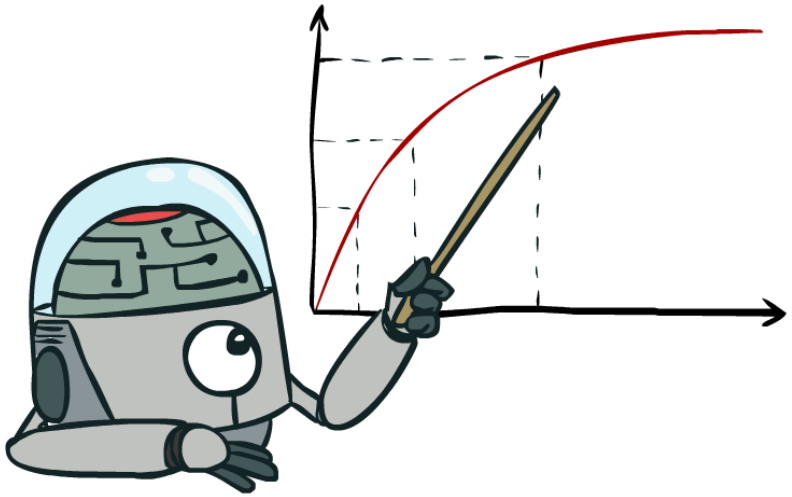
What should we build?

Should we make machines that…



Rational Decisions

- We'll use the term **rational** in a very specific, technical way:
  - Rational: *maximally achieving pre-defined goals*
  - Goals are expressed in terms of the **utility** of outcomes
  - World is uncertain, so we'll use **expected** utility
  - Being rational means acting to **maximize your expected utility**



Rational Decisions

- We'll use the term **rational** in a very specific, technical way:
  - Rational: *maximally achieving pre-defined goals*
  - Goals are expressed in terms of the **utility** of outcomes
  - World is uncertain, so we'll use **expected** utility
  - Being rational means acting to **maximize your expected utility**

A better title for this course might be:

Computational Rationality

Perspectives on Intelligence

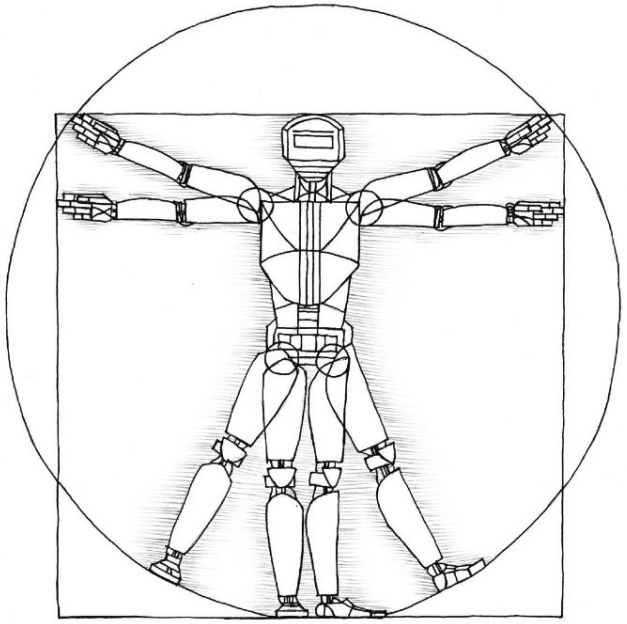
- Skills-based perspective
- "A system is only intelligent if it can do [X]."
  - Play chess?
  - Learn from experience?
  - Use words properly?
  - Make mistakes?
  - Not make mistakes?

Perspectives on Intelligence

- **Embodiment perspective** (Rodney Brooks)

A concept rooted in cognitive science, philosophy, and robotics that emphasizes the ***role of the body in shaping the mind***

  - Bodily Integrations:
  - Sensorimotor Feedback:
  - Role in Robotics & AI

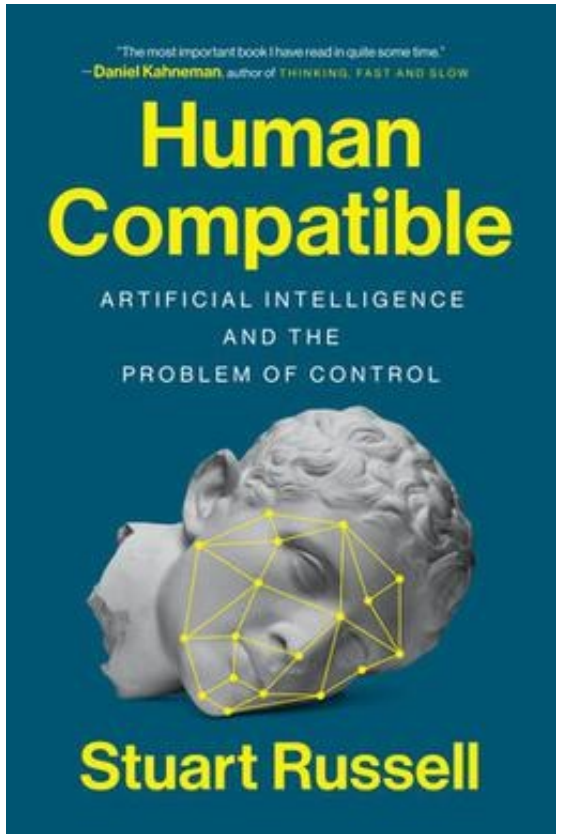


Perspectives on Intelligence

- **Psychometrics perspective** (François Chollet)
- "Measuring abilities, not skills [...] across a broad range of tasks, including tasks that were previously unknown to the ability-enabled system and its developers."

Perspectives on Intelligence

- **Human-compatible perspective** (Stuart Russell)



1. Machine's objective is to maximize <u>human utility</u>.
2. Initially <u>uncertain</u> about human preferences.
3. Must learn about preferences from human <u>behavior</u>.

Perspectives on Intelligence

A human being should be able to change a cloth, plan an invasion, steer a ship, design a building, write a poem, balance accounts, build a wall, set a bone, comfort the dying, take orders, give orders, cooperate, act alone, solve equations, analyze a new problem, pitch manure, program a computer, cook a tasty meal, fight efficiently, die gallantly. Specialization is for insects.

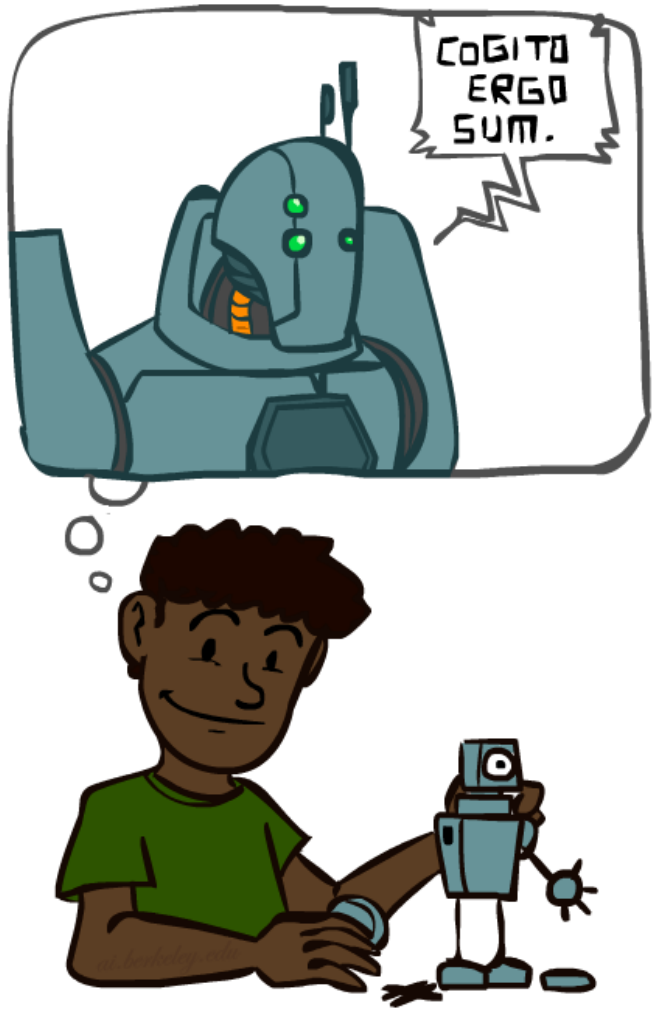
—Robert A. Heinlein

What About the Brain?

- Brains (human minds) are very good at making rational decisions, but not perfect
- Brains aren't as modular as software, so hard to reverse engineer!
- AI may be better than brains at some tasks
- *"Brains are to intelligence as wings are to flight"*
- We can't yet build AI on the scale of the brain
  - ~100T synapses in the human brain vs ~1.8T weights in GPT4
- Still, the brain can be a great inspiration for AI!

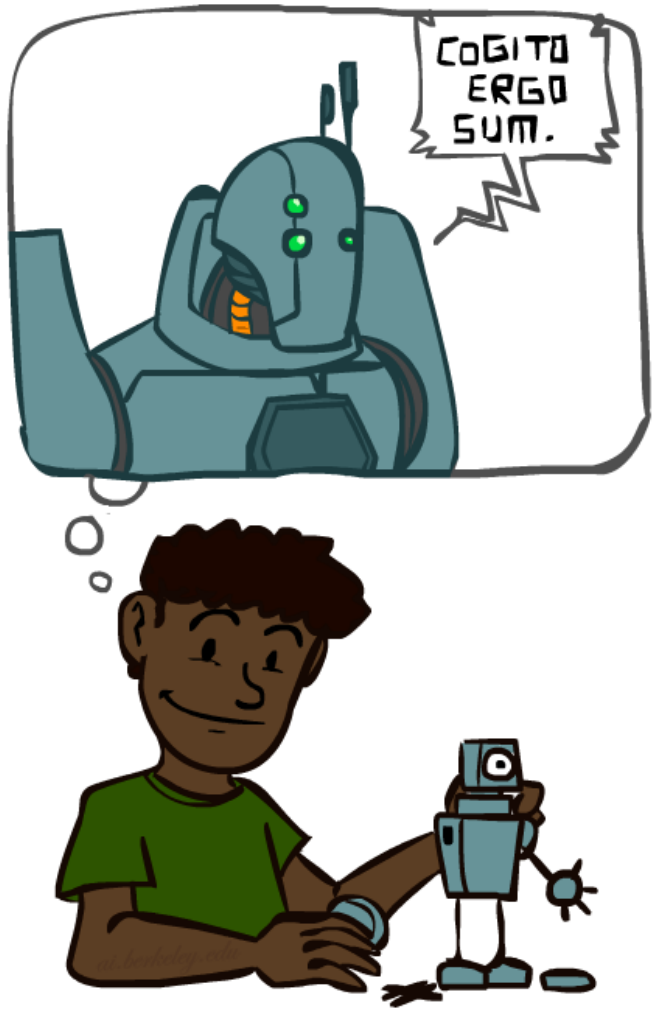
A (Short) History of AI

- 1940-1950: Early days: neural and computer science meet
  - **1943: McCulloch & Pitts: Perceptron–boolean circuit model of brain**
  - **1950: Turing's "Computing Machinery and Intelligence"**
- 1950—70: Excitement! Logic-driven
  - 1950s: Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
  - 1956: Dartmouth meeting: "Artificial Intelligence" adopted
  - **1969: Minsky & Papert: Perceptrons can't learn XOR/parity!**
- 1970—90: Knowledge-based approaches
  - 1969—79: Early development of knowledge-based systems
  - 1980—88: Expert systems industry booms; backpropagation makes it feasible to train multi-layer neural networks
  - 1988—93: Expert systems industry busts: "**AI Winter**"
- 1990—2010: Statistical approaches, agents
  - Resurgence of probability, focus on uncertainty
  - **Agents and learning systems… "AI Spring"?**
  - 1992: TD-Gammon achieves human-level play at backgammon
  - **1997: Deep Blue defeats Gary Kasparov at chess**
  - 2002: Embodied AI; Roomba vacuum invented



A (Short) History of AI

- 2010—2017: Big Data, GPUs, Deep Learning
  - 2011: Apple releases Siri
  - 2012: AlexNet wins ImageNet competition
  - **2015: DeepMind achieves "human-level" control in Atari games**
  - **2016: DeepMind's AlphaGo defeats Lee Sedol at Go**
  - 2016: Google Translate migrates to neural networks
- 2017—: Scaling Up, Large Language Models
  - 2017: Google invents Transformer architecture
  - 2017: DeepStack/Libratus defeat humans at poker
  - **2018-2020: AlphaFold predicts protein structure from amino acids**
  - 2021-2022: Modern text-to-image generation
  - **2022: OpenAI releases ChatGPT**
  - 2023: Every other company also releases a chatbot



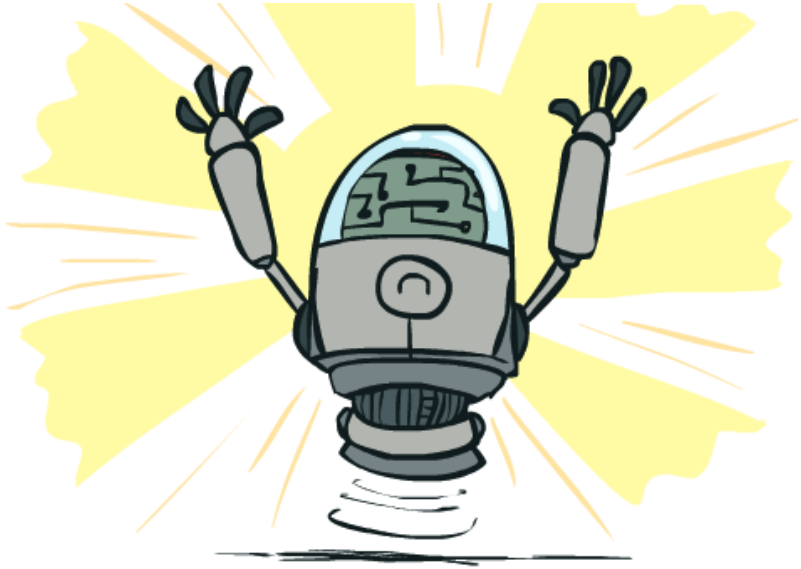
What Can AI Do?

Quiz: Which of the following can be done at present?

- ✓ Win against any human at chess?
- ✓ Win against the best humans at Go?
- Play a decent game of table tennis?
- Unload any dishwasher in any home?
- Drive safely along the highway?

- Buy a week's worth of groceries on the web?

- Discover and prove a new mathematical theorem?
- Perform a surgical operation?
- Translate spoken Chinese into spoken English in real time?
- Win an art competition?
- Write an intentionally funny story?
- Construct a building?



What Can AI Do?

Quiz: Which of the following can be done at present?

- ✓ Win against any human at chess?
- ✓ Win against the best humans at Go?
- Play a decent game of table tennis?
- Unload any dishwasher in any home?
- Drive safely along the highway?

- Buy a week's worth of groceries on the web?

- Discover and prove a new mathematical theorem?
- Perform a surgical operation?
- Translate spoken Chinese into spoken English in real time
- Win an art competition?
- Write an intentionally funny story?
- Construct a building?

Google's AlphaGo beats Lee Sedol at Go in 2016

But … a plot twist in 2023!

DEEP BLUE WAS JUST THE START —

Man beats machine at *Go* in human victory over AI

Amateur exploited weakness in systems that have otherwise dominated grandmasters.

RICHARD WATERS, FINANCIAL TIMES - 2/19/2023, 4:51 AM

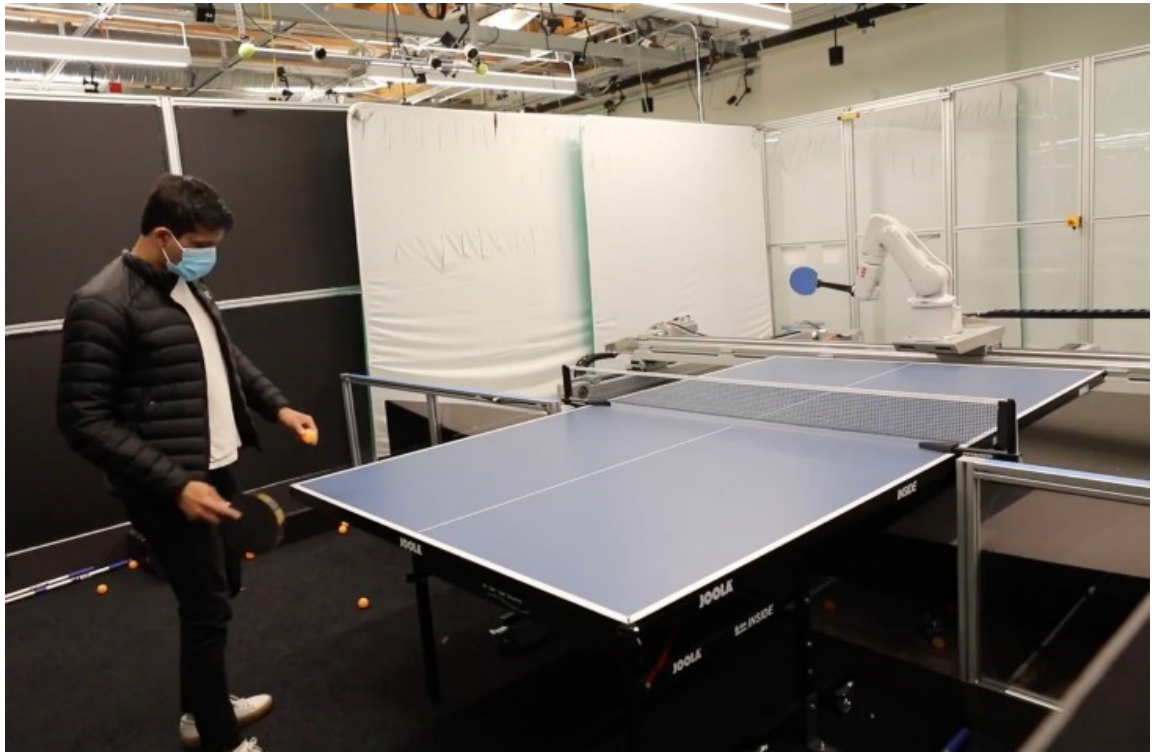
What Can AI Do?

Quiz: Which of the following can be done at present?

- ✓ Win against any human at chess?
- ✓ Win against the best humans at Go?
- ✓ Play a decent game of table tennis?
- Unload any dishwasher in any home?
- Drive safely along the highway?

- Buy a week's worth of groceries on the web?

- Discover and prove a new mathematical theorem?
- Perform a surgical operation?
- Translate spoken Chinese into spoken English in real time?
- Win an art competition?
- Write an intentionally funny story?
- Construct a building?



What Can AI Do?

Quiz: Which of the following can be done at present?

- ✓ Win against any human at chess?
- ✓ Win against the best humans at Go?
- ✓ Play a decent game of table tennis?
- ✗ Unload any dishwasher in any home?
- ✓ Drive safely along the highway?

- ✓ Buy a week's worth of groceries on the web?

- ? Discover and prove a new mathematical theorem?
- ✗ Perform a surgical operation?
- ✓ Translate spoken Chinese into spoken English in rea
- ✓ Win an art competition?
- Write an intentionally funny story?
- Construct a building?

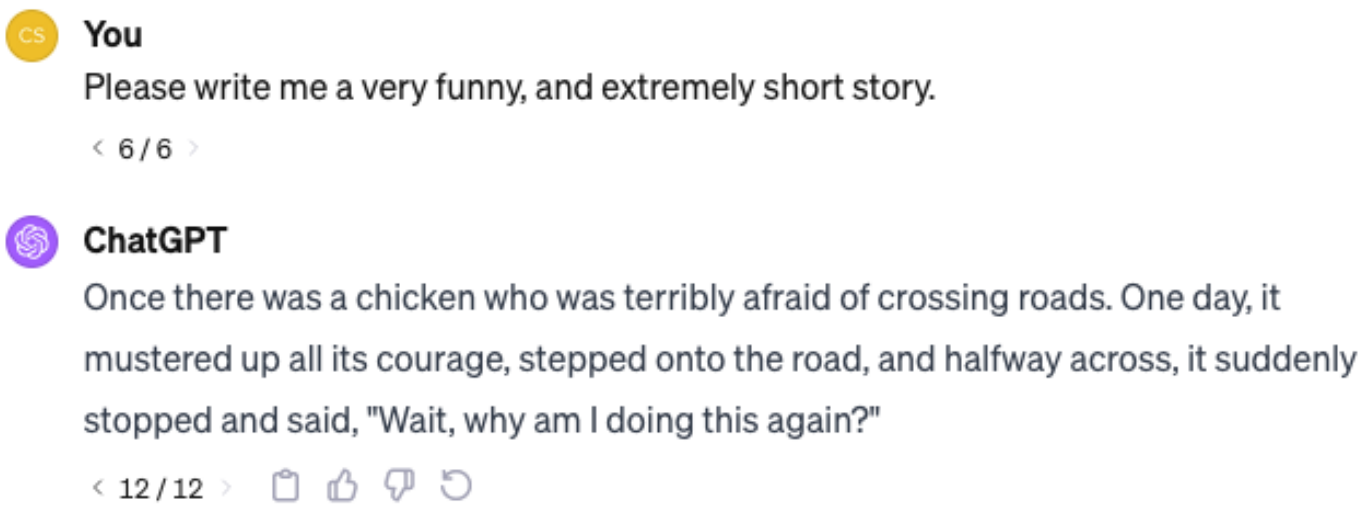
What Can AI Do?

Quiz: Which of the following can be done at present?

- ✓ Win against any human at chess?
- ✓ Win against the best humans at Go?
- ✓ Play a decent game of table tennis?
- ✗ Unload any dishwasher in any home?
- ✓ Drive safely along the highway?

- ✓ Buy a week's worth of groceries on the web?

- ? Discover and prove a new mathematical theorem?
- ✗ Perform a surgical operation?
- ✓ Translate spoken Chinese into spoken English in real time?
- ✓ Win an art competition?
- ? Write an intentionally funny story?
- Construct a building?



What Can AI Do?

Quiz: Which of the following can be done at present?

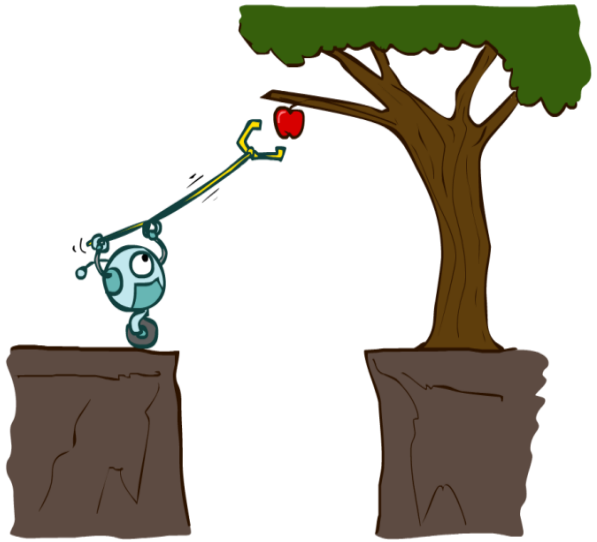
- ✓ Win against any human at chess?
- ✓ Win against the best humans at Go?
- ✓ Play a decent game of table tennis?
- ✗ Unload any dishwasher in any home?
- ✓ Drive safely along the highway?

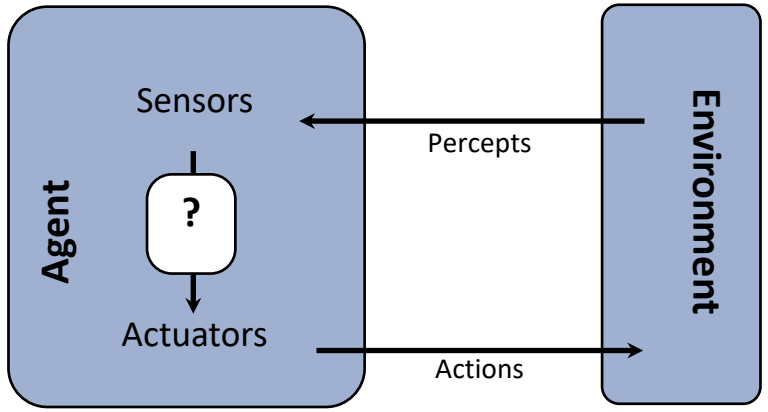
- ✓ Buy a week's worth of groceries on the web?

- ? Discover and prove a new mathematical theorem?
- ✗ Perform a surgical operation?
- ✓ Translate spoken Chinese into spoken English in real time?
- ✓ Win an art competition?
- ? Write an intentionally funny story?
- ✗ Construct a building?

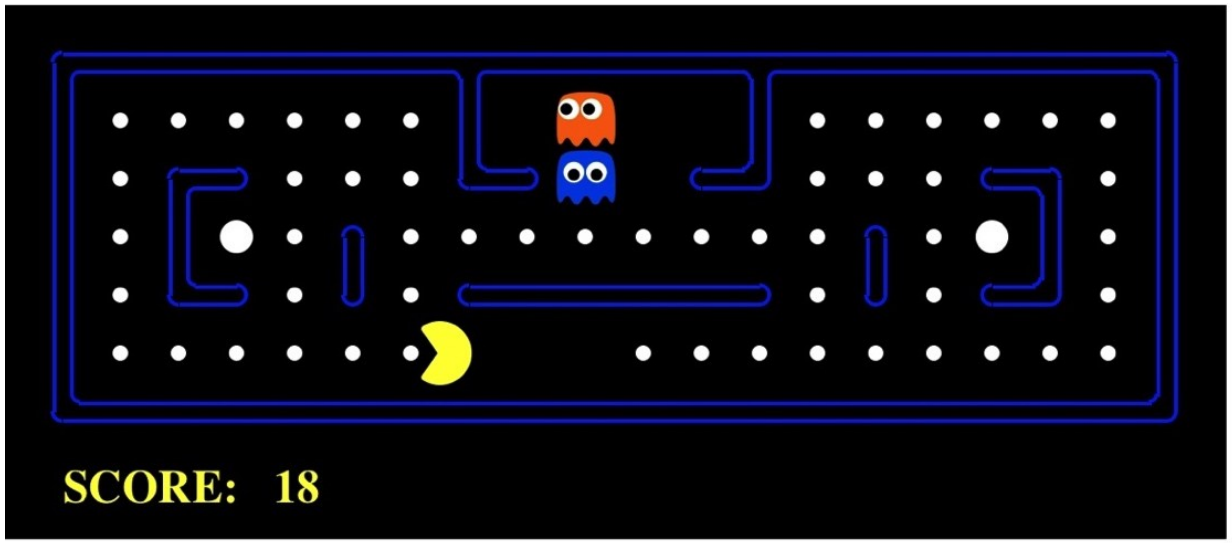
This Course: Designing Rational Agents

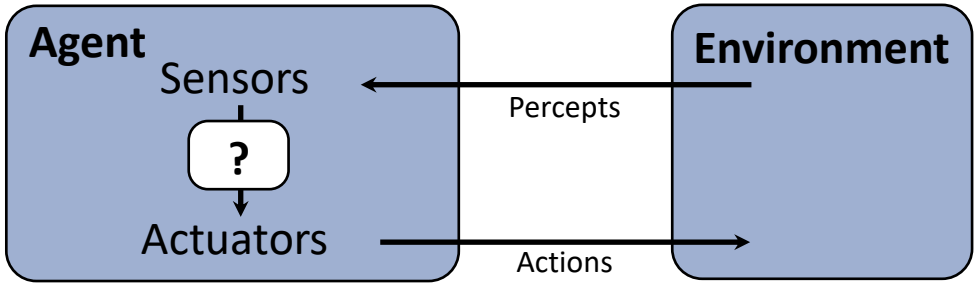
- An **agent** is an entity that perceives and acts.
- A **rational agent** selects actions that maximize its (expected) **utility**.
- Characteristics of the **percepts**, **environment**, and **action space** dictate techniques for selecting rational actions
- This course is about:
  - General AI techniques for a variety of problem types
  - Learning to recognize when and how a new problem can be solved with an existing technique





Pac-Man as an Agent





Pac-Man is a registered trademark of Namco-Bandai Games, used here for educational purposes

Course Topics

Core Components of Rational Agents:

- Search & Planning
- Probability & Inference
- Supervised Learning
- Reinforcement Learning

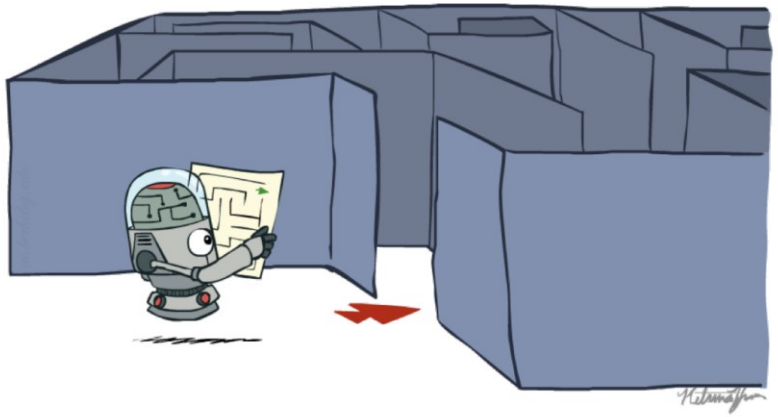
Course Topics

Search & Planning

Probability & Inference

Supervised Learning

Reinforcement Learning



How can I use my ***model*** of the world to find a ***sequence of actions*** to achieve my ***goal***?

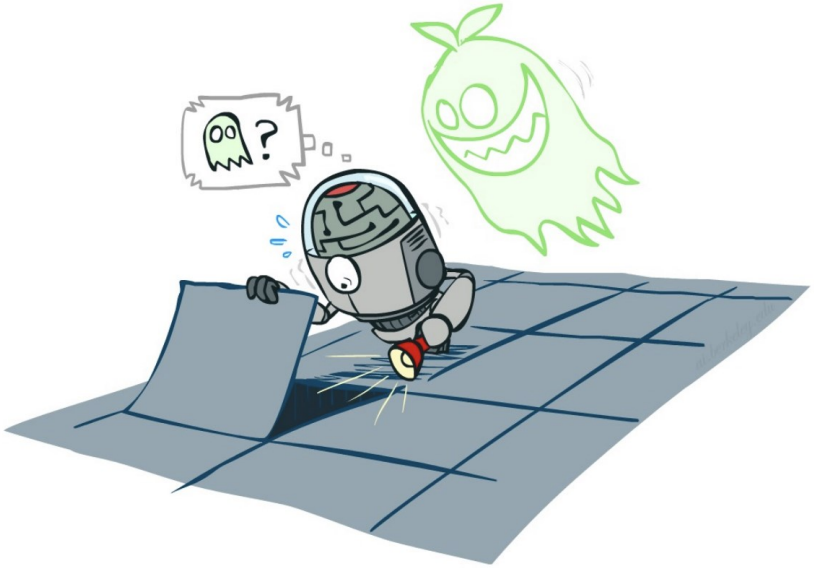
Course Topics

Search & Planning

Probability & Inference

Supervised Learning

Reinforcement Learning



How can I make sense of ***uncertainty***?

Course Topics

Search & Planning

Probability & Inference

Supervised Learning

Reinforcement Learning

How can I learn a ***model*** of the world from ***data***?

Course Topics

Search & Planning

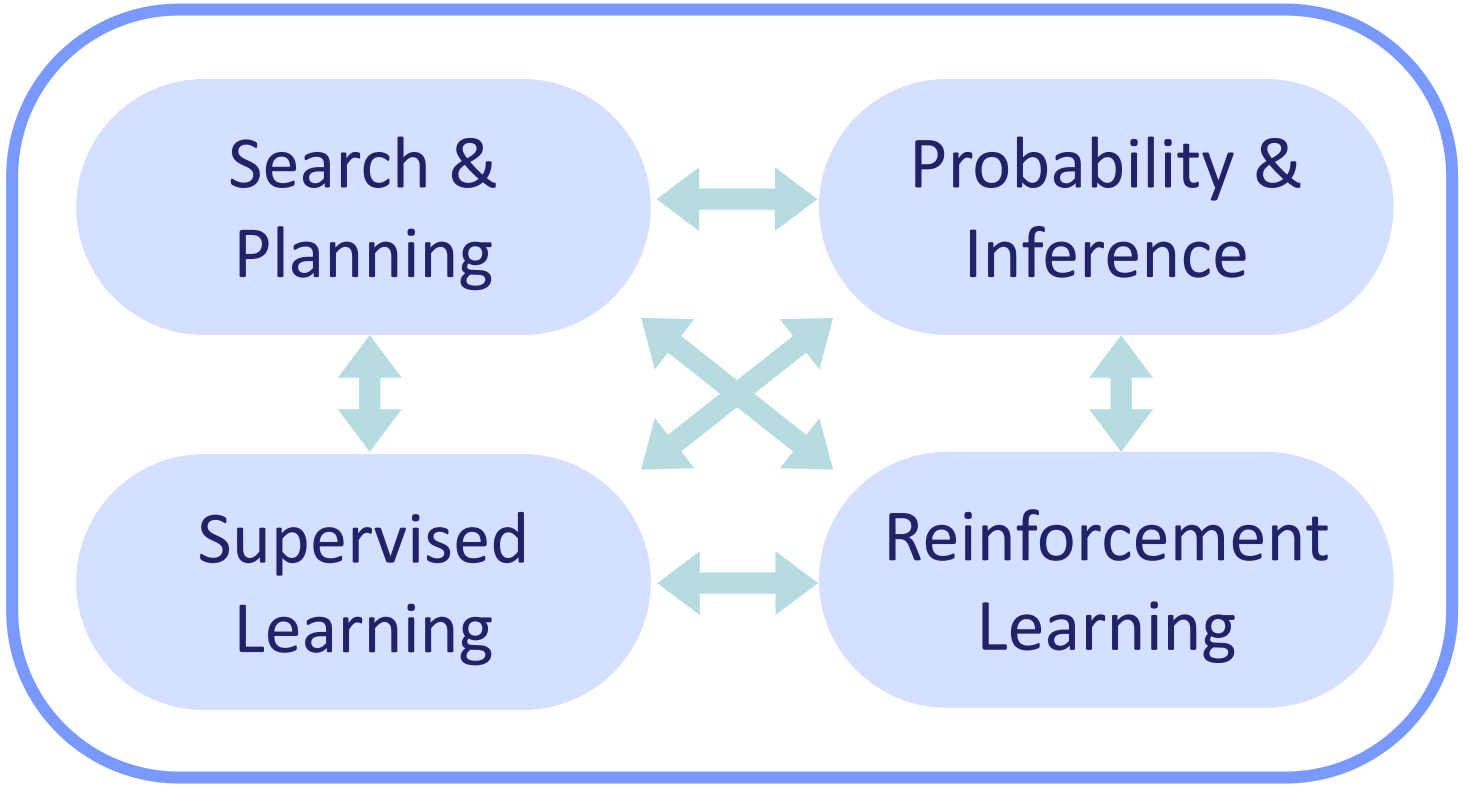
Probability & Inference

Supervised Learning

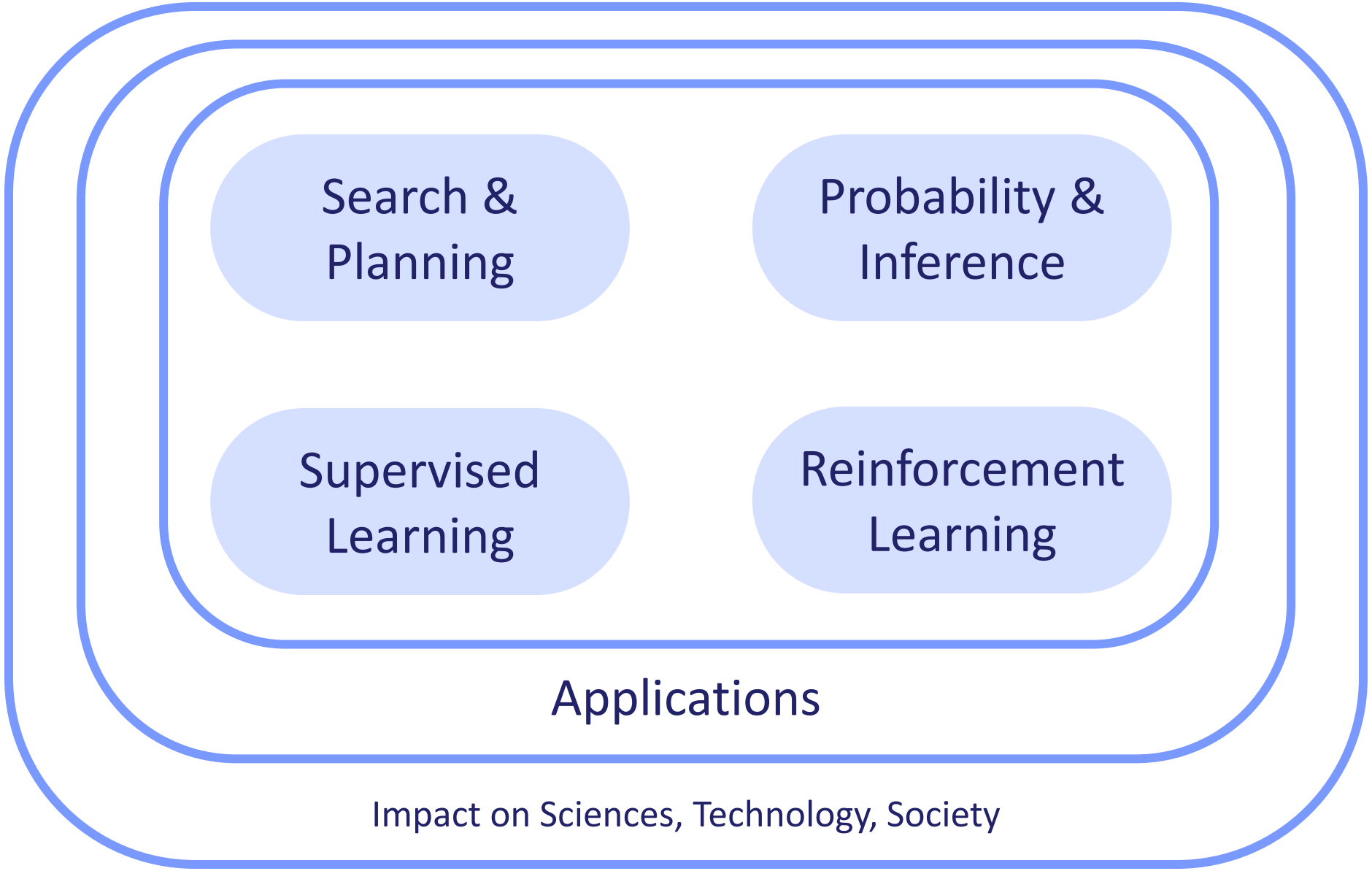
Reinforcement Learning

How can I learn a ***policy*** for any situation so that I can ***maximize utility***?

Course Topics

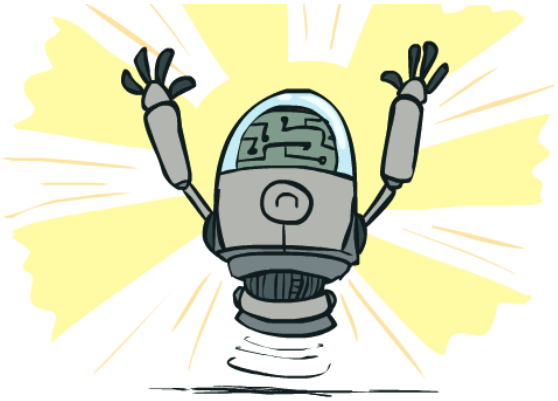


Course Topics



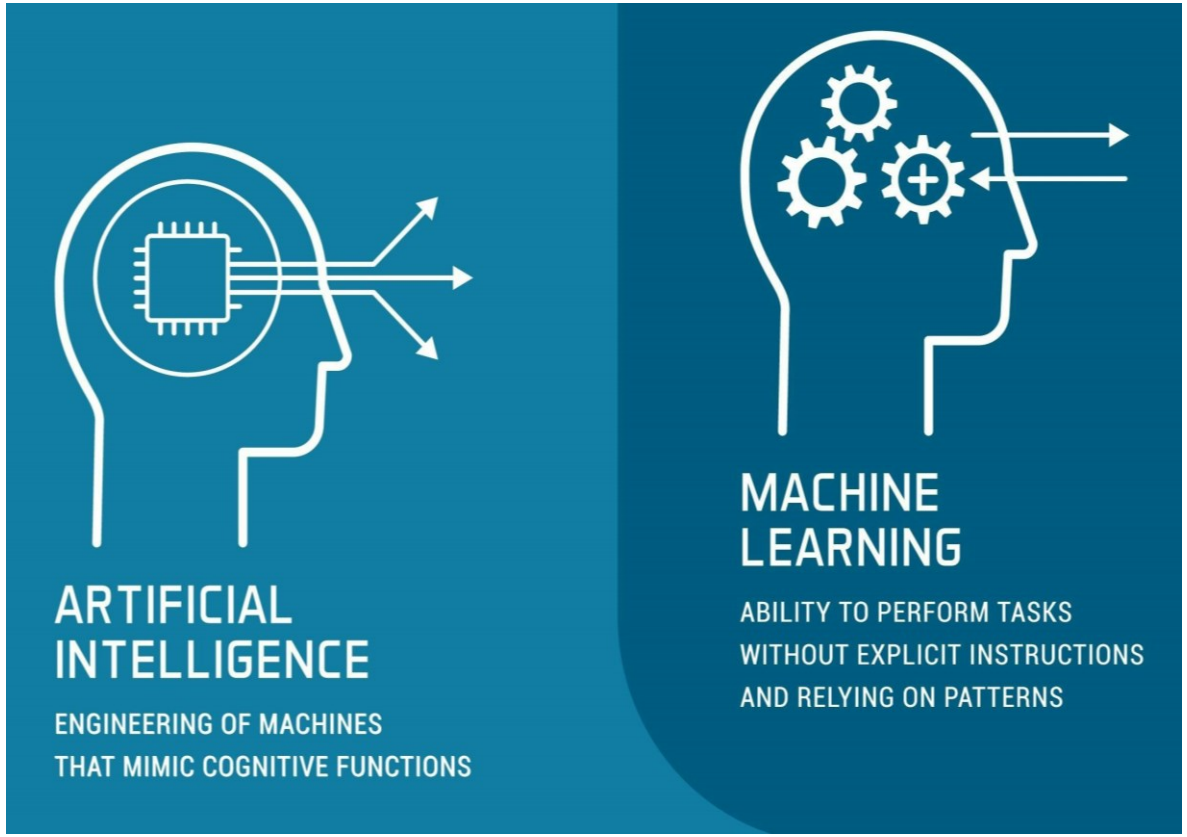
By the end of this course you'll:

- Build and understand math of rational, learning agents
- Select and apply the right AI methods for wide range of problems
- Recognize how these methods are used in modern AI systems
- Be prepared to make decisions on how AI is used in society



Course Contents

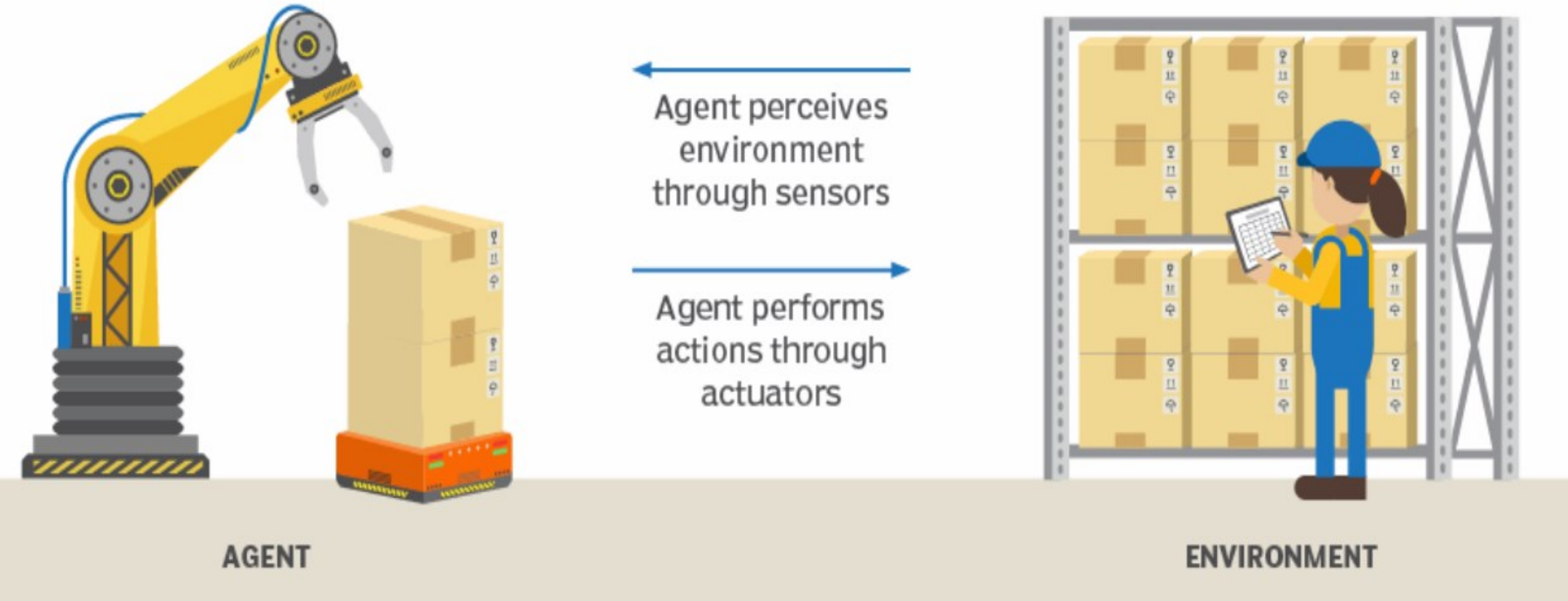
- Module 1: Foundations of AI
- Module 2: Problem-solving by Searching
- Module 3: Knowledge Representation
- Module 4: Probability reasoning and uncertainty.
- Module 5: Data Preparation for Machine Learning
- Module 6: Learning from Examples
- Module 7: Deep Learning



Module 1
Foundations of AI

Module 1: Foundations of AI

- Introduction
- Agents and rationality
- Task environment
- Agent Architecture Types



Basic Terminologies

Intelligence

- Ability to **apply knowledge** in order to **perform better in an environment**.

Artificial Intelligence

- Study and construction of **agent programs** that perform well in a given environment, for a given agent architecture. **(or)**
- **Engineering of Machines that Mimic Cognitive Functions**

Machine Learning

- Ability to perform tasks without explicit instructions and relying on pattern

Deep Learning

- **Engineering of Machines that Mimic Cognitive Functions**

Agent

- An entity that takes action in response to precepts from an environment.

Rationality

- Property of a system which does the "right thing" given what it knows.

Human Intelligence Vs. AI

"Artificial Intelligence is the ability of a computer to act like a human being". i.e.
Engineering of Machines that Mimic Cognitive Functions

	Artificial Intelligence	Human Intelligence
Processing	Based on **algorithms** and **mathematical models**	Based on **cognitive processes** and **biological structures**
Learning	Based on **data** and **feedback loops**	Based on **experience**, **intuition**, and **creativity**
Speed	Can **process data** and **perform tasks** much **faster** than **humans**	**Slower than AI** in processing **large amounts of data**, but can make **complex decisions quickly**
Adaptability	Can **quickly adapt** to **new data** and **situations**	Can **adapt to new situations**, **learn from experience**, and **make decisions based** on **context**
Emotions	**Lacks emotions** and **empathy**	Capable of **feeling emotions** and **empathy**
Creativity	**Limited** ability to be **creative** or **think outside** of the **box**	Capable of **creativity**, **imagination**, and **innovation**
Ethics	Does **not** have a **moral code** or **conscience**	Has a **moral code** and **conscience** that **guides decision-making**
Physical Limitations	Does **not have physical limitations**, can **operate 24/7**	**Limited** by **physical capabilities** and **requires rest** and **maintenance**

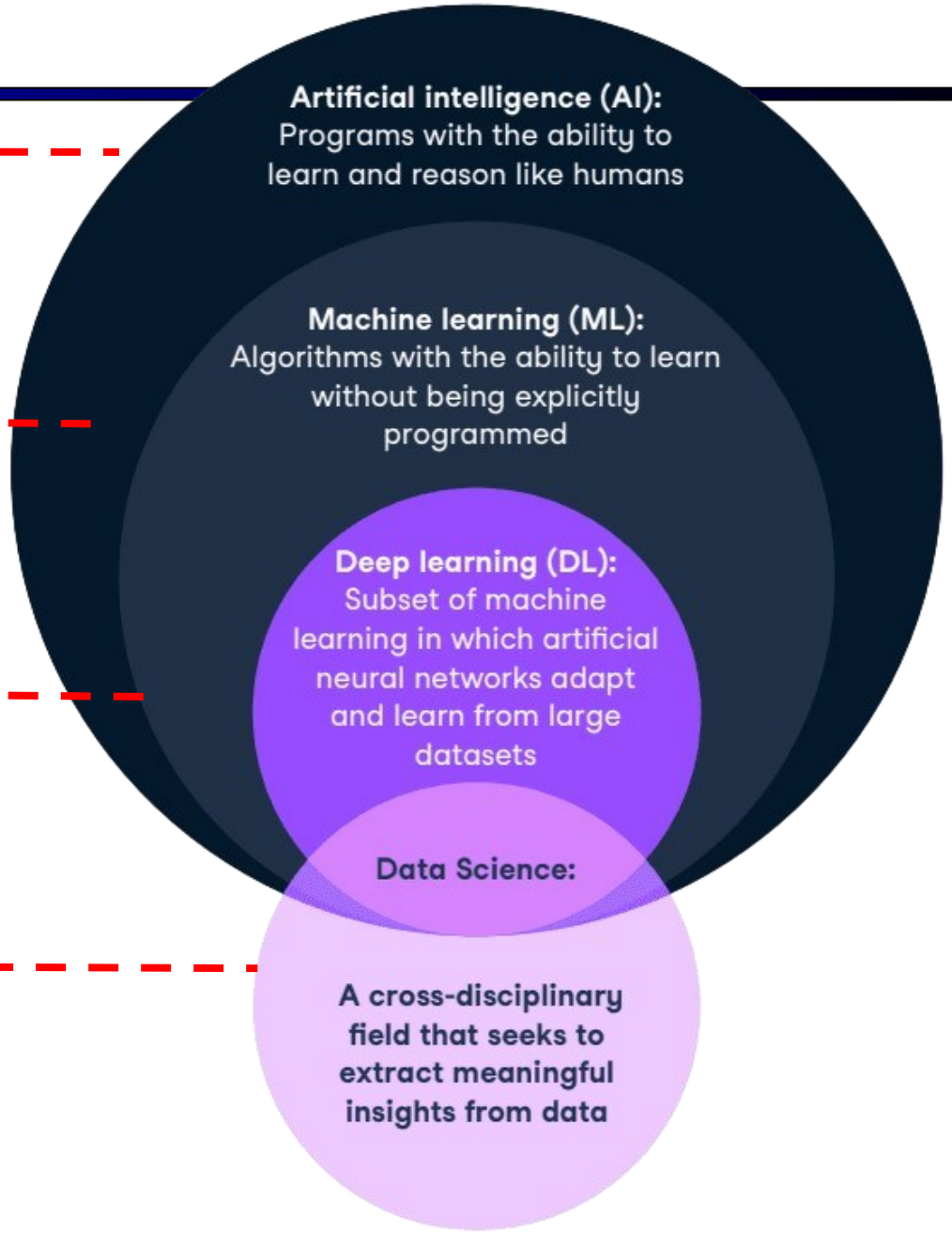
AI Subsections

- Technology for ***machines to understand/interpret, learn and make intelligent decisions***.

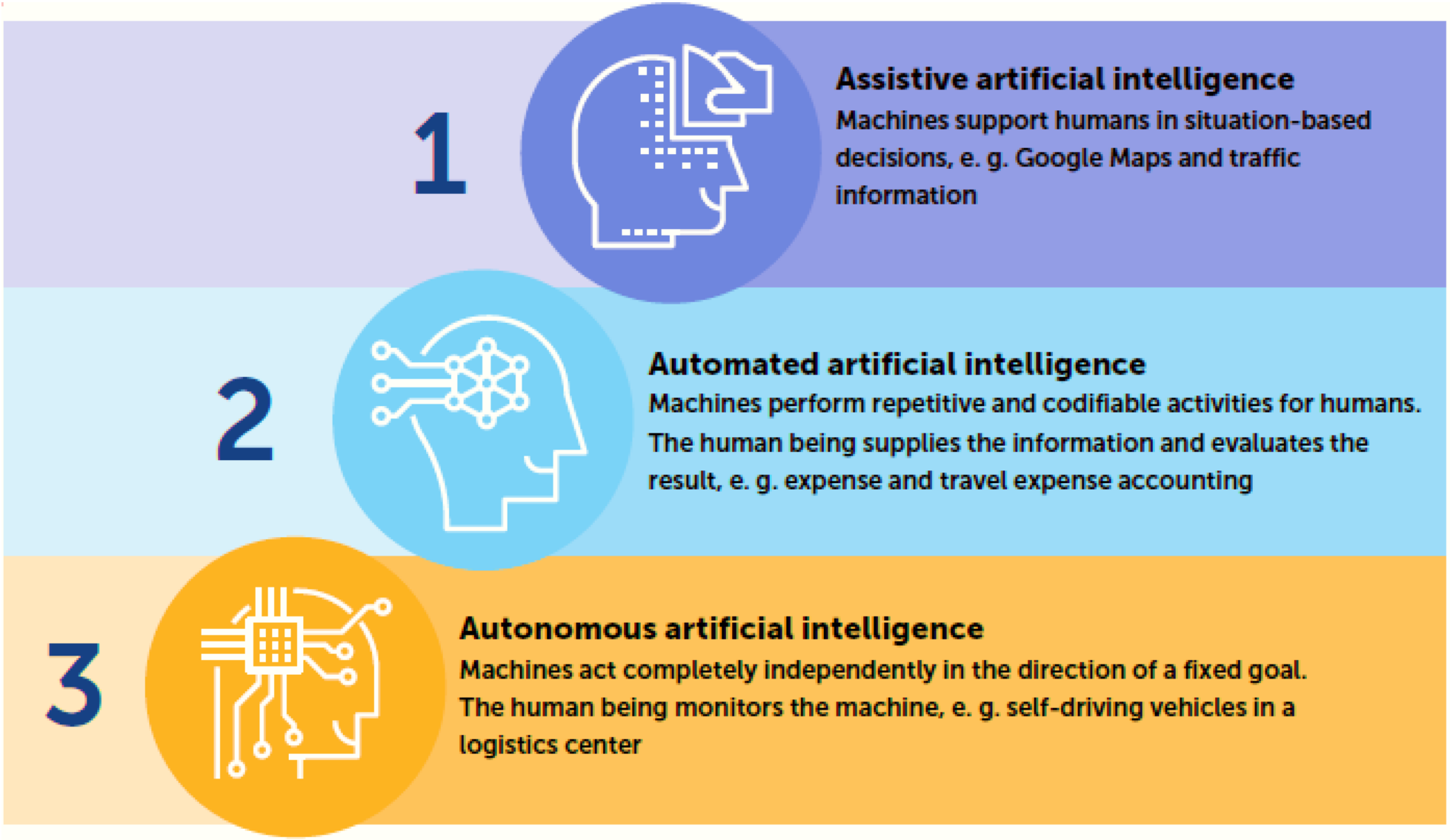
- **Algorithms** that **helps machines improve** through supervised, unsupervised and reinforcement learning.

- Subsection of ML which make ***multi-neural network feasible***

- ***Collection***, ***Preparation*** and ***analysis*** of ***Data***
- ***Leverages AI/ML***, ***research***, ***Industrial expertise*** and ***statistics*** to make ***business decisions***



3 A's of AI


1
Assistive artificial intelligence
Machines support humans in situation-based decisions, e. g. Google Maps and traffic information
2
Automated artificial intelligence
Machines perform repetitive and codifiable activities for humans. The human being supplies the information and evaluates the result, e. g. expense and travel expense accounting
3
Autonomous artificial intelligence
Machines act completely independently in the direction of a fixed goal. The human being monitors the machine, e. g. self-driving vehicles in a logistics center

Need for AI

Expert Systems:

- ES that exhibit **intelligent behavior** with the **capability** to **learn**, **demonstrate**, **explain**, and **advise** its **users**.
- Helping **machines find solutions to complex problems like humans** do and applying them as algorithms in a computer-friendly manner.

Improved efficiency:

- Can **streamline time-consuming tasks**,
- Enhancing efficiency and productivity,
- Enabling humans to concentrate on more creative, high-level tasks.

Better decision-making:

- Artificial intelligence aids in decision-making **by analyzing vast amounts of data**, particularly in ***finance***, ***healthcare***, and ***logistics***, where decisions significantly influence outcomes.

Enhanced accuracy:

- Artificial intelligence algorithms can process **data quickly and accurately**, reducing **the risk of errors** that can occur in manual processes.
- This can improve the reliability and quality of results.

Personalization:

- Artificial intelligence can enhance customer satisfaction and loyalty by personalizing user experiences, providing **personalized recommendations**, and enhancing interactions based on individual preferences and behaviors.

Exploration of new frontiers:

- Artificial intelligence can unlock previously unexplored knowledge, leading to significant advancements in fields like **astronomy**, **genetics**, and **drug discovery**.

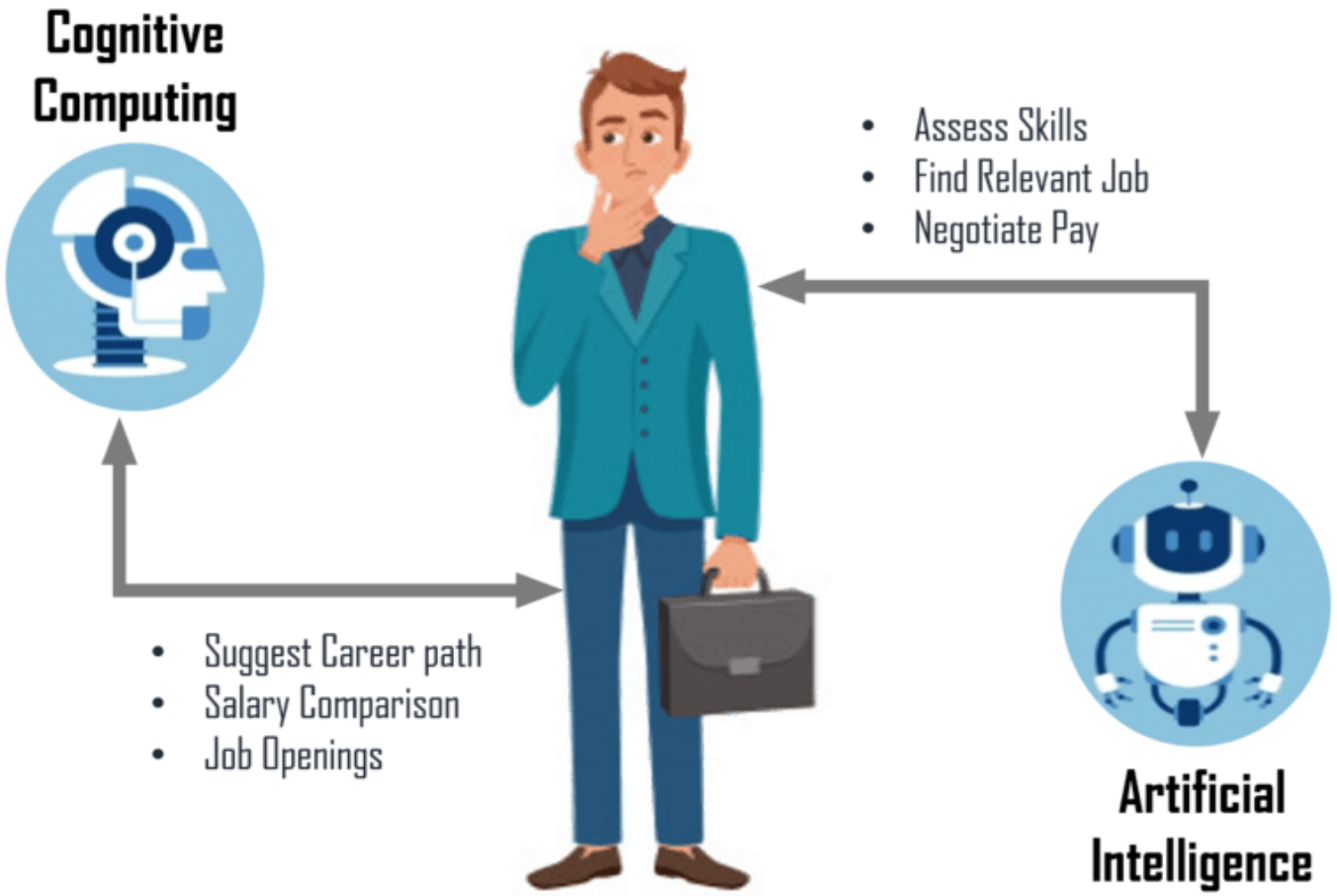
Approaches in AI

Approaches to AI

1. ***Think like humans***

The cognitive modelling approach

- ***Formulate*** a ***theory*** of ***mind***/brain
- *Express the theory in a computer program*
- Two Approaches
  - **Cognitive Science and Psychology**
    - Testing/ predicting responses of human subjects
  - **Cognitive Neuroscience**
    - Observing neurological data

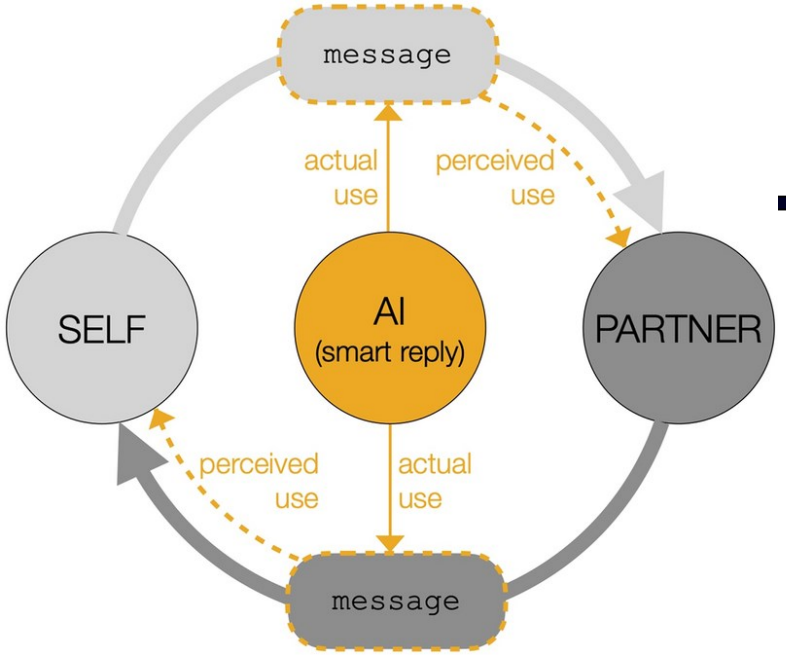


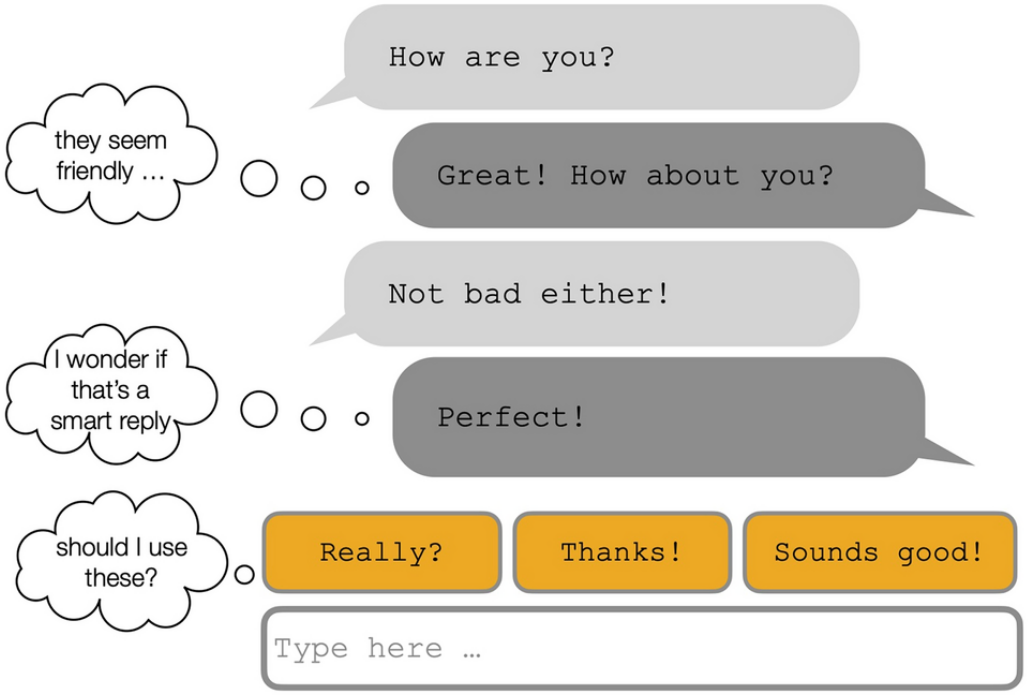
Approaches to AI

2. *Think Rationally*

The laws of thought process/approach

- "**Rational**" -> **ideal intelligence**
  - Contrast with **human** intelligence
- Rational thinking **governed** by **precise** "***laws of thought***"
  - syllogisms
  - notation and
  - logic
- Systems (in theory) can solve problems using such laws





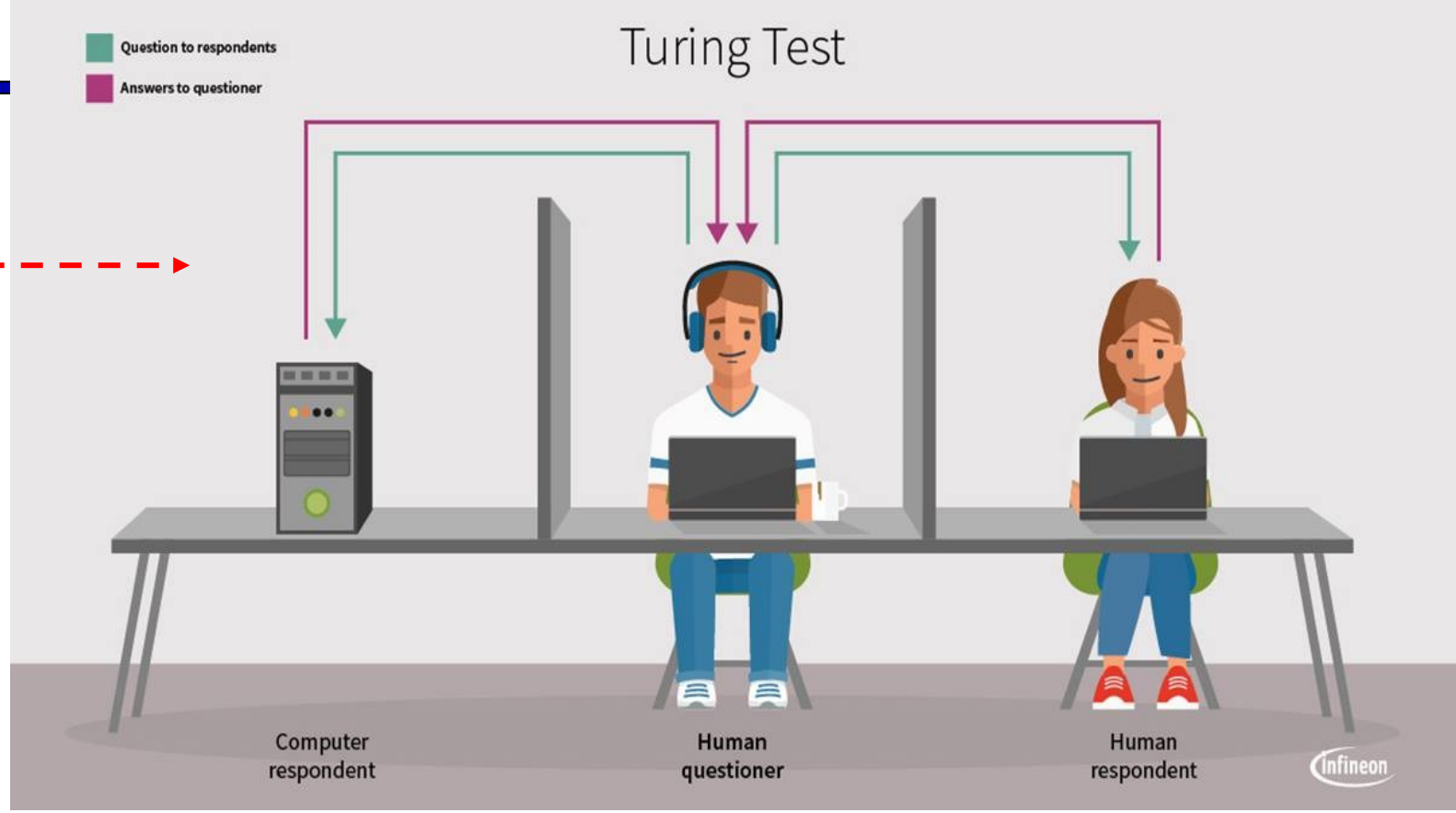
Approaches in AI

3. *Act Like Humans*

Turing Test Approach

Test for intelligent behaviour

- Interrogator writes questions and receives answers
- System providing the answers passes the test if interrogator cannot tell whether the answers come from a person or not.



Necessary components of such a system form major AI sub-disciplines:

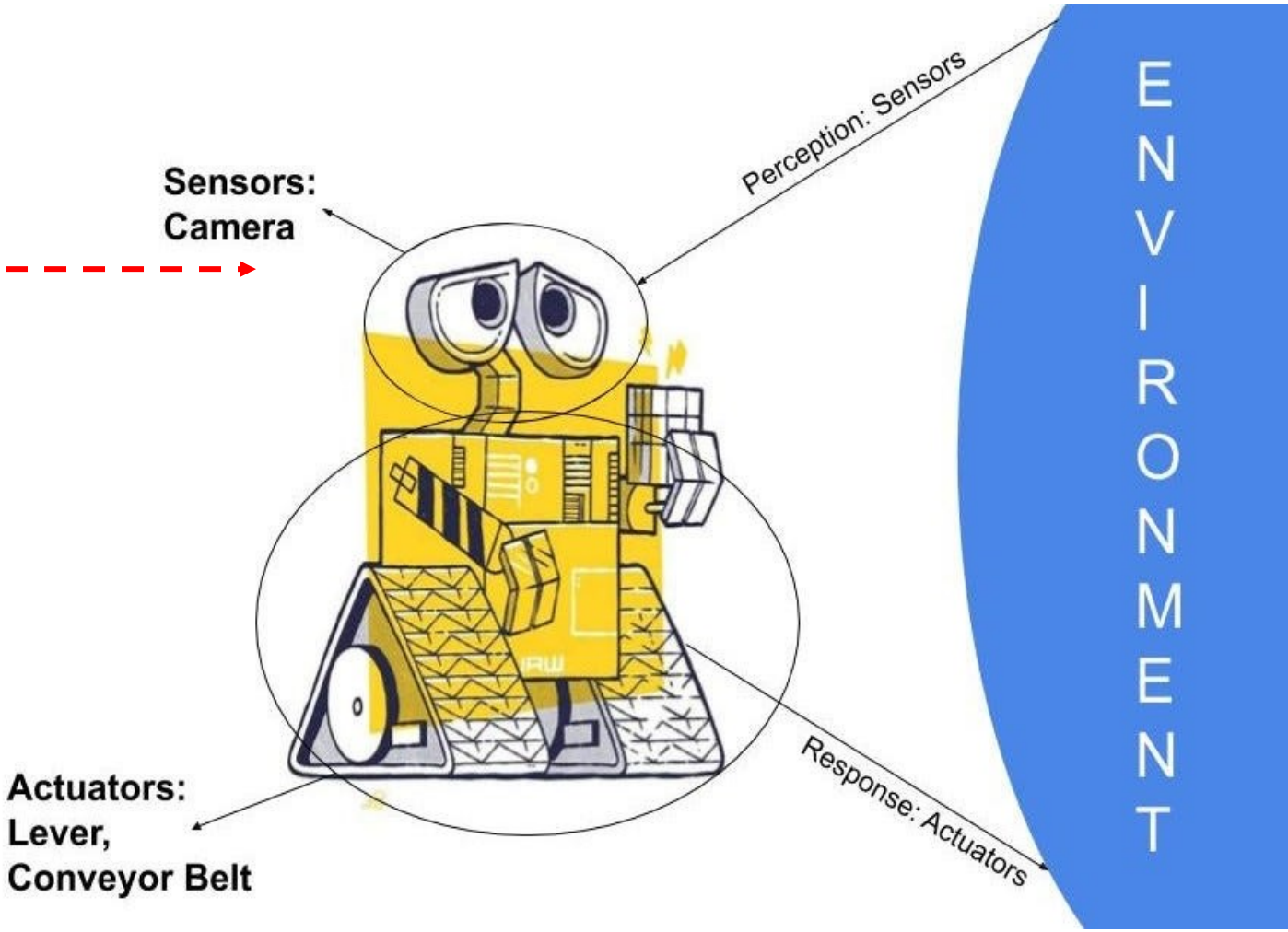
Natural language, knowledge representation, automated reasoning, machine learning

Approaches in AI

4. ***Act Rationally***

Rational Agent Approach

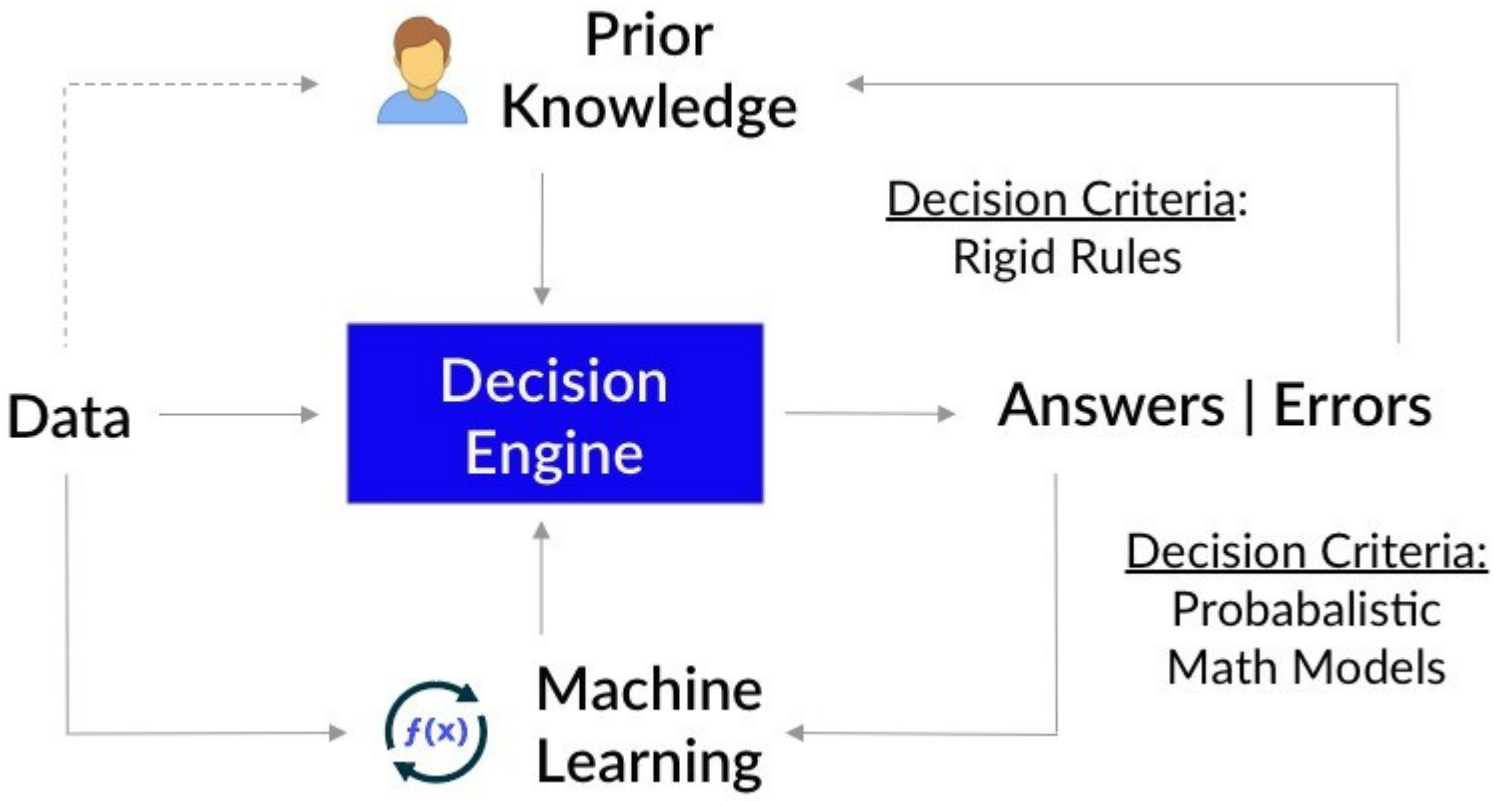
- Building systems that carry out actions to achieve the **best outcome**
- Rational **behaviour**
- May or may not involve rational thinking
  - i.e., consider reflex actions



Other Approaches in AI

Machine Learning approach:

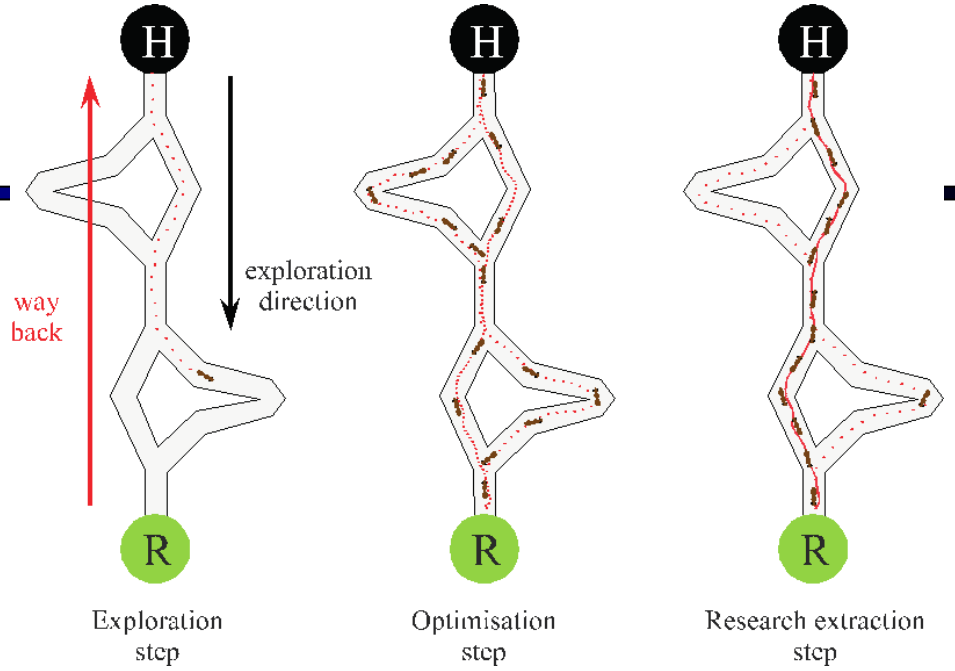
- Involves training machines to *learn from data* and *improve performance* on *specific tasks over time*.
- It is widely used in areas such as *image and speech recognition*, *natural language processing*, and *recommender systems*.



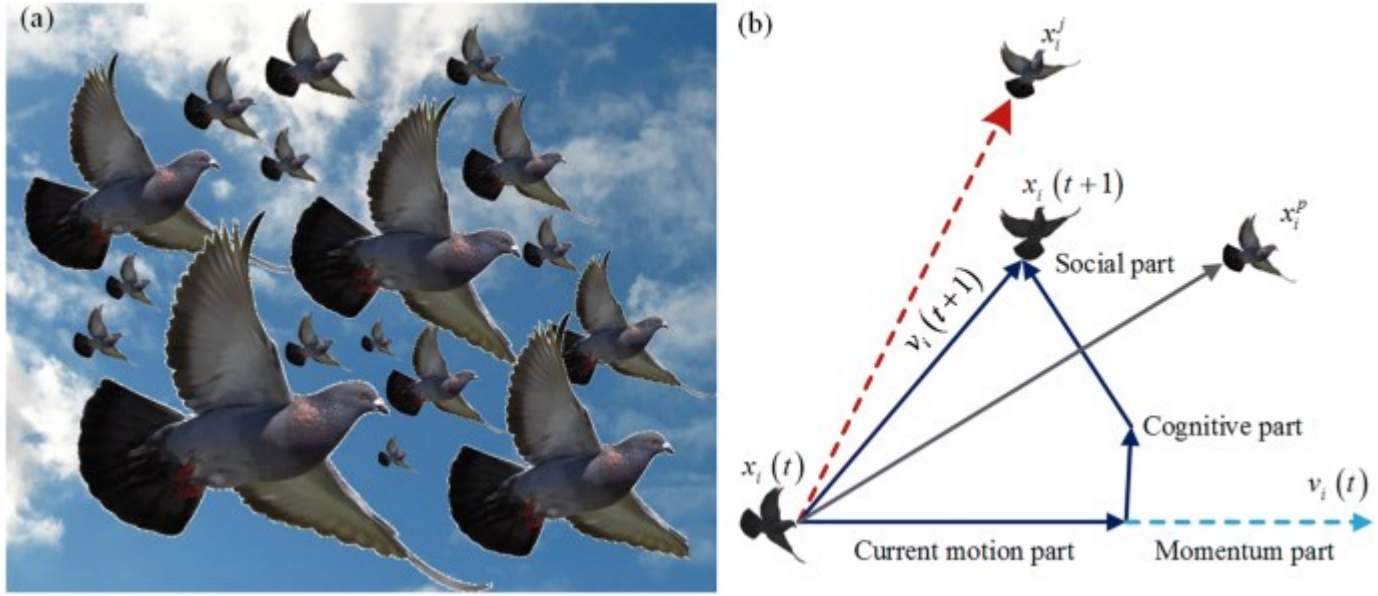
Other Approaches in AI

Evolutionary Intelligence Approach:

- This approach is **inspired** by the **natural process**.
- It involves generating and testing a large number of variations of a solution to a problem, and then selecting and combining the most successful variations to create a new generation of solutions.



Ant Colony Optimization



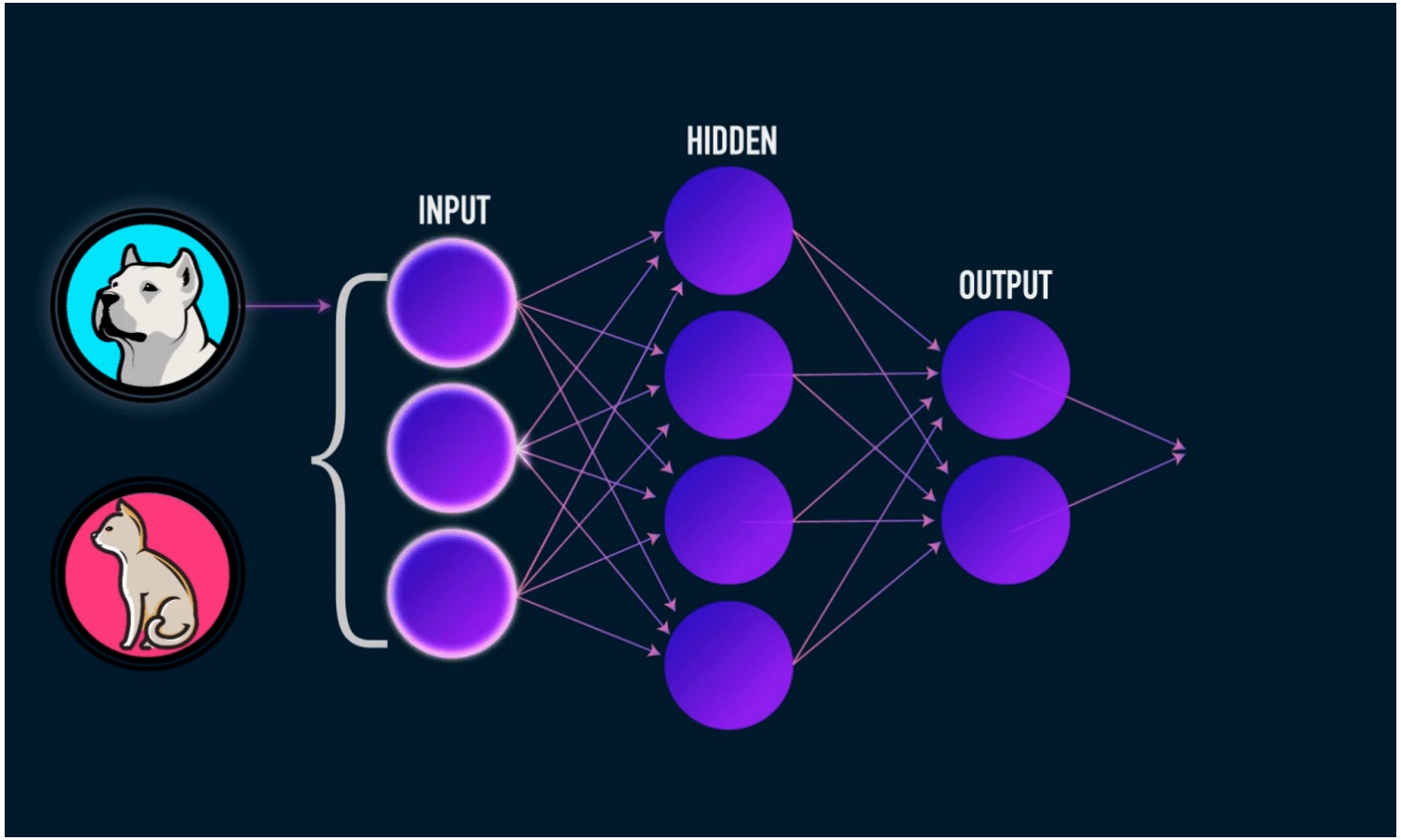
PSO Algorithm

Other Approaches in AI

Neural Networks:

- Involves building artificial neural networks that are modeled after the ***structure and function of the human brain***.

- Neural networks can be used for tasks such as *pattern recognition*, *prediction*, and *decision-making*.

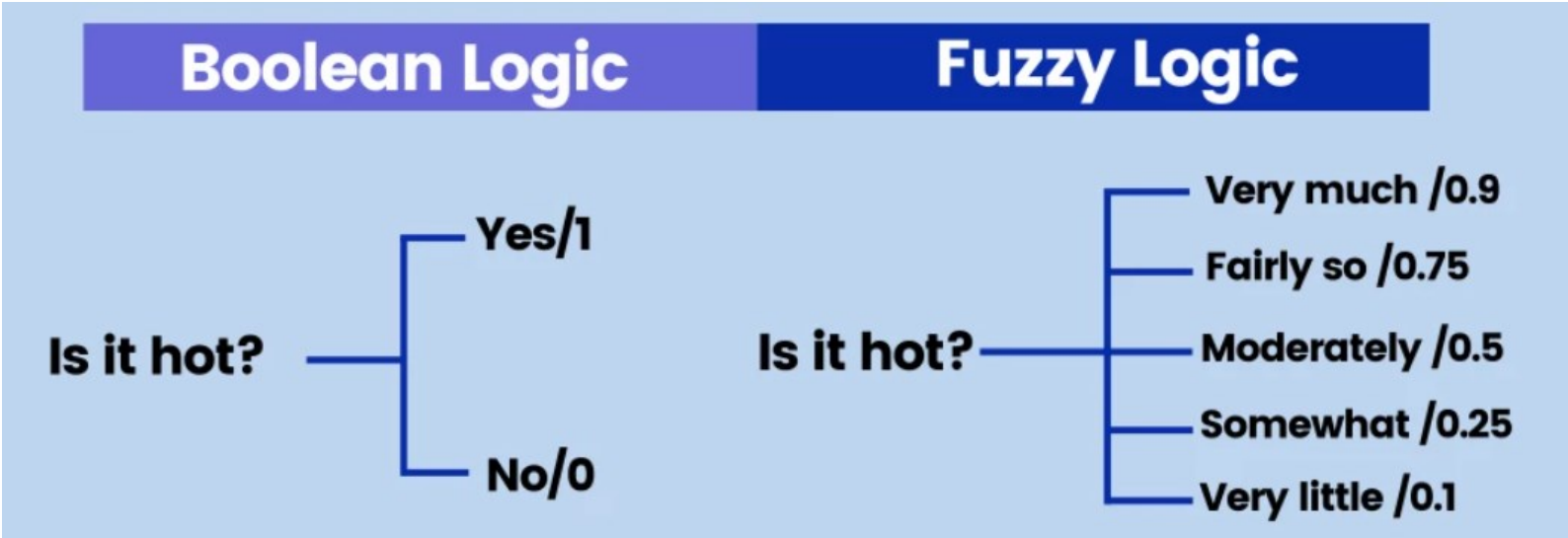


Other Approaches in AI

Fuzzy Logic Approach:

- Involves *reasoning with uncertain* and *imprecise information*, which is common in real-world situations.

- Fuzzy logic can be used to *model and control complex systems* in areas such as *robotics, automotive control*, and *industrial automation*.

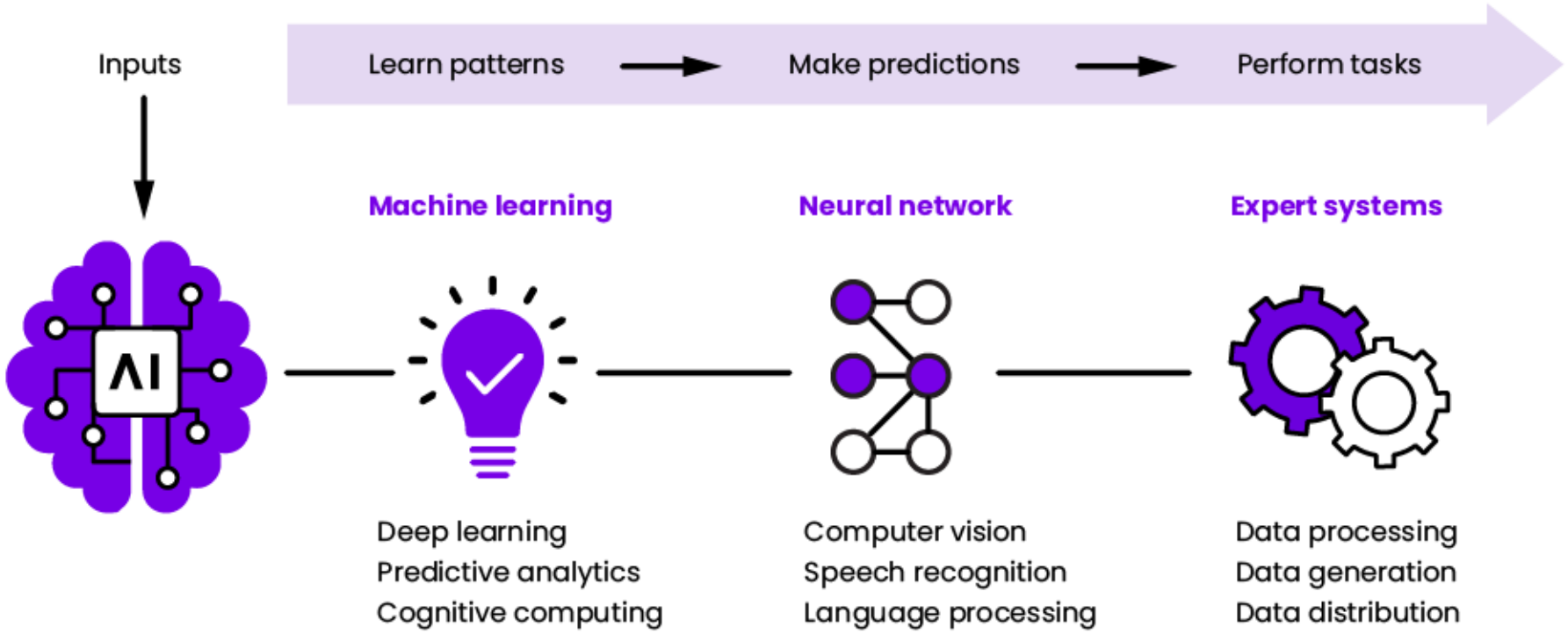


Other Approaches in AI

Hybrid Approach:

Combines multiple AI techniques to solve complex problems

For example, a **hybrid approach** might use ***machine learning to analyze data*** and ***identify patterns***, and then use ***logical reasoning*** to make ***decisions*** based on those patterns.



Drawbacks of AI

Bias and unfairness:

- AI systems can perpetuate and amplify existing biases in data and decision-making.

Lack of transparency and accountability:

- Complex AI systems can be difficult to understand and interpret, making it challenging to determine how decisions are being made.

Job displacement:

- AI has the potential to automate many jobs, leading to job loss and a need for reskilling.

Security and privacy risks:

- AI systems can be vulnerable to hacking and other security threats, and may also pose privacy risks by collecting and using personal data.

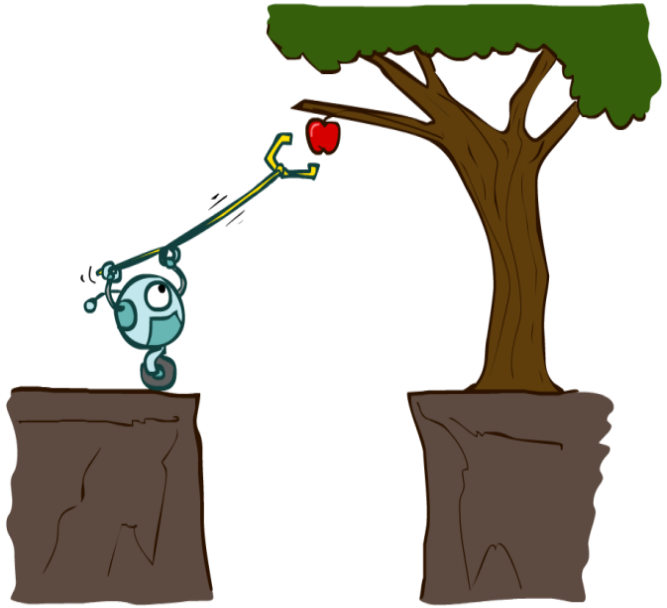
Ethical concerns:

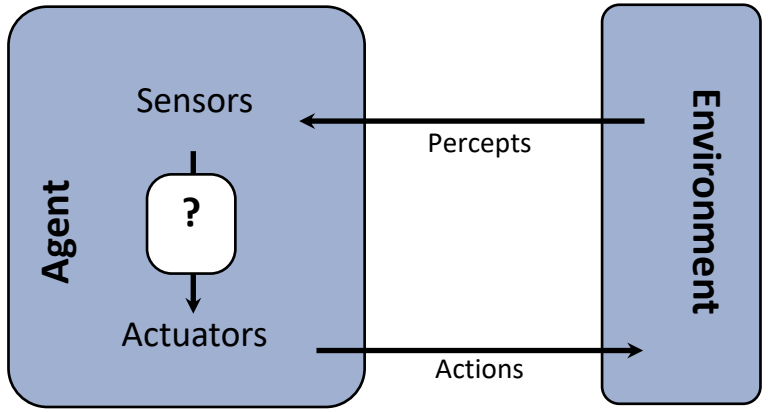
- AI raises important ethical questions about the use of technology for decision-making, including issues related to autonomy, accountability, and human dignity.

Agents

Agents

- **An agent can perceive its environment through sensors and acts upon that environment through effectors.**
  - ***Human Sensors:***
    - Eyes, ears, and other organs for sensors.
  - ***Human Actuators:***
    - Hands, legs, mouth, and other body parts.
  - ***Robotic Sensors:***
    - Mic, cameras and infrared range finders for sensors
  - ***Robotic Actuators:***
    - Motors, Display, speakers etc





Agents

Human-Agent:

- A human agent has eyes, ears, and other organs which work for sensors and
- Hand, legs, vocal tract work for actuators.

Robotic Agent:

- A robotic agent can have cameras, infrared range finder, NLP for sensors and
- Various motors for actuators.

Software Agent:

- Software agent can have keystrokes, file contents as sensory input and
- Act on those inputs and display output on the screen.

Agents

Sensor:

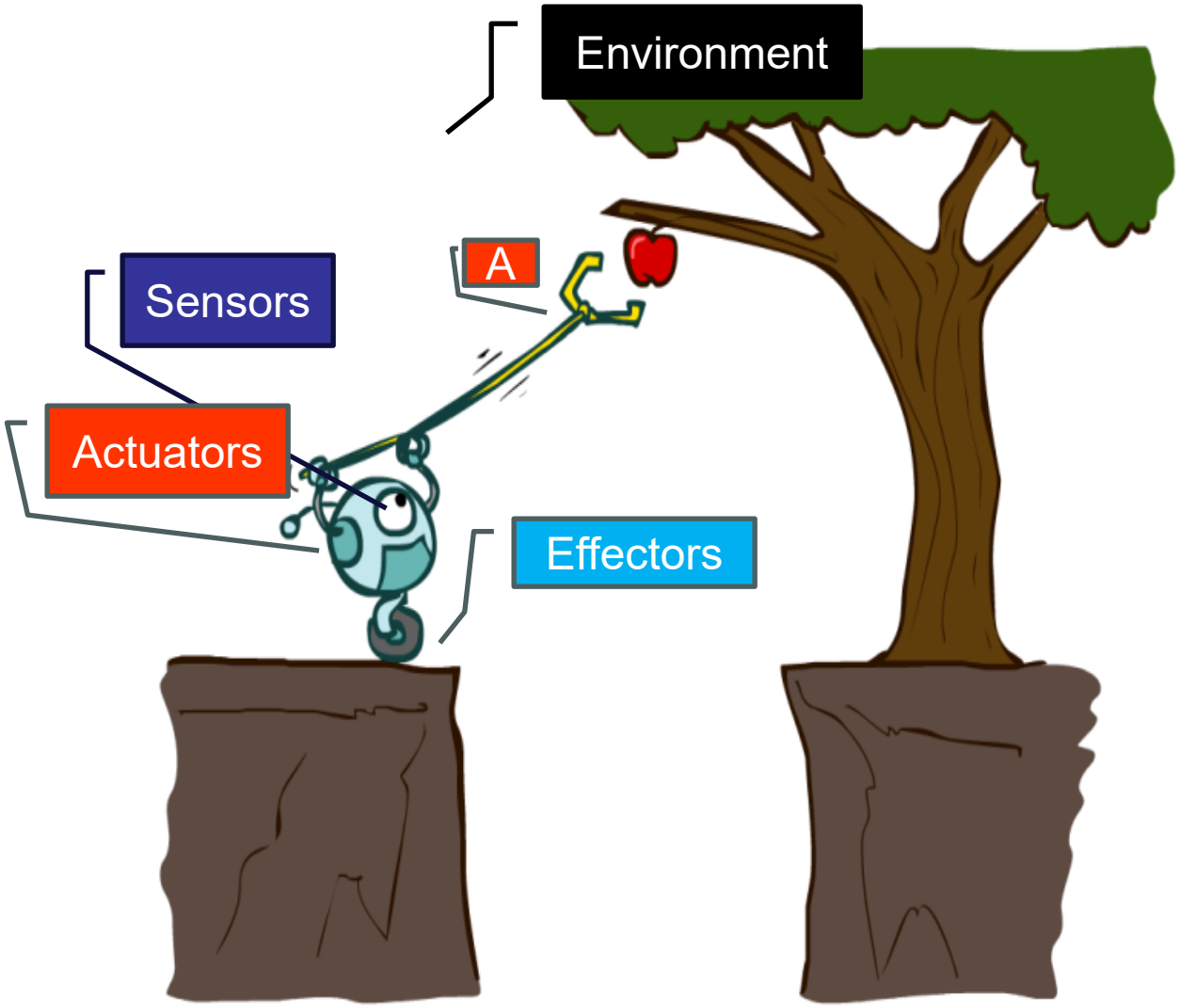
- Sensor is a device which **detects the change** in the environment and sends the information to other electronic devices.
- **An agent observes its environment through sensors.**

Actuators:

- Actuators are the component of machines that **converts energy into motion**.
- Only responsible for moving and controlling a system.
- Can be an electric motor, gears, rails, etc.

Effectors:

- Effectors are the devices which affect the environment. Effectors can be legs, wheels, arms, fingers, wings, fins, and display screen.



Intelligent Vs. Rational Agents

Intelligent Agents

- An intelligent agent is an **autonomous entity which act upon an environment** using **sensors and actuators for achieving goals**.
- An intelligent agent may learn from the environment to achieve their goals.
- Thermostat is an example of an intelligent agent.

4 Rules for an AI Agent

Rule 1: AI agent must have the ***ability to perceive the environment***.
Rule 2: ***Observation*** must be ***used*** to make ***decisions***.
Rule 3: ***Decision*** should ***result*** in an ***action***.
Rule 4: ***Action*** taken by an ***AI agent*** must be a ***rational action***.

Rational Agents

- An agent which has ***clear preference***, ***models uncertainty***, and ***acts in a way to maximize*** its ***performance*** measure with ***all possible actions***.
- Rationality of an agent is measured by its ***performance measure***.

Rationally can be judged:

- Performance measure which defines the **success criterion**.
- Agent **prior knowledge** of its environment.
- **Best possible actions** that an agent can perform.
- The **sequence** of **percepts**.

Agents: Functions

In general, an action may depend on all precepts observed so far, not just the current percept.

$$\boldsymbol{a} = \boldsymbol{F}(\boldsymbol{p})$$

where $\boldsymbol{p}$ is the current percept, $\boldsymbol{a}$ is the action carried out, and $\boldsymbol{F}$ is the agent function

$\boldsymbol{F}$ maps Percepts to Actions $\boldsymbol{F}: \boldsymbol{P} \rightarrow \boldsymbol{A}$

where $\boldsymbol{P}$ is the set of all percept, $\boldsymbol{A}$ is the set of all action

$$\boldsymbol{a_k} = \boldsymbol{F}(\boldsymbol{p_0}, \boldsymbol{p_1}, \boldsymbol{p_2}, \dots, \boldsymbol{p_k})$$

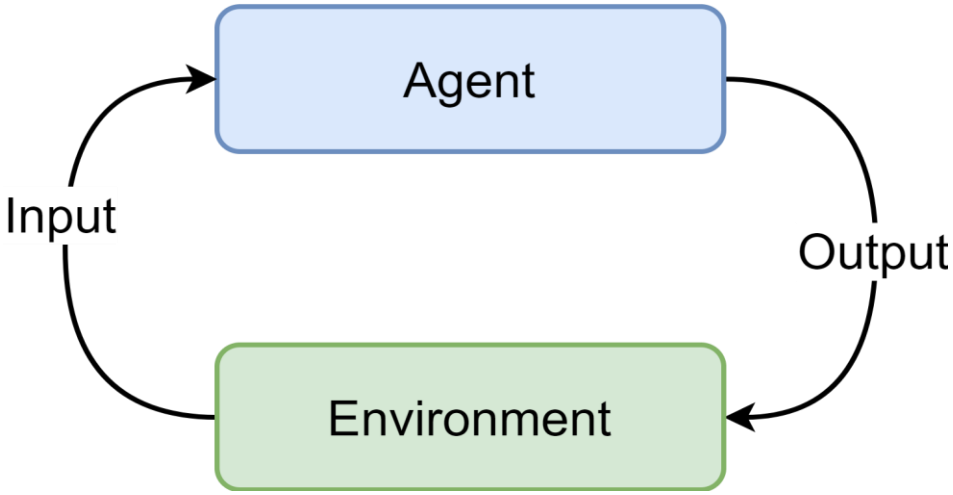
Where $\boldsymbol{p_0}, \boldsymbol{p_1}, \boldsymbol{p_2}, \dots, \boldsymbol{p_k}$ is the sequence of the percepts observed to date, a_k is the resulting action applied.

$\boldsymbol{F}$ now maps Percept Sequences to Actions $\boldsymbol{F}: \boldsymbol{P}^* \rightarrow \boldsymbol{A}$

Structure of Agents and Specifying Tasks

- **Agent** = **Architecture** + **Program**
- **Architecture**
  - Device with sensors and actuators
  - e.g., A robotic car, a camera, a PC, …
- **Program**
  - implements the agent function on the architecture

An abstract view of the agent



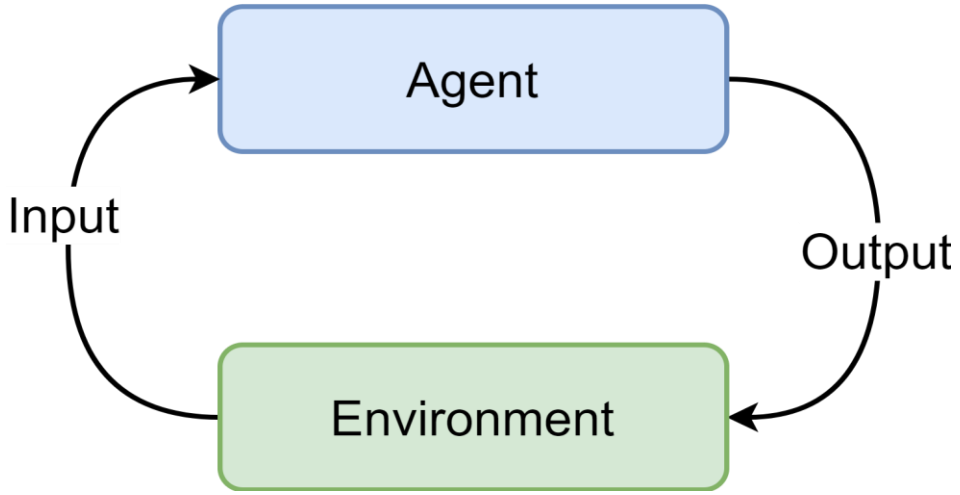
PEAS

- **Performance Measure**: Captures agent's aspiration, e.g.: safety, destination.
- **Environment**: Context, restrictions, e.g.: streets, traffic.
- **Actuators**: Indicates what the agent can carry out, e.g.: steering, brakes.
- **Sensors**: Indicates what the agent can perceive, e.g.: GPS, engine sensors,

Simple Problem Solving by a Agent

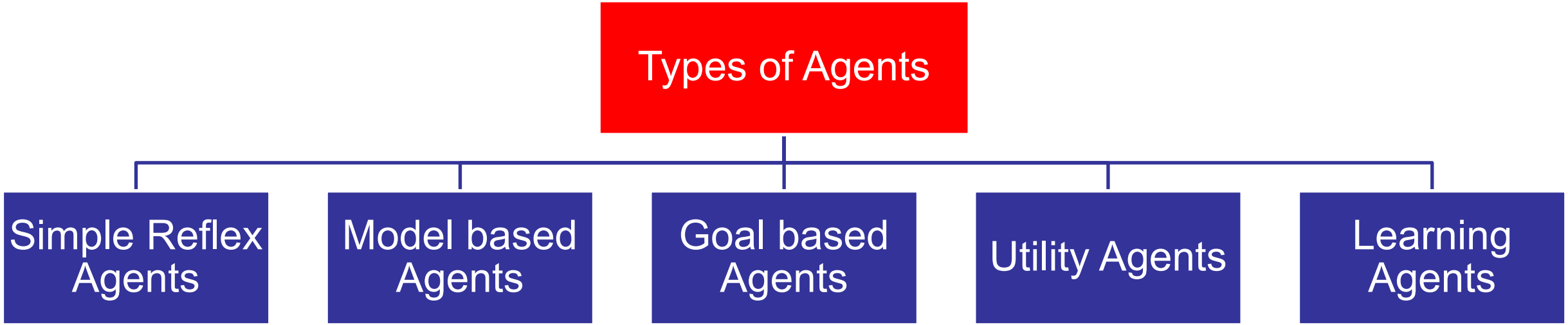
```
function SIMPLE-PROBLEM-SOLVING-AGENT( percept) returns an action
    static: seq, an action sequence, initially empty
            state, some description of the current world state
            goal, a goal, initially null
            problem, a problem formulation

    state ← UPDATE-STATE(state, percept)
    if seq is empty then do
        goal ← FORMULATE-GOAL(state)
        problem ← FORMULATE-PROBLEM(state, goal)
        seq ← SEARCH( problem)
    action ← FIRST(seq)
    seq ← REST(seq)
    return action
```

An abstract view of the agent



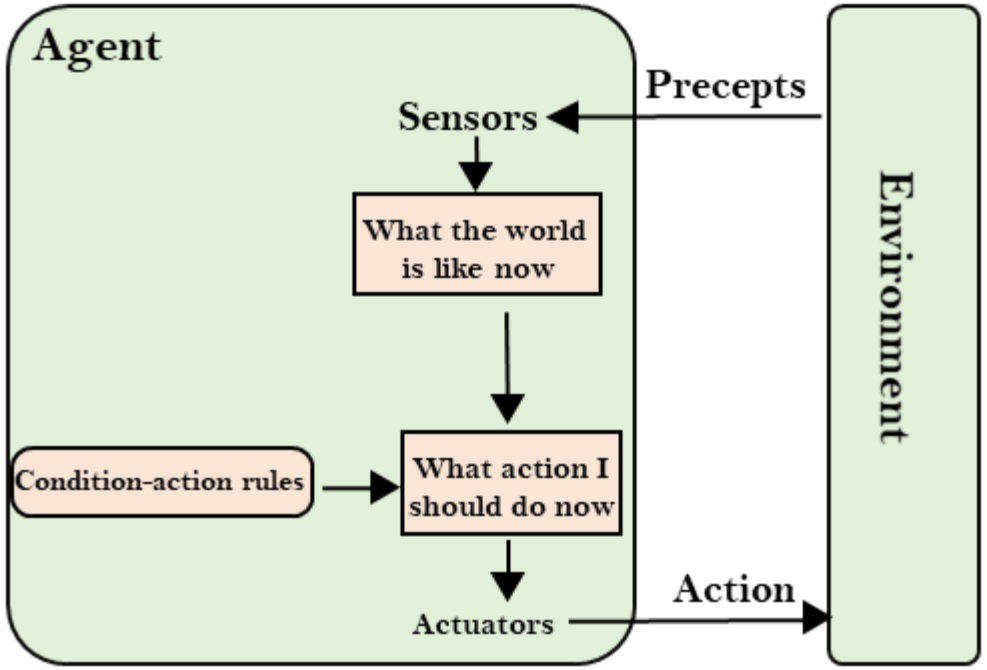
Types of Agents

- Agents can **improve their performance** and **generate better action over the time**.
- Agents can be **grouped into five classes** based on ***their degree of perceived intelligence and capability.***



a. Simple Reflex Agent

- Such agents **take decisions** on **basis of current percepts** and **ignore the rest of the percept history**.
- Such agents ***only succeed*** in the ***fully observable environment***.
- Such agents works on ***Condition-action Rule***,
  - Which means it maps the current state to action.
  - Such as a Room Cleaner agent, it works only if there is dirt in the room.

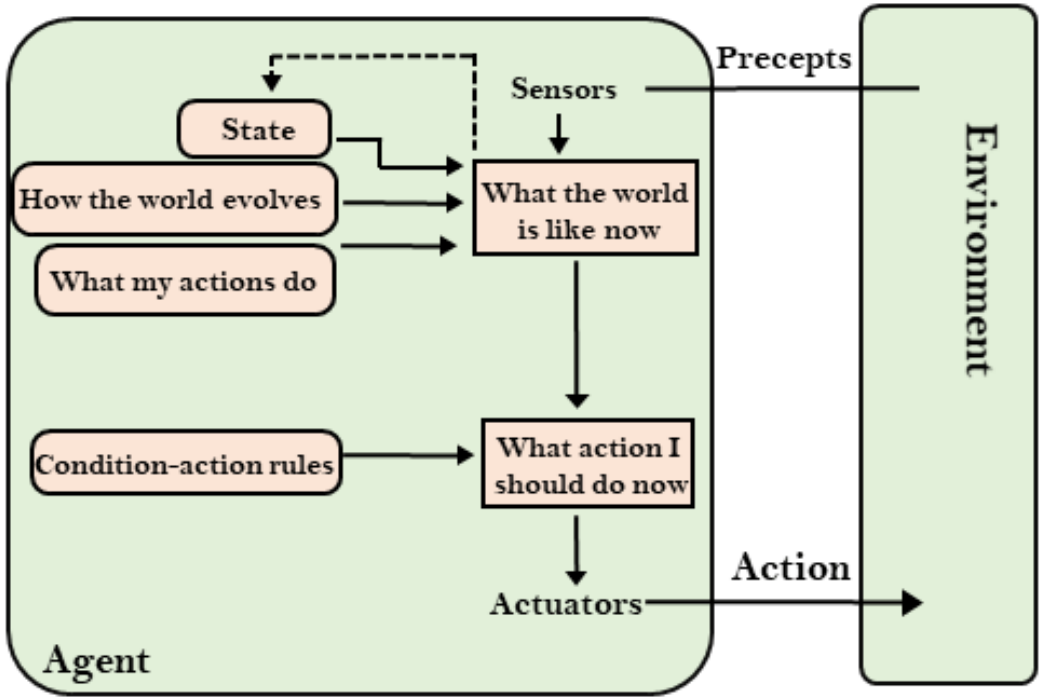


Problems for the simple reflex agent design approach:

- They have very ***limited intelligence***
- They ***do not have knowledge*** of ***non-perceptual parts*** of the ***current state***
- Mostly too big to generate and to store.
- Not ***adaptive to changes in the environment***.

b. Model Based Agent

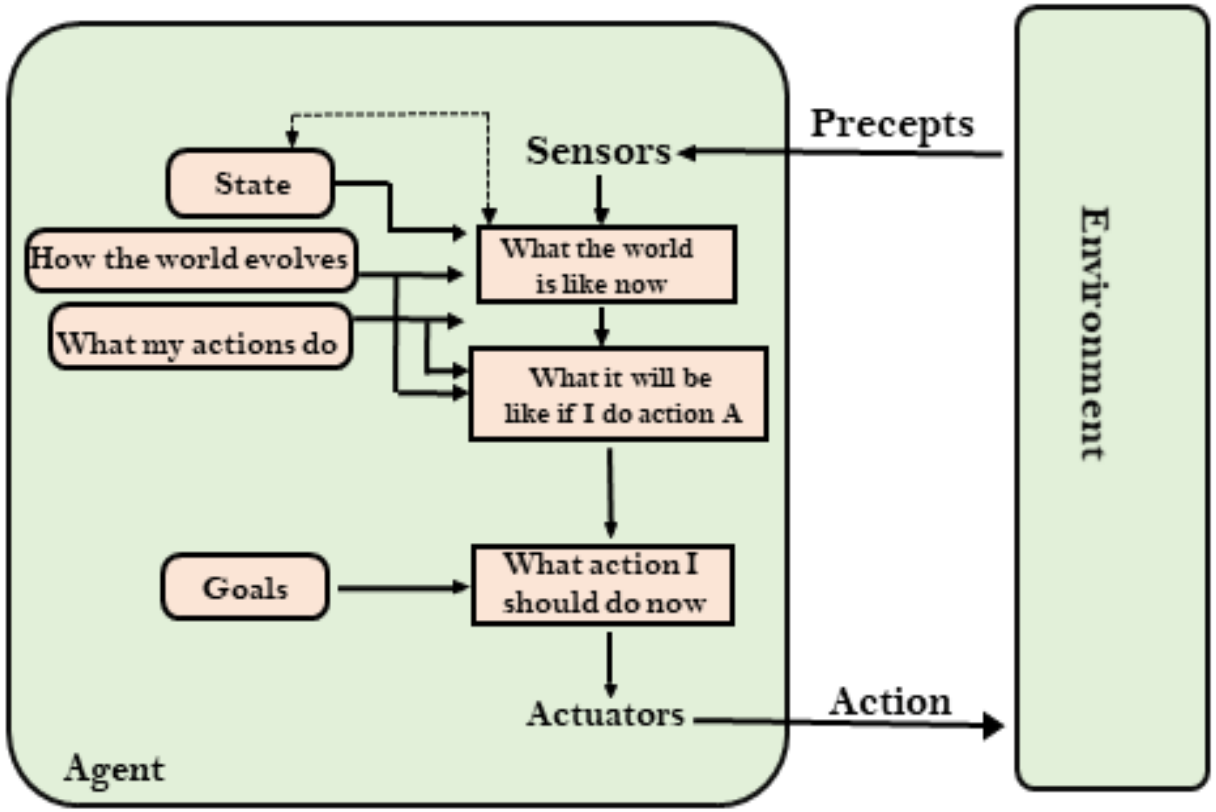
- These agents can work in a ***partially observable environment***, and ***track the situation***.
- *This agent has two important factors:*
- ***Model***: It is knowledge about "***how things happen in the world***," so it is called a Model-based agent.
- ***Internal State***: It is a ***representation*** of the ***current state based*** on ***percept history***.



- These agents have **model**, "**which is knowledge of the world**" and **based on the model they perform actions**.
- **Updating the agent state requires information about:**
  - ***How the world evolves?***
  - ***How the agent's action affects the world?***

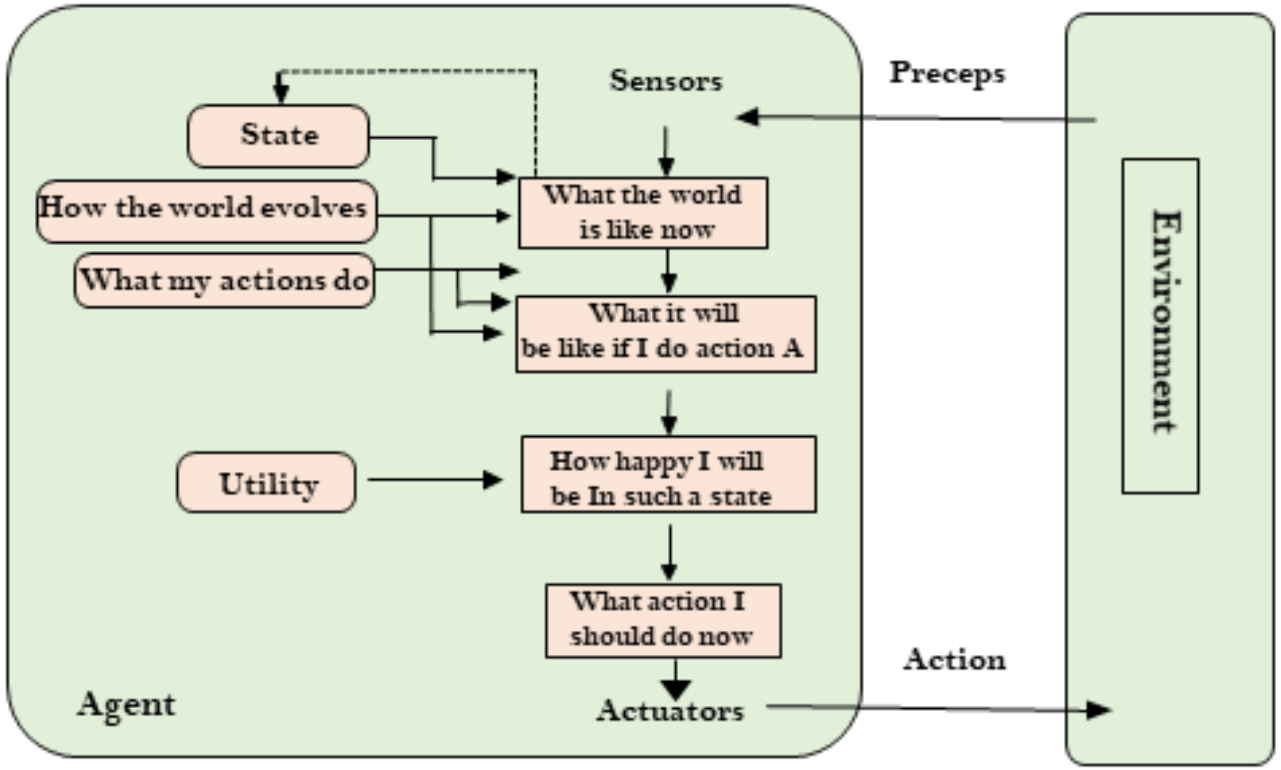
c. Goal Based Agent

- **Knowledge** of the **current state environment** is **not always sufficient** to **decide** for an **agent to what to do**.
- Agent needs to **know its goal** which describes desirable situations.
- Goal-based agents **expand the capabilities** of the **model-based agent** by having the "**goal**" information.
- These agents may have to **consider a long sequence of possible actions before deciding whether the goal is achieved or not**.
- Such considerations of different scenario are called **searching and planning**, which makes an **agent proactive**.



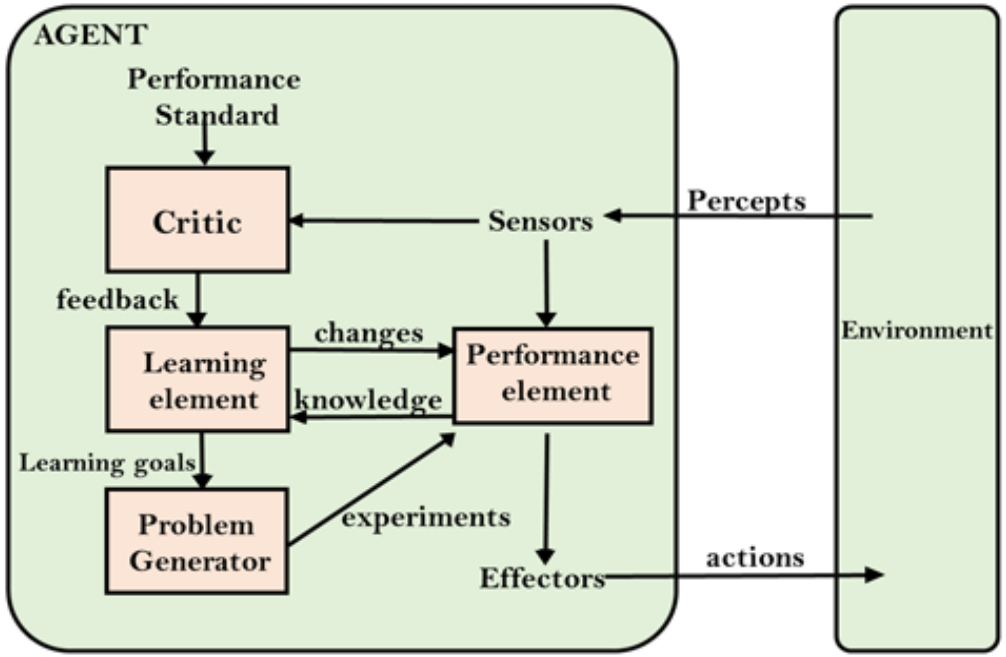
d. Utility Based Agent

- **Similar** to the **goal-based** agent but provide an ***extra component*** of **utility measurement** which makes them different by **providing a measure** of **success at a given state**.
- Such agents act based not only goals but also the best way to achieve the goal.
- Such a agent is **useful** when there are **multiple possible alternatives**, and an ***agent has to choose*** in ***order to perform the best action***.
- **Utility function** ***maps each state*** to a ***real number*** to ***check how efficiently*** each ***action achieves the goals***.



e. Learning Agent

- Such agent in AI is the type of agent which **can learn from its past experiences**, or it **has learning capabilities**.
- It **starts** to **act** with **basic knowledge** and then **able** to **act** and **adapt automatically through learning**.

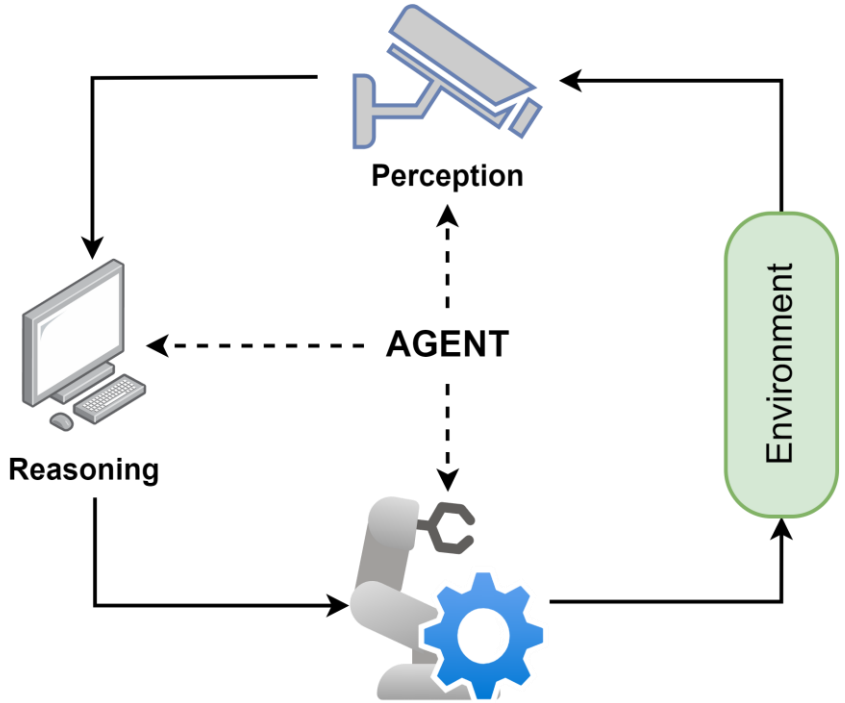


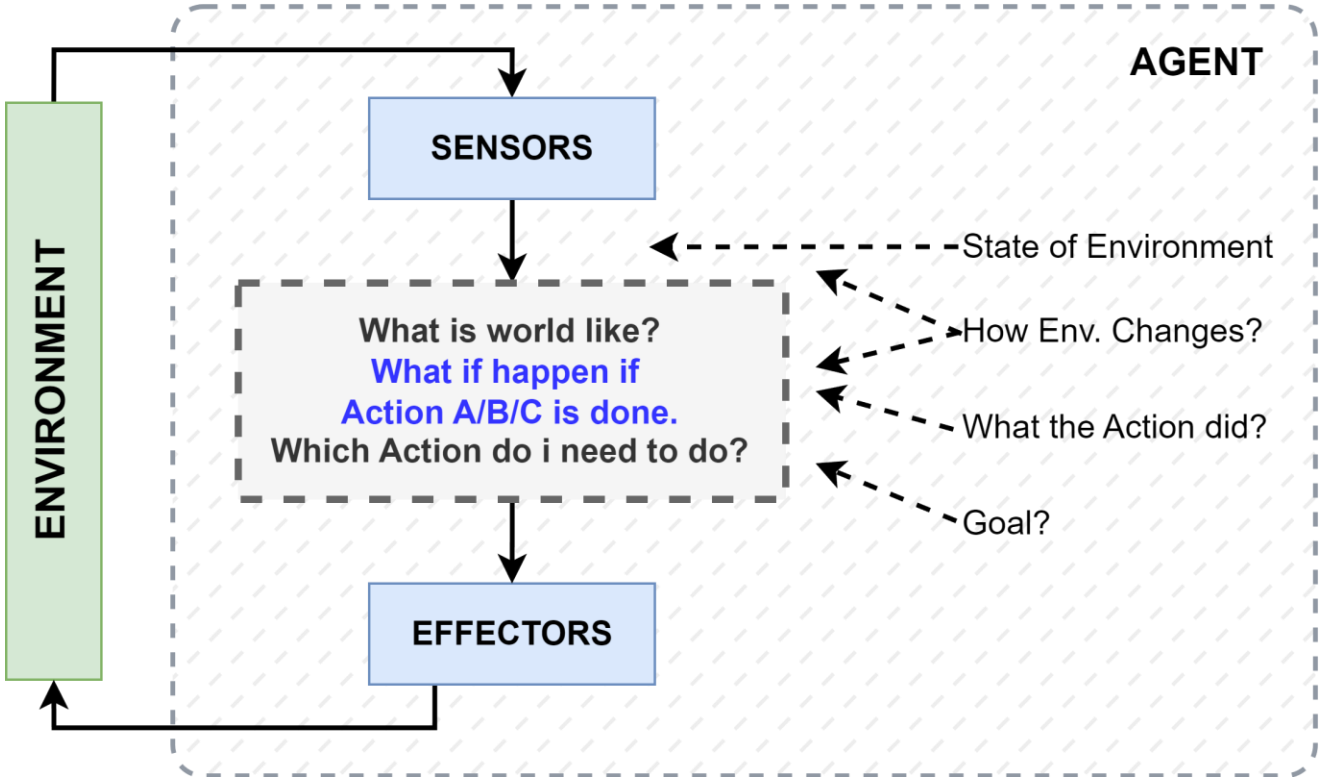
Has **four** conceptual components:

1. **Learning element**: Responsible for **making improvements** by **learning** from **environment**
2. **Critic**: Learning element takes **feedback from critic** which describes that **how well the agent is doing** with **respect to a fixed performance standard**.
3. **Performance element**: It is responsible for **selecting external action**
4. **Problem generator**: Is responsible for **suggesting actions** that will lead to **new and informative experiences**.

Rationality based Agent

- Rationality in AI refers to the principle that **such agents should consistently choose actions** that are **expected to lead to the best possible outcomes**, given their **current knowledge** and the **uncertainties present** in the **environment**.





Rationality based Agent

- Principle of rationality ***guides the behavior of intelligent agents*** in the following ways:

Perception and Information Processing:

Rational agents **strive to perceive** and **process information efficiently** to **gain the most accurate understanding** of their **environment**.

Reasoning and Inference:

Employ logical reasoning and **probabilistic inference** to make **informed decisions** based on **available evidence** and **prior knowledge**.

Decision-Making Under Uncertainty:

When **faced with uncertainty**, rational agents **weigh the probabilities of different outcomes** and **choose actions** that **maximize** their **expected utility** or **achieve** the **best possible outcome** given the **available information**.

Adaptation and Learning:

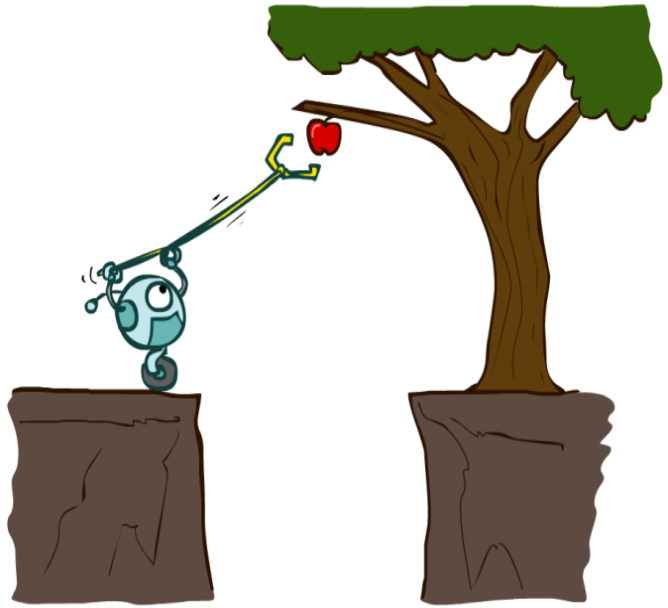
Rational agents **adapt** their **behavior over time** based on **feedback** and **experience**, **continuously refining** their **decision-making strategies** to **improve performance** and **achieve** their **goals more effectively**.

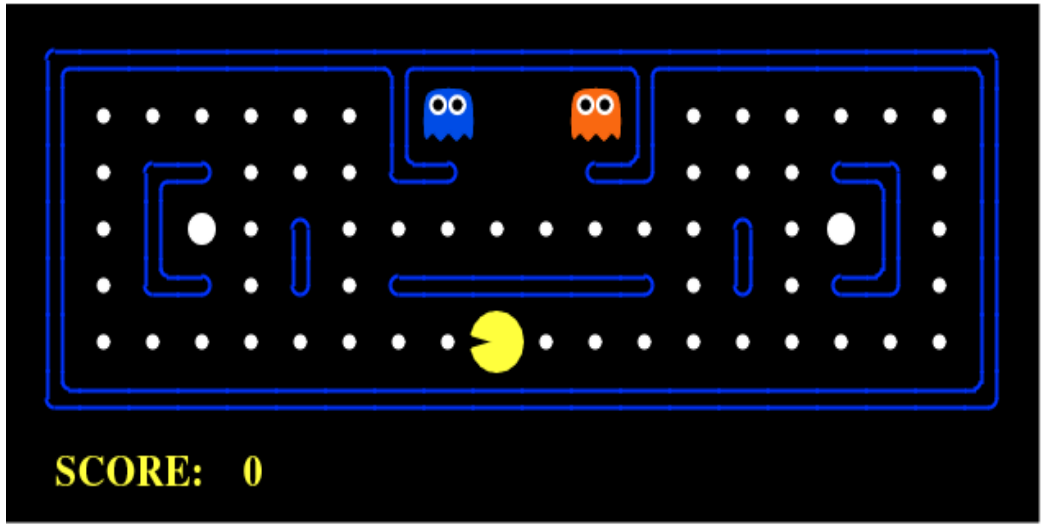
Environment

Environment

- **Environment** is ***everything in the world*** which ***surrounds*** the ***agent***, but it is **not a part of an agent itself**.
- An **environment** can be described as a **situation in which an agent is present**.
- **Environment** is ***where agent lives***, ***operate*** and ***provide*** the ***agent*** with ***something*** to ***sense*** and ***act upon*** it.
- **Examples:**

Agents		Environments
Robot	→	Room
Chatbot	→	Chatting
Vehicle	→	Road
Program	→	Data & Rules
Machine	→	Working Field



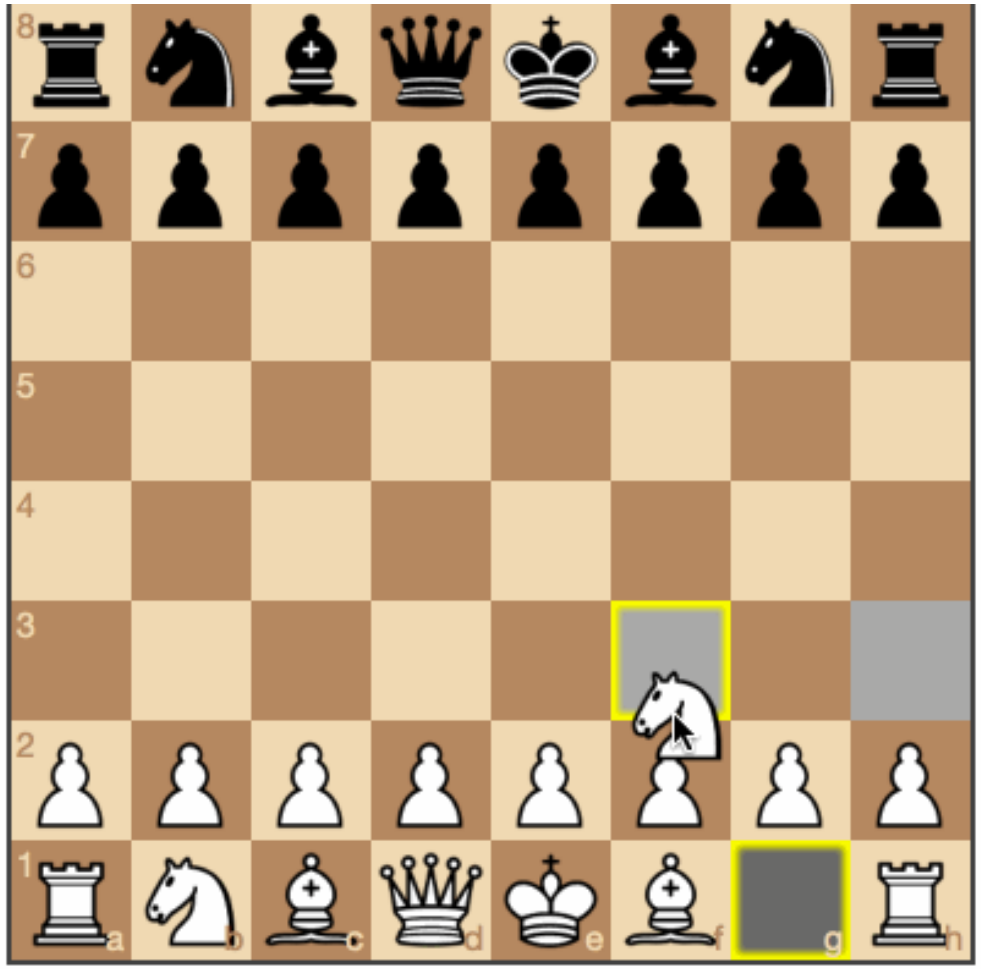


Fully v/s Partially Observable Environment

- If an **agent sensor** can **sense or access the complete state** of an **environment** at **each point of time** then it is **a fully observable** environment, else it is **partially observable**.
- **Fully observable environment** is **easy** as ***there is no need to maintain*** the ***internal state*** to ***keep track history*** of the ***world***.
- An **agent with no sensors** in all environments then such an environment is called as **unobservable**.

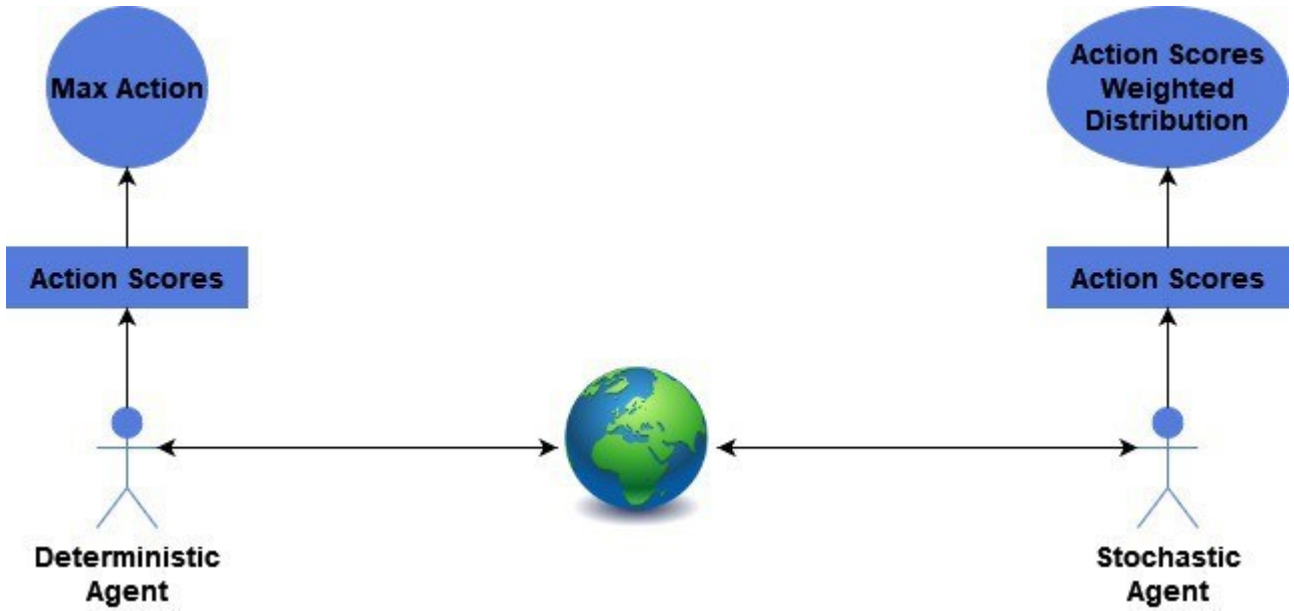
Example:

- **Chess** – the board is fully observable, as are opponent's moves.
- **Driving** – what is around the next bend is not observable and hence partially observable.



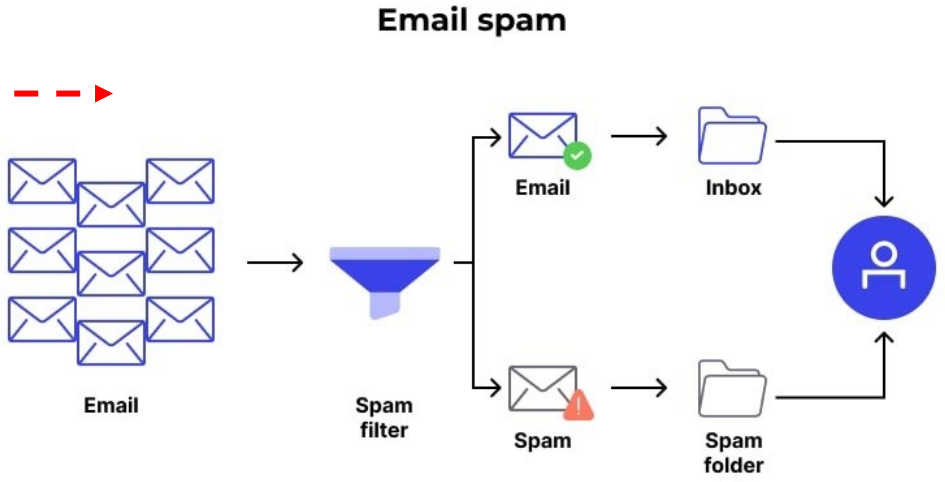
Deterministic v/s Stochastic Environment

- **Deterministic Environment:**
  - If an **agent's current state** and **selected action** can **completely determine** the **next state** of the **environment.**
- **Stochastic Environment:**
  - Is **random** in **nature** and **cannot** be **determined completely** by an **agent**.
- In a **deterministic, fully observable environment**, agent **does not need** to **worry about uncertainty**.

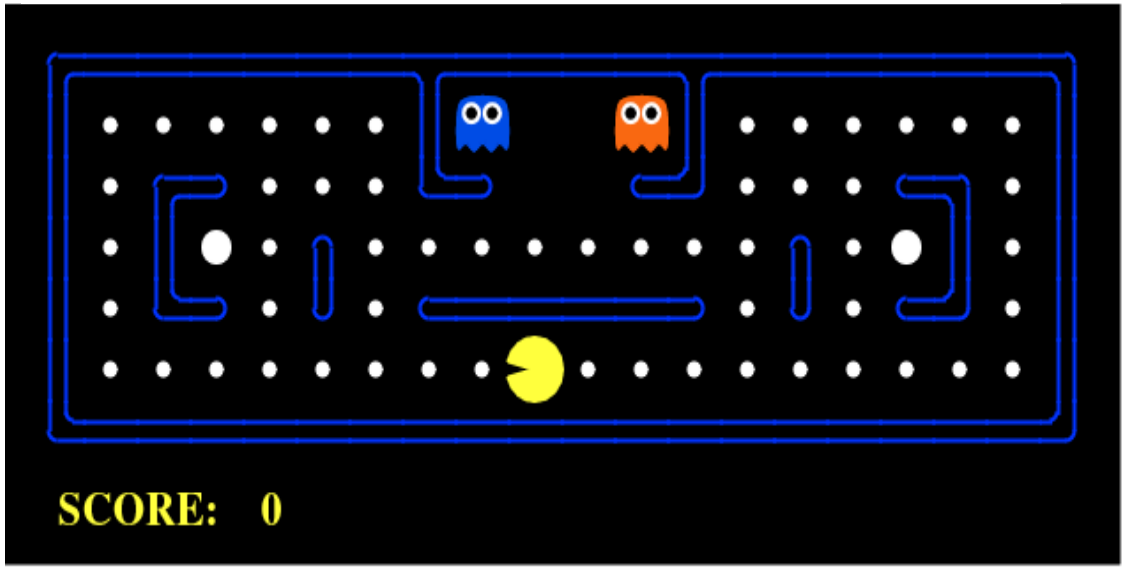


Episodic v/s Sequential Environment

- **Episodic Environment**
  - There is a **series of one-shot actions**, and **only the current percept** is **required for the action**.



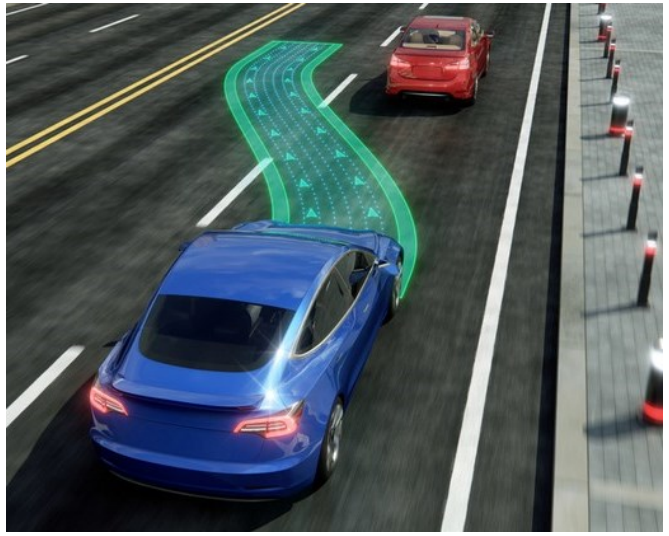
- **Sequential Environment:**
  - An agent **requires memory of** past **actions** to **determine** the **next best actions**.



Static v/s Dynamic Environment

- If the ***environment can change itself*** while an ***agent is deliberating*** then such environment is called a ***dynamic environment*** else it is called a **static environment**.
- **Static environments** are **easy to deal** because an **agent does not need to continue looking at the world** while **deciding for an action**.
- However for **dynamic environment**, **agents need** to **keep looking** at the **world** at **each action**.

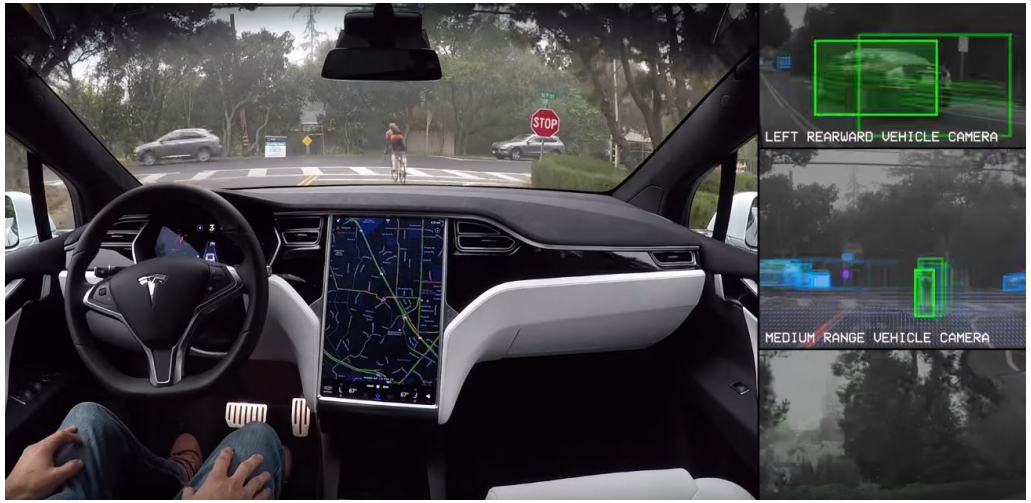
For example: Driving is an example of a dynamic environment whereas Crossword puzzles are an example of a static environment.



Discrete v/s Continuous Environment

- **Discrete Environment:**
  - If in an environment there are a **finite number of precepts and actions** that can be performed within it.
  - A chess game comes under discrete environment as there is a finite number of moves that can be performed.

- **Continuous Environment:**
  - If in an environment there are a **infinite number of precepts** and **actions**
  - A self-driving car is an example of a continuous environment



Single Agent v/s Multi Agent Environment

Single Agent Environment

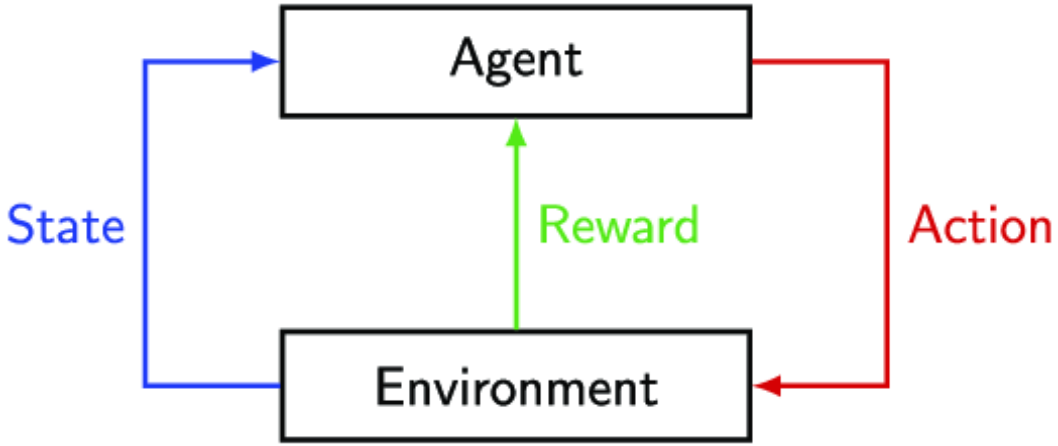
- If only **one agent** is **involved** in an **environment**, and **operating** by **itself**.

Multi Agent Environment

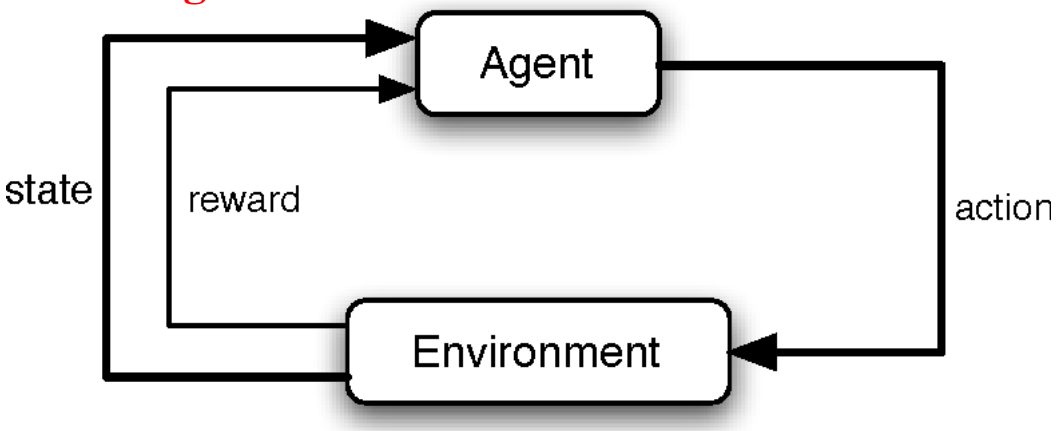
- If **multiple agents** are **operating** in an **environment**.

The agent design problems in the multi-agent environment are different from single agent environment.

Single Agent



Multi-Agent



Other Environments

Known vs. Unknown

- **Known** and **unknown** are **not** actually a **feature** of an **environment**, but it is **an agent's state of knowledge to perform an action**.
- **In known environment:** ***Results*** for all ***actions*** are ***known*** to the ***agent***.
- **In unknown environment:** agent ***needs*** to ***learn how*** it ***works*** in ***order*** to ***perform*** an action.
- ***Quite possible*** that a ***known environment*** to be ***partially observable*** & an ***Unknown environment*** to be ***fully observable***.

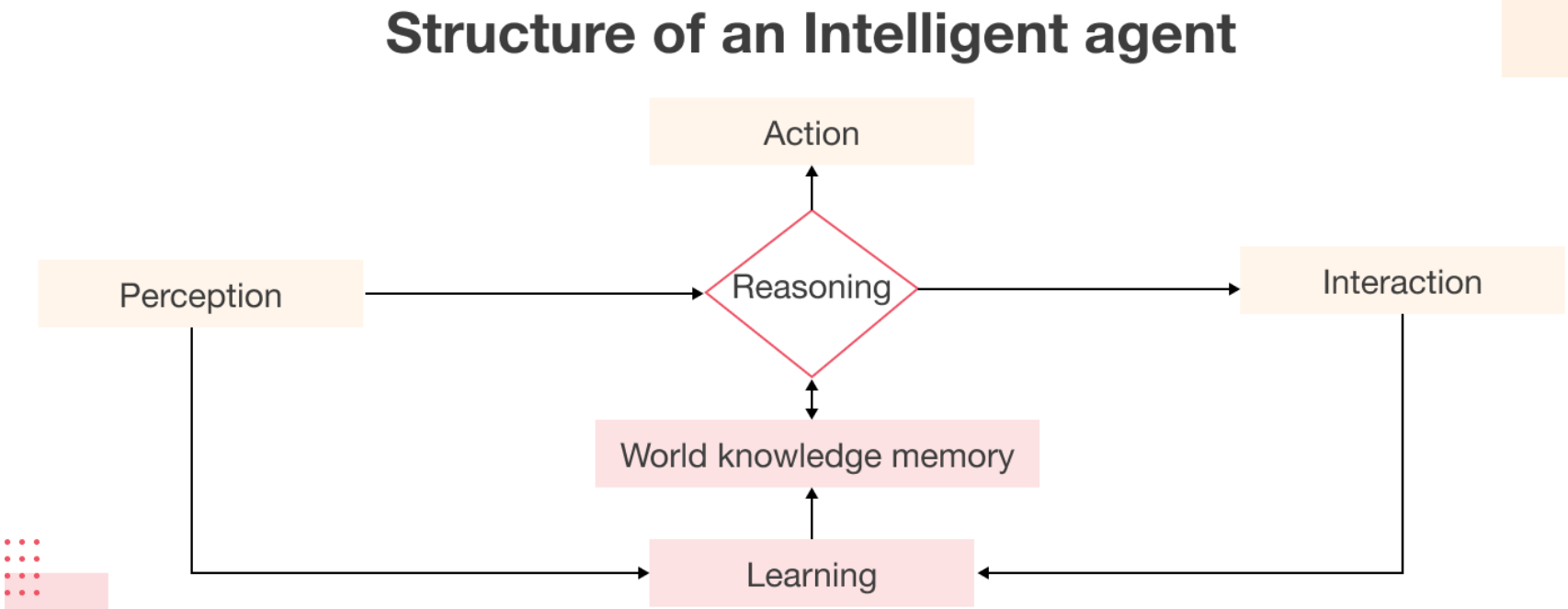
Accessible vs. Inaccessible

- **Accessible environment**: If an **agent** can **obtain complete** and **accurate information** about the **state's environment**, else it is called inaccessible.
- An **empty room** whose **state** can be **defined** by its **temperature** is an example of an **accessible environment**.
- **Information** about an **event** on **earth** is an example of **Inaccessible environment**.

Agent Architecture

Agent Architecture

- **It examine the complex structures that shape how machines perceive, reason, and act in their environments in the pursuit of autonomous intelligence.**
- Offers a ***methodical framework*** for ***creating***, ***putting into practice***, and ***comprehending agents*** that may ***independently interact with their surroundings*** to ***accomplish predetermined goals***.



Characteristics of a Agent

Autonomy:

- An AI virtual agent is capable of **performing tasks independently** without requiring constant human intervention or input.

Perception:

- The agent function **senses and interprets the environment** they operate in through various sensors, such as cameras or microphones.

Reactivity:

- AI agent can **assess the environment** and **respond accordingly** to achieve its goals.

Reasoning and decision-making:

- AI agents are **intelligent tools** that can **analyze data** and **make decisions** to **achieve goals**. They use **reasoning techniques** and **algorithms** to **process information** and take **appropriate actions**.

Learning:

- They can **learn and enhance their performance** through **machine**, **deep**, and **reinforcement learning** elements and techniques.

Communication:

- AI agents **can communicate** with **other agents** or **humans** using different methods, like understanding and responding to natural language, recognizing speech, and exchanging messages through text.

Goal-oriented:

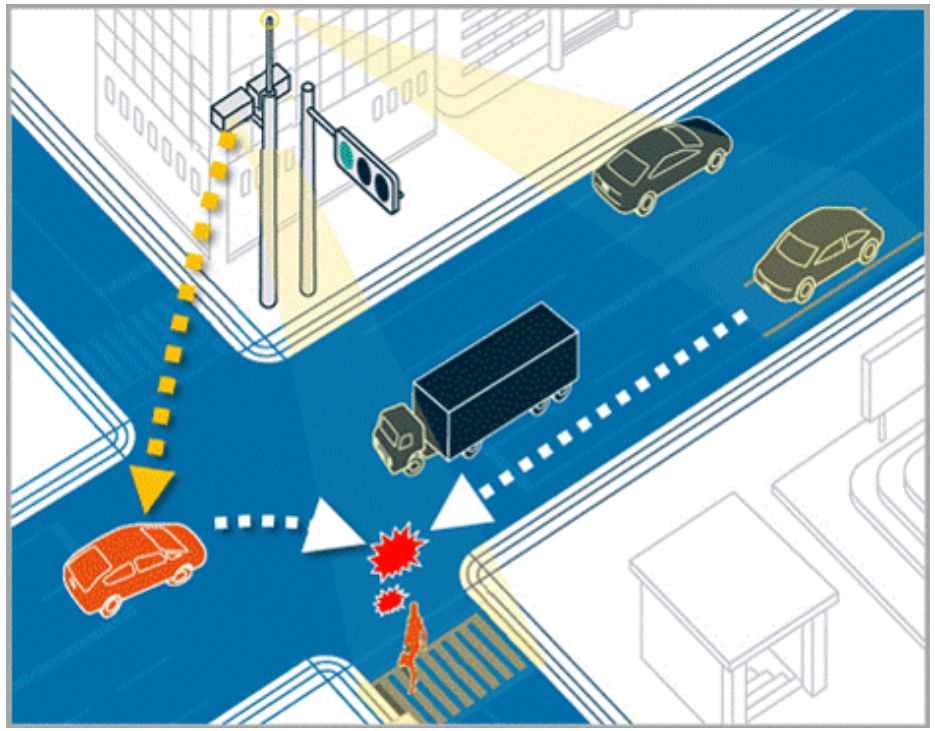
- They are designed to **achieve specific goals**, which can be pre-defined or **learned through interactions with the environment**.

1. Profiling Module:

The Eyes and Ears of the Agent

- Agent is **endowed** with **sensory expertise** via the **profiling module**, it is also known as the **perception module**.
- With the help of this, the **agent may collect and analyze information** from its **environment** like how human senses work.
- It ***interprets unprocessed sensory data***, ***eliminates superfluous details***, and ***extracts important characteristics*** that are essential for **making decisions**.
- The agent's **accurate and thorough grasp** of its **surroundings** is ensured by the **profile module**, which helps it comprehend *visual signals*, *recognize speech patterns*, and *sense tactile inputs*.

Example: a **self-driving automobile** that has **radar**, **lidar**, and **camera sensors**. Through the processing of data from various sensors, the **profile module recognizes *barriers***, ***traffic lights***, and ***lane markers***, allowing the ***vehicle*** to ***go safely***.

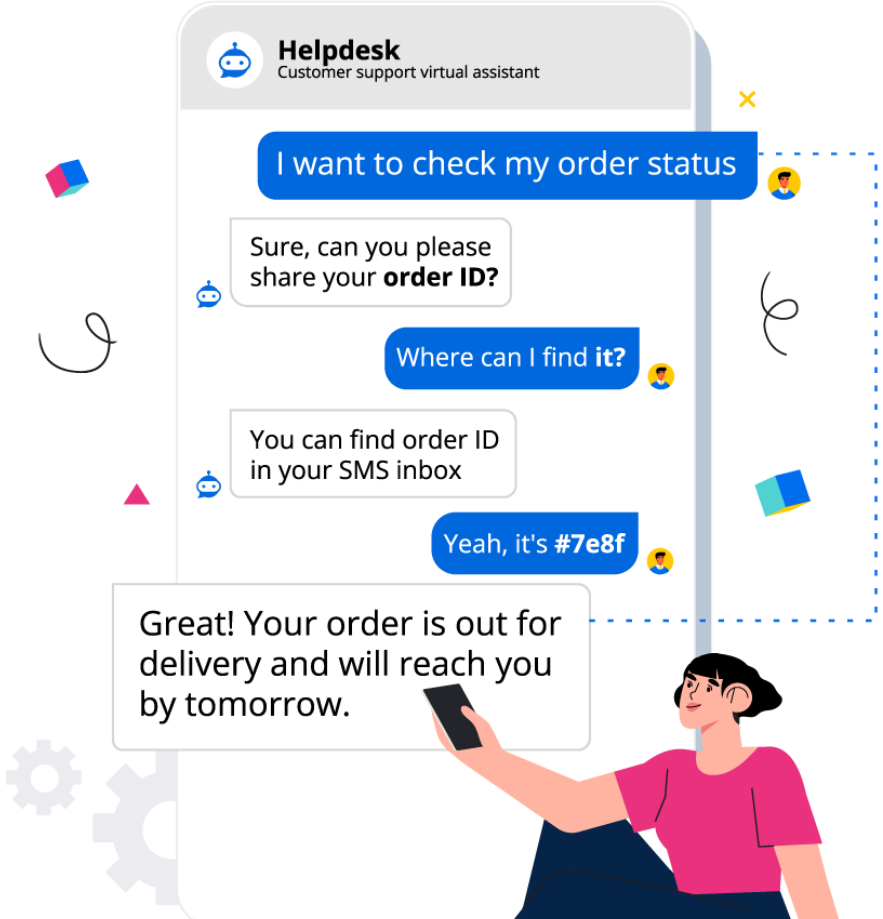


2. Memory Module:

The Repository of the Knowledge

- Memory module is **essential** for **organizing** and **storing data** as it serves as the **agent's knowledge base**.
- **Allows** the **agent** to **remember information**, **rules**, **patterns**, and **events** from the past.
- Functions similarly to the **agent's short- and long-term memory**.
- **For the agent to make wise judgments**, **learn from its interactions**, and **have a sophisticated grasp of its surroundings**, this module is crucial.

Example: To **provide timely and correct replies during discussions**, a **virtual assistant chatbot**, for example, uses its **memory module** to **remember client requests**, **product specs**, and **commonly asked questions**.

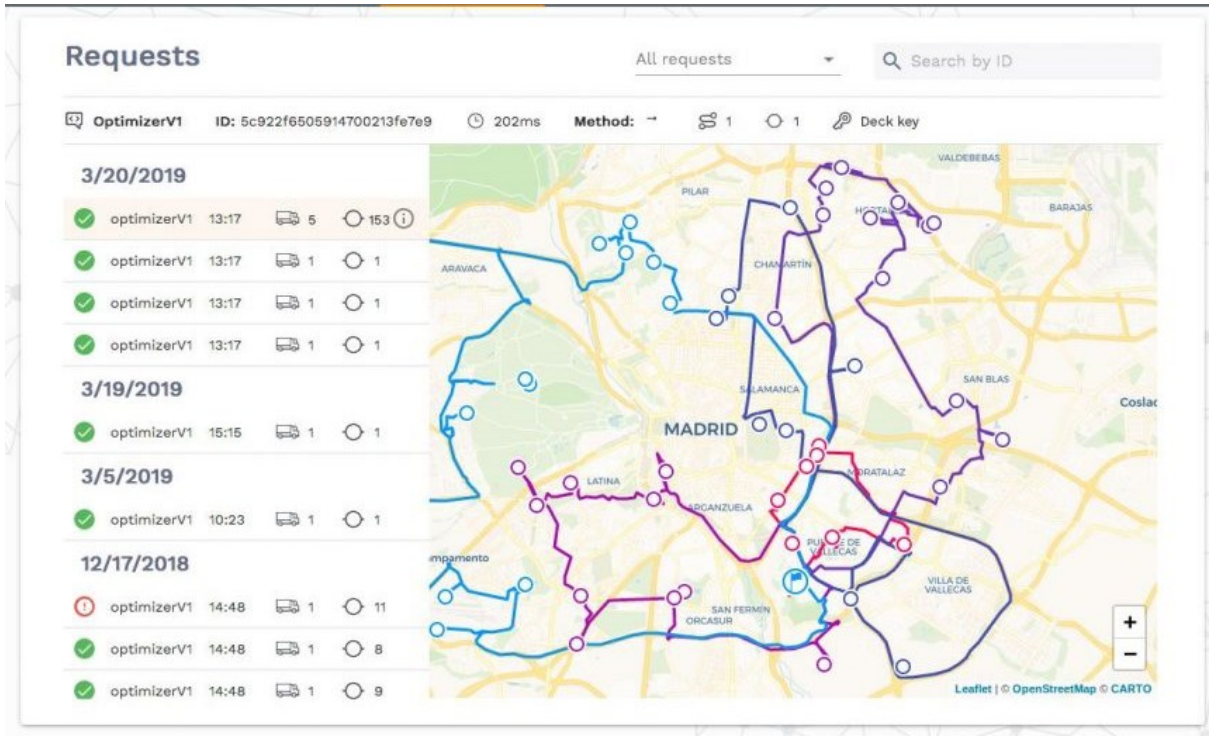


3. Planning Module:

The Strategist behind Decisions

- Agent's **central decision-making command post** for **making decisions** and acting in a **goal-oriented manner** is the planning module.
- **Examines** the **present situation** and **determines** the **best course** of **action** to **accomplish** the **agent's goals**, taking into **account information** from the **memory** and **profiling modules**.
- Using strategies like ***optimization approaches***, ***search algorithms***, ***or rule-based systems***, the planning module figures out the best course of action.

Example: A **delivery routing agent** utilizes the planning module to **optimize** the **most effective routes** and **timetables** for its **fleet** by **taking** into **account variables** like **vehicle capacity**, **traffic patterns**, and **delivery priority**.

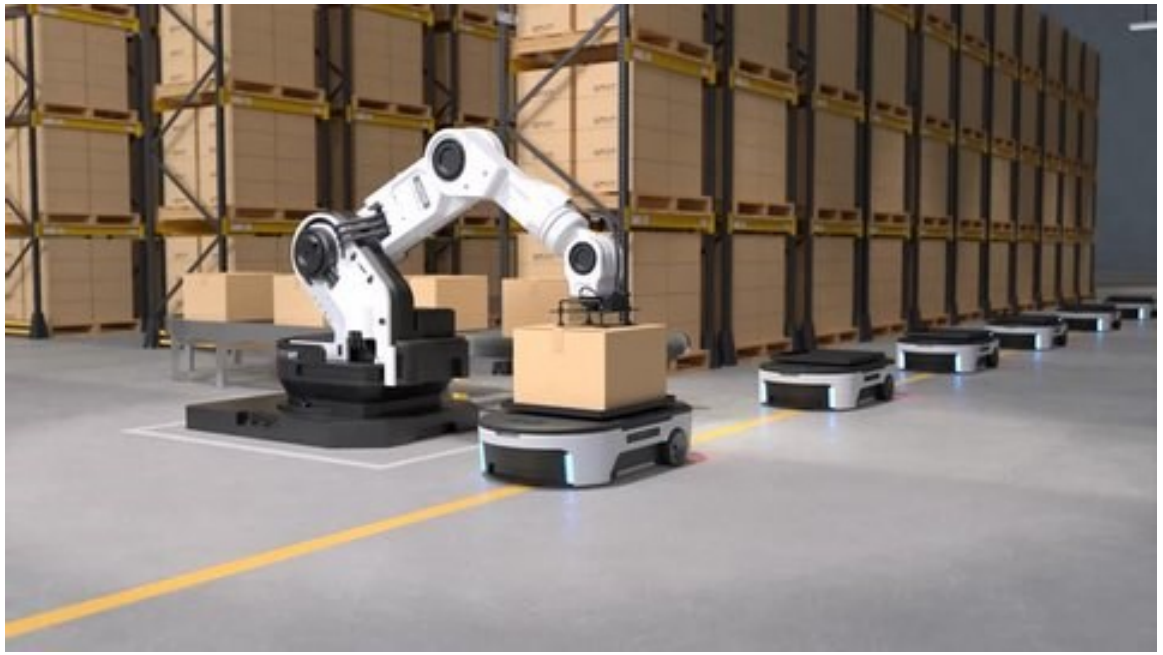


4. Action Module:

From Planning to Actions

- The **choices** and **plans established** by the **planning module** are **implemented** by the **action module**.
- **Converts plans** into **executable commands** and **serves** as the **agent's interface** with the **outside world**.
- **Operates** in **conjunction** with **actuators** or output devices, **carrying out** and **overseeing** the **selected course** of **action**.
- **Feedback** from the action module helps the **planning module** make **necessary modifications** and **fixes**.

Example: When a **robotic arm** is used in a **factory**, the action module gets instructions on how to pick up and move things. It then sends signals to the robot's motors to carry out the exact motions required to ensure precision assembly jobs.

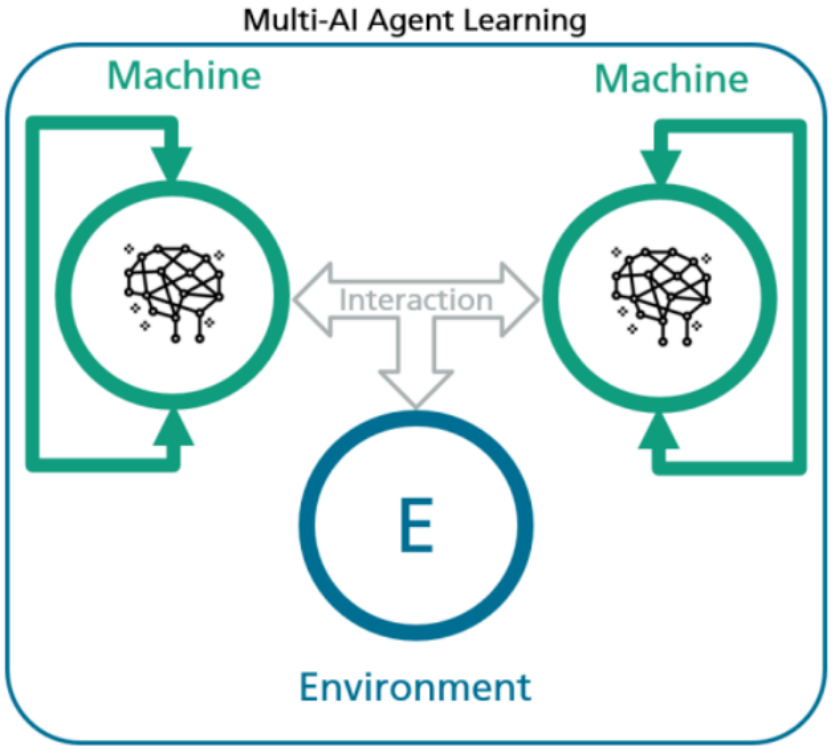


4. Learning Strategies:

Engine for Adaptation

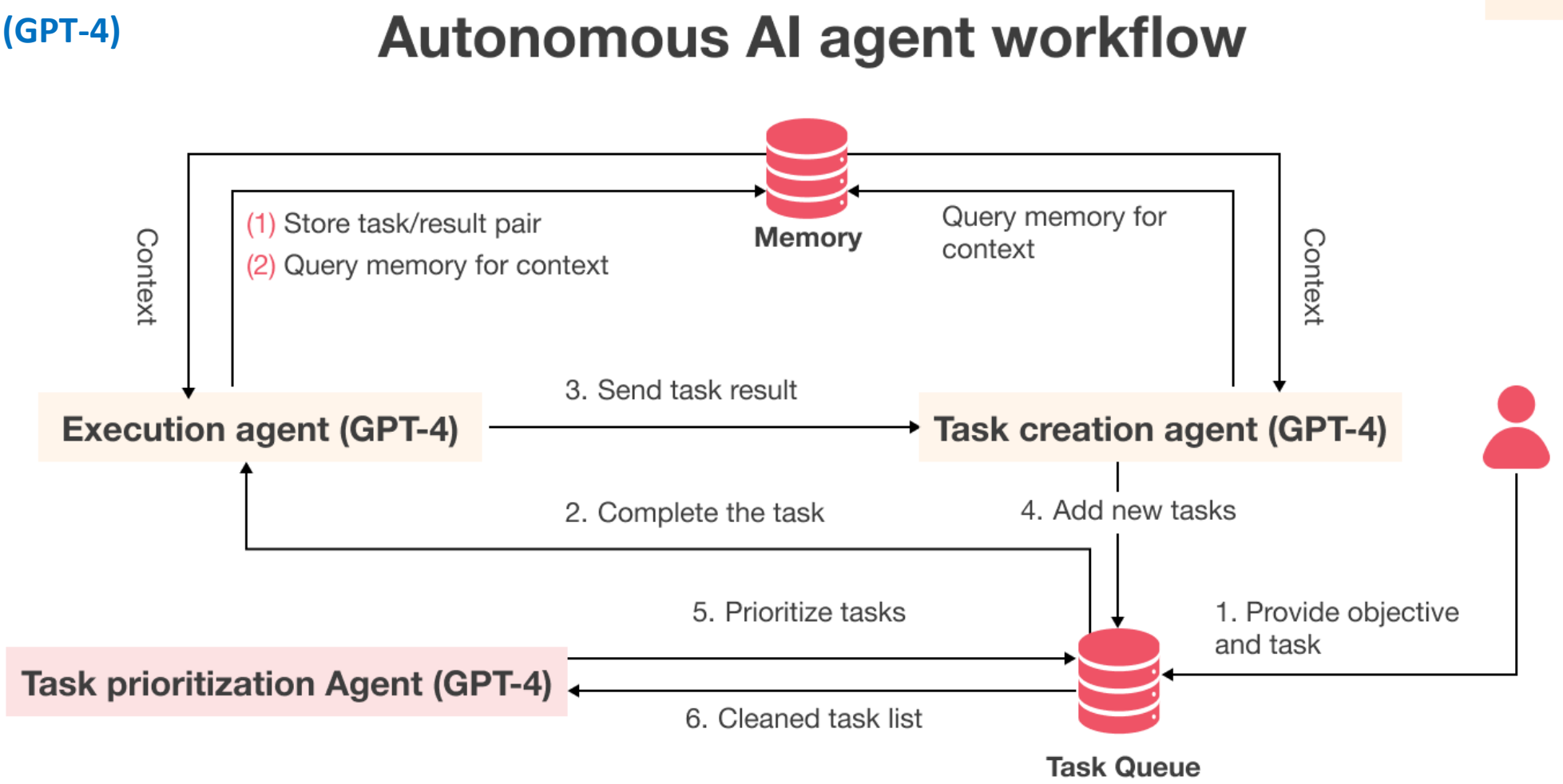
- The incorporation of **learning strategies**—**mechanisms** that **let agents change**, **become better**, and **learn new things**—is essential to agent architecture.
- Common methods include ***supervised learning***, ***unsupervised learning***, and ***reinforcement learning***.
- ***Reinforcement learning*** directs agents' behavior toward the best results by providing ***feedback in the form of incentives or punishments***.

Example: By ***improving the agent's capacity to draw conclusions from the past and adjust to new circumstances***, these learning techniques help them improve their performance over time.



A Sample Agent Architecture – Use Case

Generative Pre-trained Transformer 4 (GPT-4)

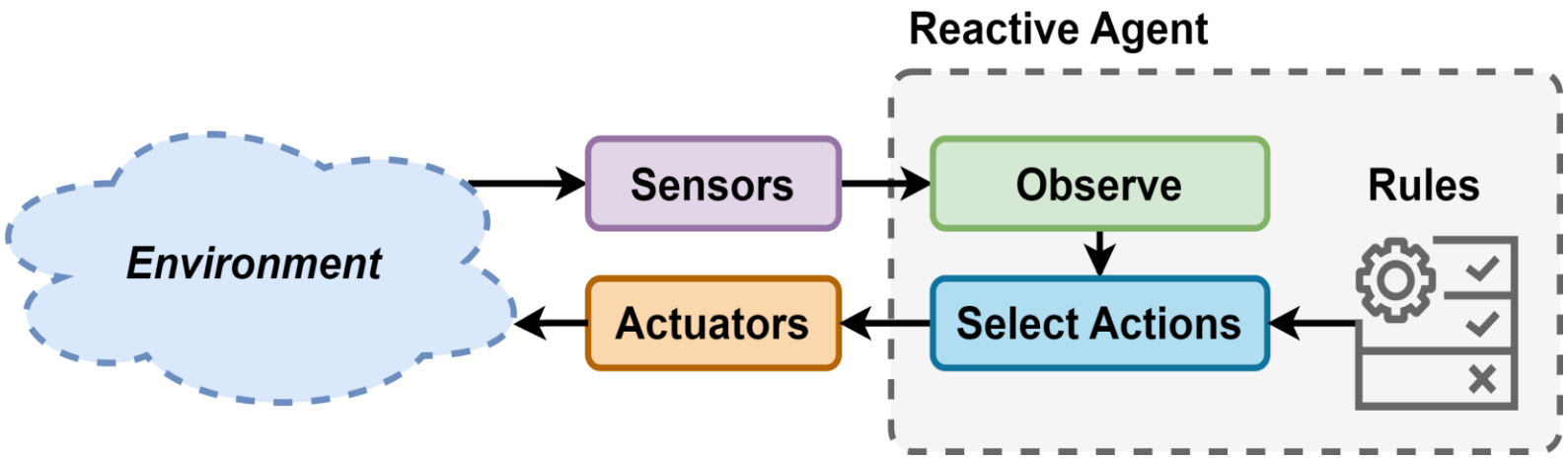


SIMFORM

Agent Architecture – Types

Reactive Architectures

- Reactive agents are the **simplest type** of **AI agent**.
- They are **solely focused** on the **immediate task** at hand and **do not take into account any long-term goals**.
- They **respond to changes** in the **environment** in a **stimulus-response manner**.

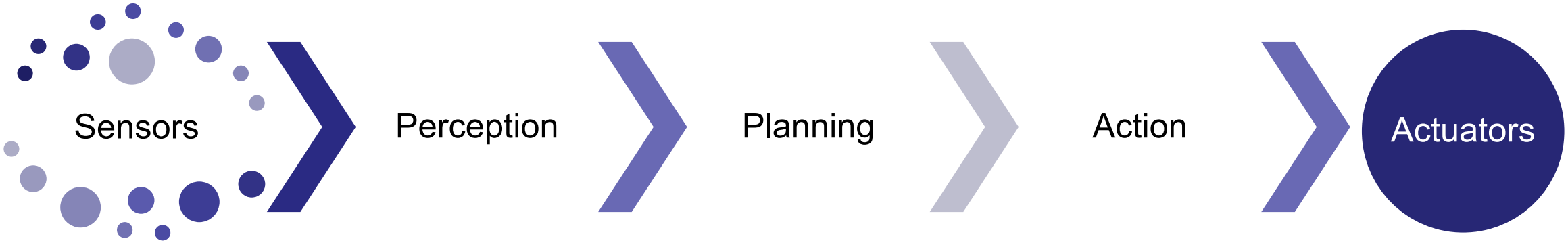


- Such architecture is based on the **direct mapping of situation to action**, and it is **different** from the **logic-based architecture**, where *no central symbolic* world *model* and *complex symbolic reasoning* are used.

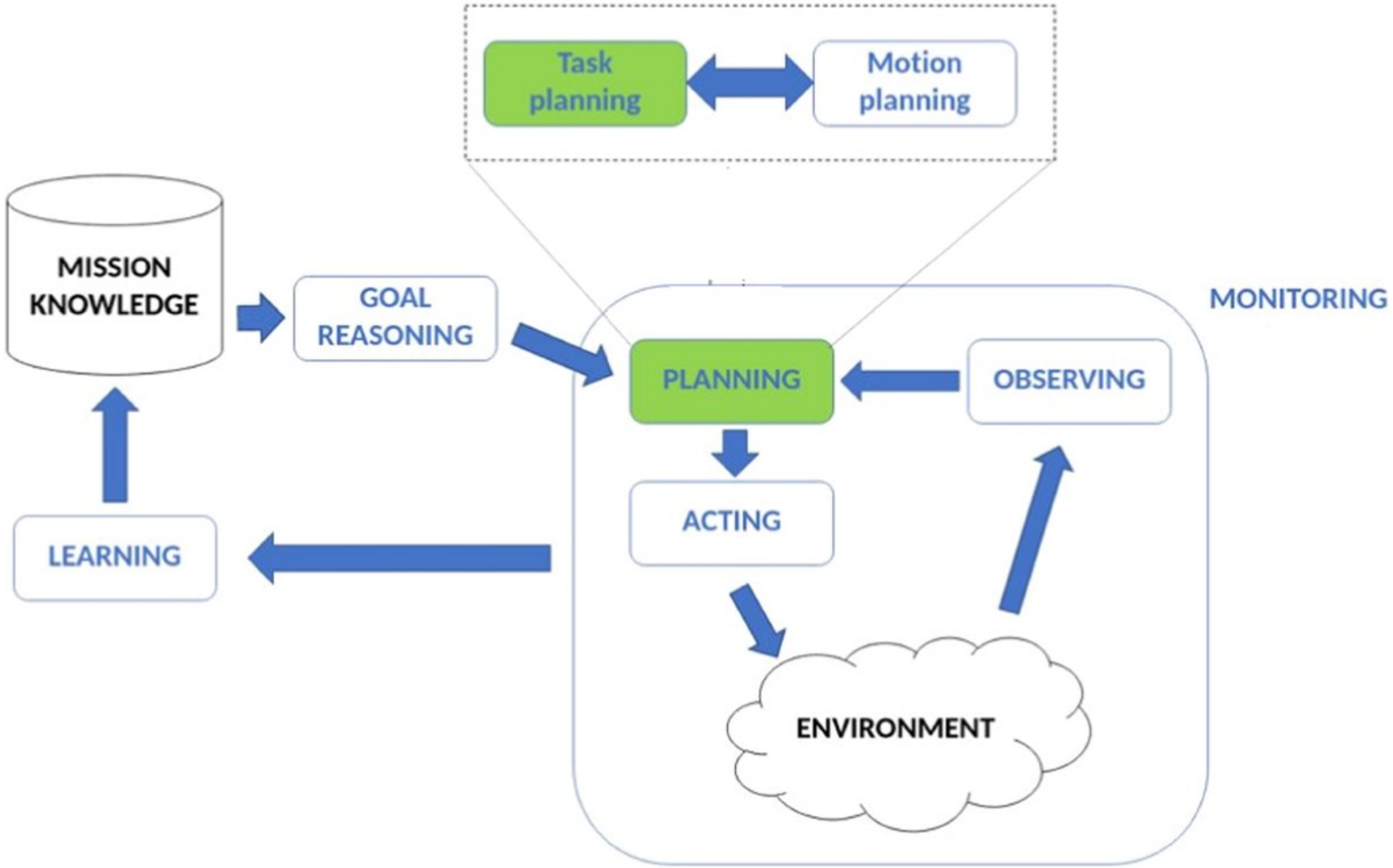
Agent Architecture – Types

Deliberative Reasoning Architectures

- Also known as intentional agents, **possess** an **explicitly represented**, **symbolic model of the world**, and **make decisions via symbolic reasoning**.
- Require **symbolic representation** with **compositional semantics** in all major functions, as their **deliberation is not limited to present facts**, but **construes hypotheses about possible future states** and **potentially** also **holds information** about **memory**.



Deliberative Reasoning Architectures – Use Case

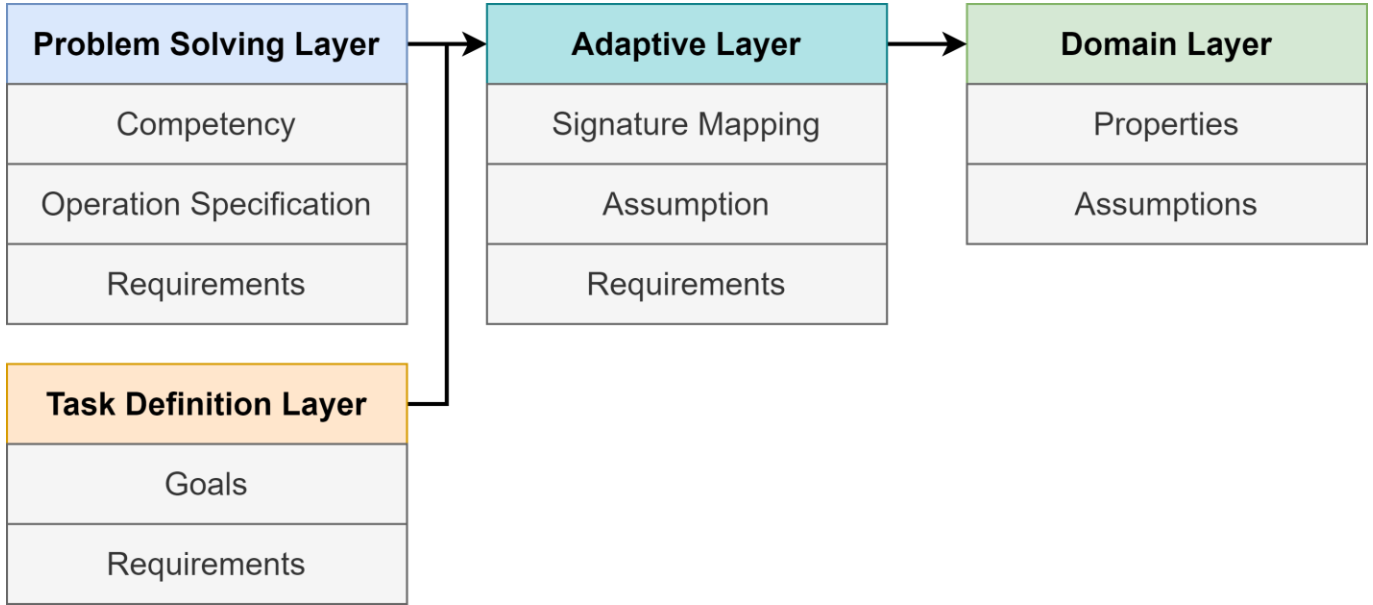


Logic programming for deliberative robotic task planning

Agent Architecture – Types

Layered/Hybrid Architectures

- Hybrid agents are the **most complex type of AI agent.**
- They take into **account both short-term and long-term goals** and **objectives** and **plan accordingly**.
- In such an architecture, an **agent's control subsystems** are arranged into a **hierarchy:**
  - With **higher layers dealing** with **information** at **increasing levels** of **abstraction**.
  - Often, the ***reactive component*** is given some kind of ***precedence*** over the ***deliberative*** one.



Agent Architecture – Types

- **Architecture** of an **agent** is **closely tied to its environment** and the **tasks it is designed to perform**. It **includes both** the **software elements** (the agent program) and the **hardware elements** (sensors and actuators) that allow the agent to interact with its environment.

Reactive Architectures

- Are **straightforward** to **design** and **implement**, offering **rapid responses to environmental stimuli without complex processing**.
- However, they **lack the capability to understand broader context** or perform complex tasks.

Deliberative architectures

- Enable **agents** to **generate optimal solutions** and **reason** about **future states**,
- But they **struggle** with **rapid changes** in the **environment** due to slower re-planning.

Hybrid architectures

- Merge the **responsiveness** of **reactive systems** with the **foresight** of **deliberative ones.**
- Yet their **complexity** can **pose significant design** and **implementation challenges**.

Conclusions

- **Introduction to AI**:
  - Artificial Intelligence (AI) is the branch of computer science focused on ***creating systems capable of performing tasks that would typically require human intelligence***. These tasks include **reasoning**, **learning**, **problem-solving**, **perception**, and **language understanding**.
  - Techniques used in AI include ***machine learning***, ***natural language processing***, ***computer vision***, ***robotics***, and ***expert systems***.
- **Agents and Rationality:**
  - An agent in AI is an **entity** that **perceives** its **environment** through sensors and **acts upon that environment** through actuators, with the **goal** of **achieving certain objectives**.
  - Rationality in the context of agents refers to the **principle of selecting actions that maximize the expected performance measure**, based on the **agent's knowledge** and **observations** of the **environment**.

Conclusions

- **Task Environment of AI**:
  - Refers to the **specific problem** or **domain** in which the **AI agent operates**. It **includes everything** that the **agent perceives** and **acts upon**.
  - **Components of a task environment include:**
    - ***Performance measure***: Defines the criterion for success or failure of the agent's actions.
    - ***Actuators***: Mechanisms through which the agent can affect its environment.
    - ***Sensors***: Devices through which the agent perceives its environment.
    - ***Environment***: The external context within which the agent operates, including other agents and objects.
  - **Task environments** can **vary widely**, from ***simple deterministic environments*** to ***complex***, ***uncertain***, and ***dynamic*** ones.

Conclusions

- **Agent Architecture Types:**
  - An **autonomous AI agent** is an **advanced software program** that can **operate independently** without **human control**.
    - *It can think, act, and learn on its own, without needing constant input from humans.*
    - *AI agents can reach higher levels of autonomy and intelligence by combining sensory skills, knowledge storage, decision-making abilities, action execution, and learning mechanisms.*
  - **AI agent architectures** offer a **organized system** for comprehending how **smart systems independently perceive**, **reason**, and **make decisions** in their surroundings.
    - *The selection of an architecture is driven by the application's needs, including real-time response requirements, task complexity, environmental dynamics, and desired autonomy levels.*