Artificial Intelligence (CS280)

Noushin Saba

About the Course

- Explores **fundamental theoretical concepts** of Artificial Intelligence (AI)
- Focuses on **foundational principles** rather than practical application
- Examines **symbolic manipulations, pattern matching, and knowledge representation**
- Highlights the distinction between **knowledge, data, and code** in an AI context
- Provides a **conceptual understanding** of AI methodologies and their role in intelligent systems

Course Objectives

- **Analyze and discuss** the theoretical foundations of AI, including core concepts like problem-solving by searching and knowledge representation.
- **Differentiate and conceptualize** the distinctions between knowledge, data, and code within the context of **AI systems**.
- **Evaluate and critique** various AI techniques and algorithms, such as **min-max algorithms, neural networks**, and **natural language processing**, purely from a theoretical and analytical perspective.

Textbook/ Reference Book

- TEXTBOOK
- Stuart Russell and Peter Norvig, Artificial Intelligence. A Modern Approach, 3rd edition, Prentice Hall, Inc., 2015.

- REFERENCE BOOK:
- Norvig, P., “Paradigms of Artificial Intelligence Programming: Case studies in Common Lisp”, Morgan Kaufman Publishers, Inc., 1992.
- Luger, G.F. and Stubblefield, W.A., “AI algorithms, data structures, and idioms in Prolog, Lisp, and Java”, Pearson Addison-Wesley. 2009.
- Severance, C.R., 2016. “Python for everybody: Exploring data using Python 3.” CreateSpace Independent Publ Platform.
- Miller, B.N., Ranum, D.L. and Anderson, J., 2019. “Python programming in context.” Jones & Bartlett Pub.
- Joshi, P., 2017. “Artificial intelligence with python.” Packt Publishing Ltd.

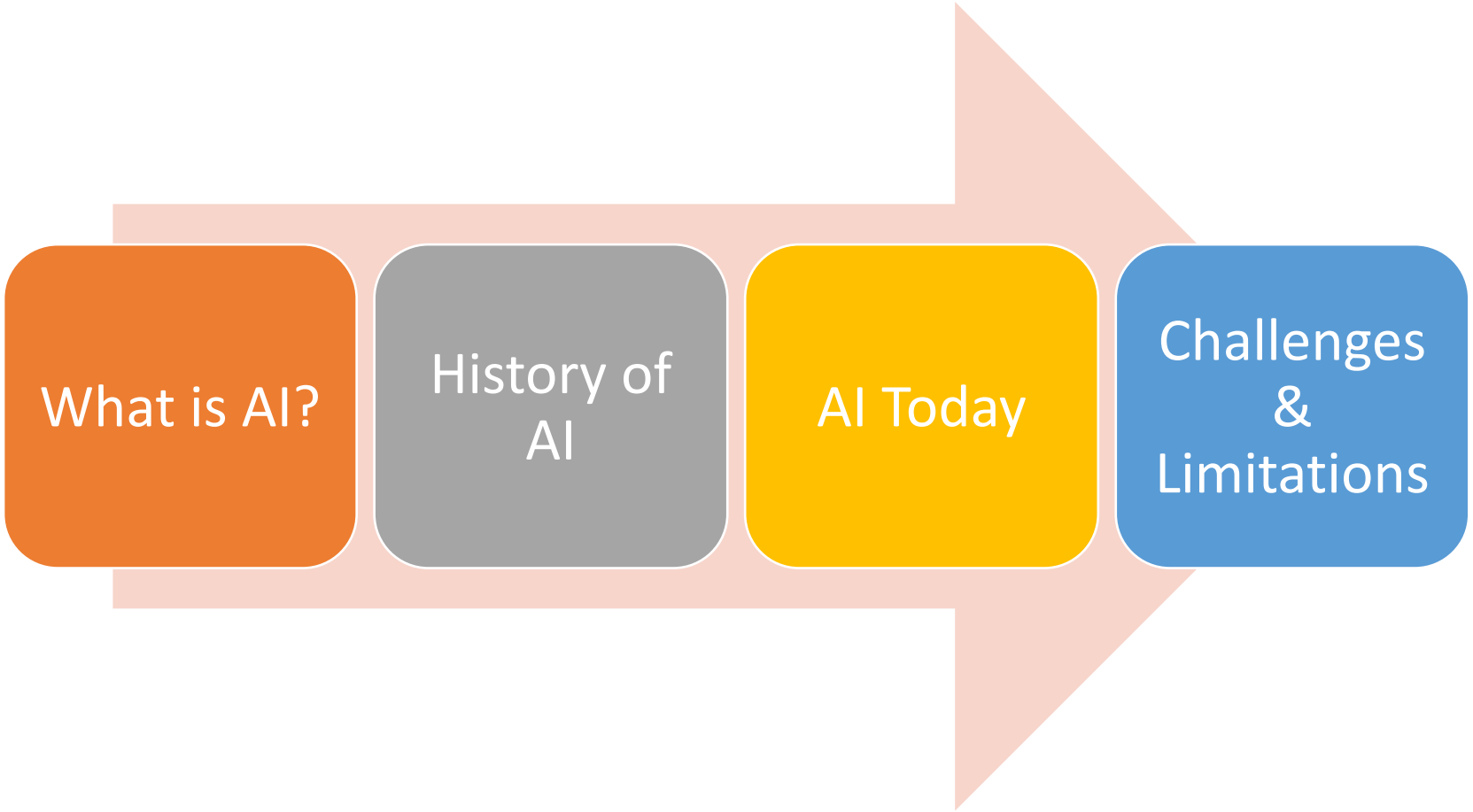
Marking Criteria

Type of Assessment	Total Number	Weightage (%)
Assignments	04	5
Quizzes	04	10
Project/Problem-Based Learning Activities PBLs	01 Project	10
Mid Term Exam	01	25
Final Term Exam	01	50

Lecture 1

Foundational Concepts: An Introduction to Artificial Intelligence and its Theoretical Applications

Outline



What is Artificial Intelligence (AI)

- John McCarthy, who coined the term Artificial Intelligence in 1956, defines it as

"The science and engineering of making intelligent machines", especially intelligent computer programs.

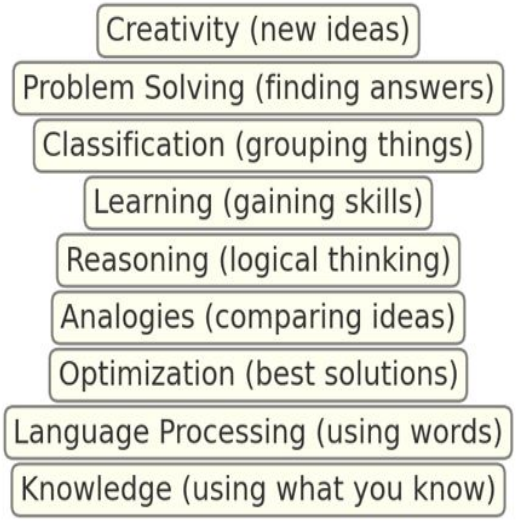
- AI is the study of how to make computers do things which, at the moment, people do better.
- AI is the study and design of ***Intelligent Agents***, where an intelligent agent is a system that perceives its environment and take actions that maximize its chances of success.
- AI is concerned with the design of ***intelligence*** in an ***artificial*** device.
- There are two ideas in the definition
  - Intelligence
  - Artificial
- The term artificial is easy to understand. But it's very difficult to define intelligence.

What is Intelligence?

- The ability to acquire and apply knowledge and skills.
  - Intelligence is what we use when we don't know what to do.
  - Intelligence relates to tasks involving higher mental processes.

Examples: Creativity, Solving problems, Classification, Learning, Reasoning, Building analogies, Optimization, Language processing, Knowledge and many more.

Intelligence = ability to learn new things and use knowledge to handle new situations



What is Intelligence Behaviour?

- Perceiving one's environment.
- Acting in complex environments.
- Learning and understanding from experience.
- Reasoning to solve problems and discover hidden knowledge.
- Knowledge applying successfully in new situations.
- Thinking abstractly, using analogies.
- Communicating with others, and more like
- Creativity, Ingenuity, Expressive-ness, Curiosity.

Goals of AI

Solve hard Problems for Human

- Automate tasks that are difficult for, time consuming, or impossible for humans
- **Examples**
  - Medical Diagnosis
  - Self Driving Cars
  - Language Translation
  - Scientific Discovery

Design Intelligent Agents

- AI is about creating agents that perceive, reason, and act
- **Framework of an Intelligent Agent**
  - **Perception:** Sense the environment
  - **Reasoning:** Process data, apply knowledge and make decisions
  - **Action:** Interact with environment to achieve goals

Key Characteristics of AI Systems

- **Autonomy**: Operate without human guidance
- **Adaptability**: Learn and improve from experience
- **Rationality:** Choose actions that maximize success

Types of AI (By Capability)

•Narrow AI (Weak AI)

- **Specialized** in a **single** task.
- Most AI today fall into this category
- **Examples**: Spam filters, Voice Assistants (Siri, Alexa)

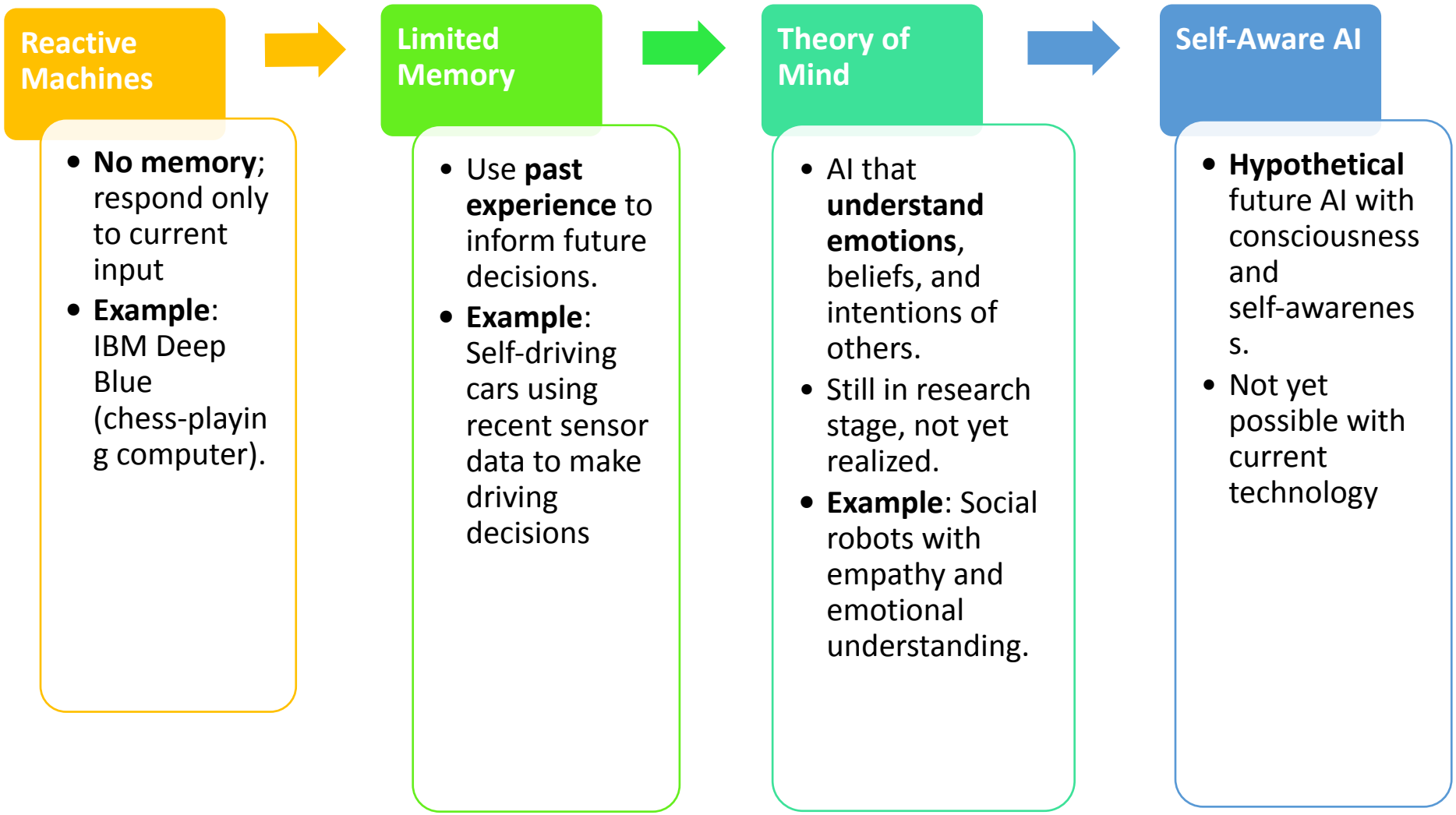
General Artificial Intelligence (Strong AI)

- **Human like intelligence** that can perform any intellectual task
- Still a **theoretical concept**, not yet achieved yet
- **Examples**: A machine that can learn physics, paint, write, and reason like a human

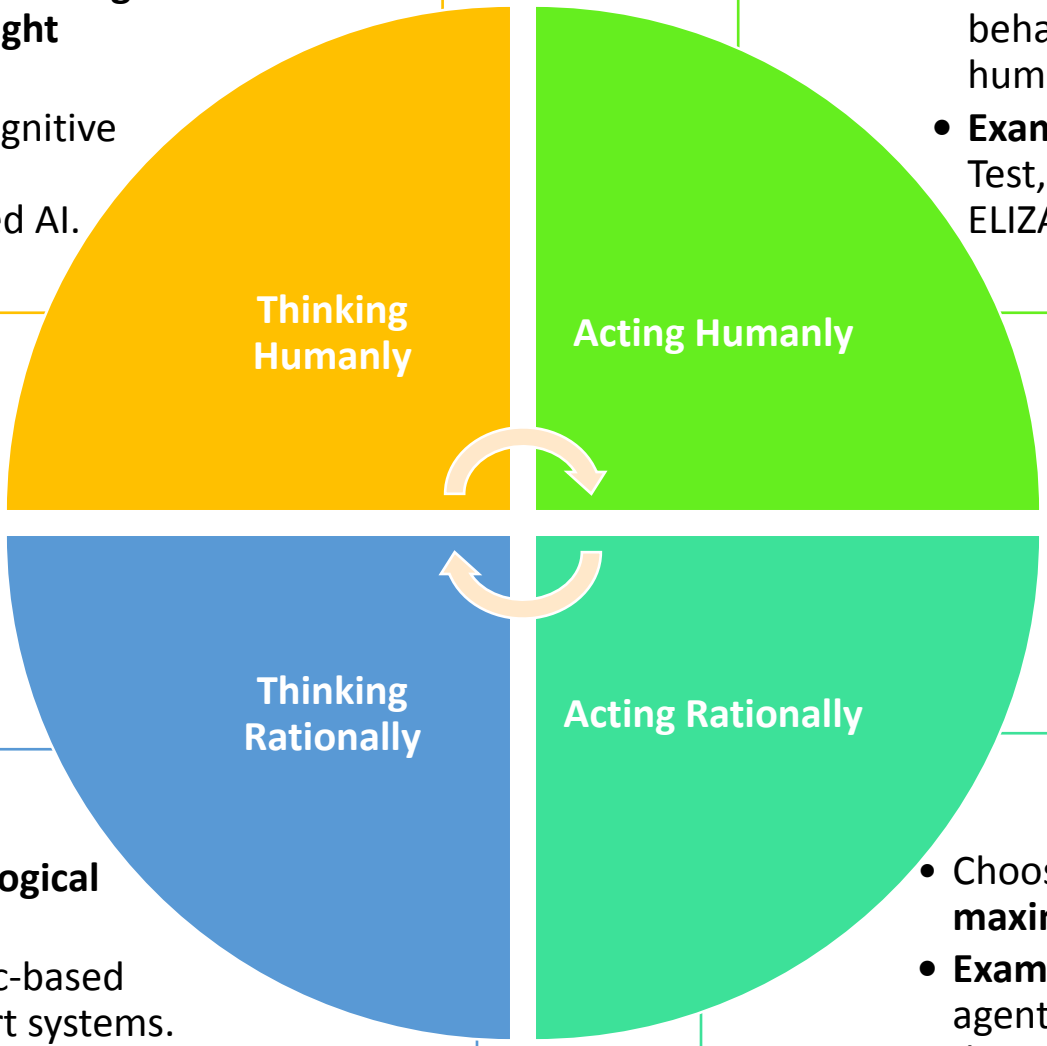
Superintelligence

- Intelligence that **surpasses** human cognitive abilities in all areas
- Theoretical and highly debated(associated with future AI speculation)
- **Example**: A hypothetical AI that can innovate new science, outsmart humans and self improve rapidly

Types of AI (By Functionality)



Approaches to Defining AI



- Focus on **simulating human thought processes**.
- **Example**: Cognitive modeling, brain-inspired AI.

- Make machines behave like humans.
- **Example**: Turing Test, chatbots like ELIZA.

- **Draw correct logical conclusions**.
- **Example**: Logic-based systems, expert systems.

- Choose **actions that maximize success**.
- **Example**: Intelligent agents, rational decision-making.

Definitions of AI: Four Approaches

Thinking Humanly

- "The exciting new effort to make computers think . . . *machines with minds*, in the full and literal sense." (Haugeland, 1985)
- "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)

Thinking Rationally

- "The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985)
- "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)

Acting Humanly

- "The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990)
- "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)

Acting Rationally

- "Computational Intelligence is the study of the design of intelligent agents." (Poole *et al.*, 1998)
- "AI . . . is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

Thinking humanly? | Acting humanly? | Thinking rationally? | Acting rationally?

The brain as an information processing machine.

- Requires **scientific theories** of how the brain works.

Note: The brain does not work like artificial neural networks from ML!

Understand cognition as a computational process

- **Introspection**: reflect on how we think.
- **Behavioral studies**: Predict and analyze human actions.
- **Brain imaging & neuroscience**: study brain activity and data

AI and Consciousness

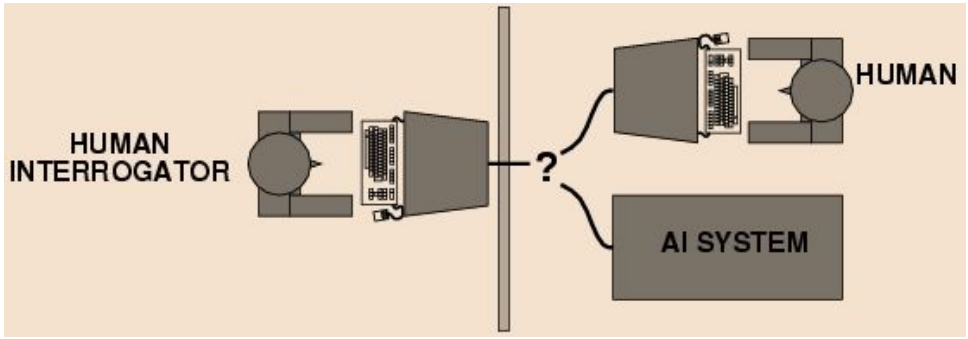
- What does it mean for a machine to be *conscious* or *sentient*?
- How could we recognize or test machine consciousness?

Cognitive Sciences

AI + psychology + neuroscience

Thinking humanly? | **Acting humanly?** | Thinking rationally? | Acting rationally?

- Alan Turing rejects the vague question "Can machines think?"
- The Turing Test tries to define what acting like a human means

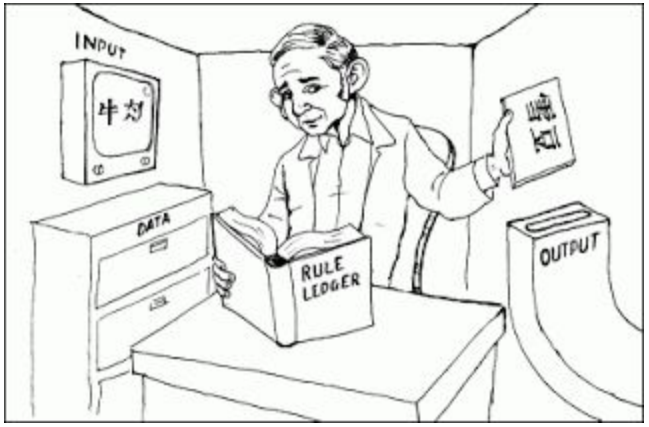


- What capabilities would a computer need to have to pass the Turing Test? These are still the core AI areas.
  - Natural language processing □ understand & generate language
  - Knowledge representation □ store and recall information
  - Automated reasoning □ use knowledge to answer and infer
  - Machine learning □ adapt and detect patterns
- **Prediction vs. Reality**
  - Turing predicted: *By 2000, machines could fool 30% of judges for 5 minutes.*
  - **Today:** Systems like **ChatGPT (2023)** likely achieve this level or more.

Turing Test: Criticism

- **Problems with the Test**
  - Some **human behavior is not intelligent**.
  - Some **intelligent behavior is not human**.
  - Human judges may be **easily fooled**.
  - Results depend heavily on **expectations**.
  - **Anthropomorphic fallacy**: humans tend to humanize machines.
  - Risk of **imitating intelligence without true understanding** (e.g., ELIZA, 1964).
- **Is Passing the Test a Good Goal?**
  - **Engineering view:** imitating humans is not always practical.
  - AI can be **useful and intelligent** without behaving exactly like humans.
  - Focus should be on solving **real-world problems** and building **intelligent agents**.

Chinese Room Argument



Thought experiment by John Searle (1980): Imitate intelligence using rules.

Thinking humanly? | Acting humanly? | **Thinking rationally?** | Acting rationally?

- **Goal:** Draw sensible conclusions from facts, logic and data.
- **Logic**:
  - A chain of argument that always yield correct conclusions.
  - Example: "Socrates is a man; all men are mortal; therefore, Socrates is mortal."
- **Logic-based approach to AI**
  - Represent problems in **formal logic notation**.
  - Use **deduction procedures** to derive solutions.
- **Challenges:**
  - Hard to express **real-world problems** in strict logic.
  - **Computational complexity** – solutions may be intractable.
  - Much rational behavior in an **uncertain world** cannot be defined by simple rules.

Thinking humanly? | Acting humanly? | Thinking rationally? | **Acting rationally?**

Acting rational means to try to achieve the "best" outcome.

- **Core Idea:** Rational action = choose the **best outcome**.
  - Use **optimization** to evaluate possible actions.
  - Desirability of outcomes measured by **utility** (economic concept).
  - Under uncertainty → maximize **expected utility**.
- Optimization has several advantages:
  - **Generality**: optimization is not limited to rules.
  - **Practicality**: can be adapted to many real-world problems.
  - **Well established**: solvers, simulation and experimentation.
  - Avoids philosophy and psychology in favor of a **clearly defined objective**.
- **Limitation: Bounded rationality**
  - In practice, rationality is limited by:
  - **Knowledge constraints** (incomplete information).
  - **Computational limits** (time, resources).
  - Agents approximate rationality instead of achieving perfection.

The 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.

1974-1980
First AI Winter

1987-1993
Second AI Winter

1980s 1989: Universal approximation theorem for neural networks.

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

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

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

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

2010

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).

2015

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

2017 Transformer architecture and large language models LLMs

2022 Generative AI models:

- DALL-E
- ChatGPT, Bard
- …

NVIDIA

Deep Learning Revolution (learning layered artificial neural networks) starts fueled by NVIDIA GPUs. enables leaps in image processing and speech recognition.

DeepMind
Now Google

OpenAI

Source: https://qbi.uq.edu.au/brain/intelligent-machines/history-artificial-intelligence

Rational Agent and Intelligent Agent

The Rational Agent Framework

- **Agent Definition**
  - perceives environment → reasons → acts.
- **Core Components**
  - Perception
  - Reasoning
  - Action
  - Learning
- **Example**
  - Self-driving car → sensors → decision-making → steering/braking
- **Purpose**
  - A unifying model for AI .

Example: Self-Driving Car (Rational Agent)

Percept: People crossing the street

Action: Stop the car

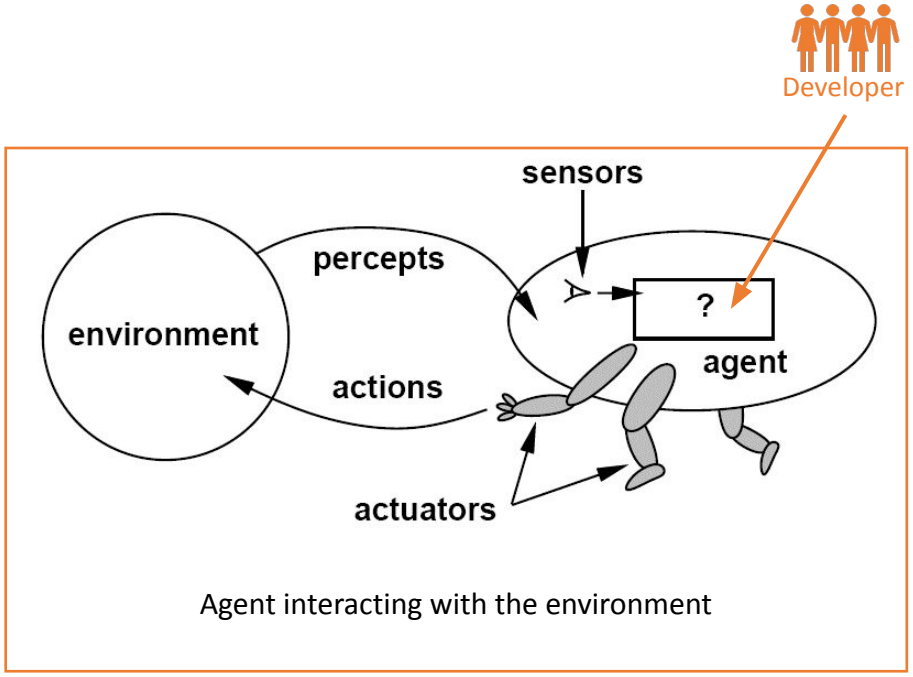
What are the Components of an Intelligent Agent?

Intelligent agents need to

- **Communicate** with the environment.
- **Represent knowledge**, **reason** and **plan** to achieve a desired outcome.

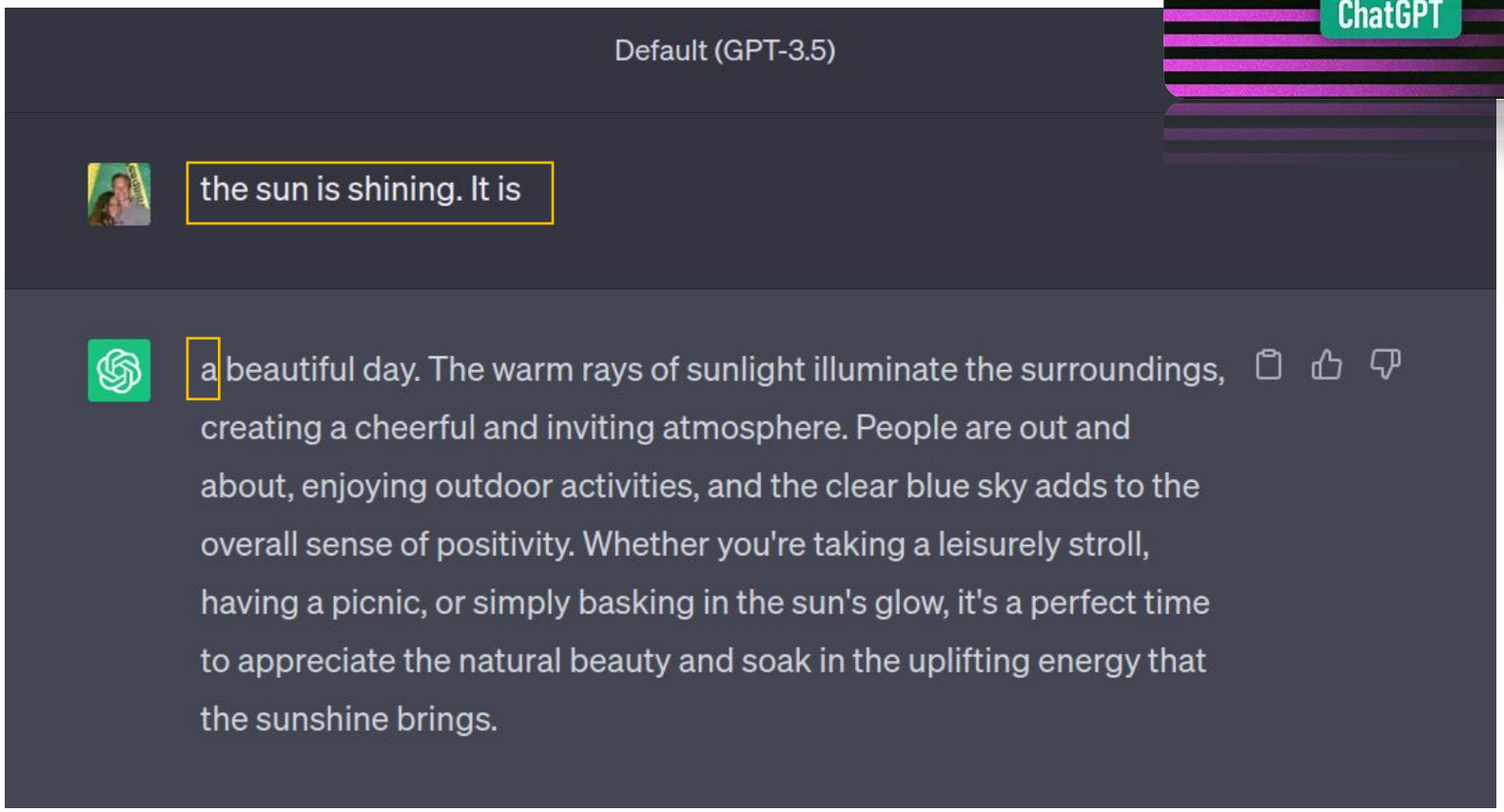
Optional

- **Learn** to improve performance**.**



Agent interacting with the environment

Example: LLM (Intelligent Agent)



Percept: prompt

Action: next most likely word

More words are created word-by-word.

Intelligent Agent vs. Rational Agent

Intelligent Agent	Rational Agent
Broad category: any system that perceives and acts	A specific type of intelligent agent
Acts intelligently but not always optimally	Always aims for the best possible outcome
Example: Chatbot imitating conversation	Example: Self-driving car choosing safest route

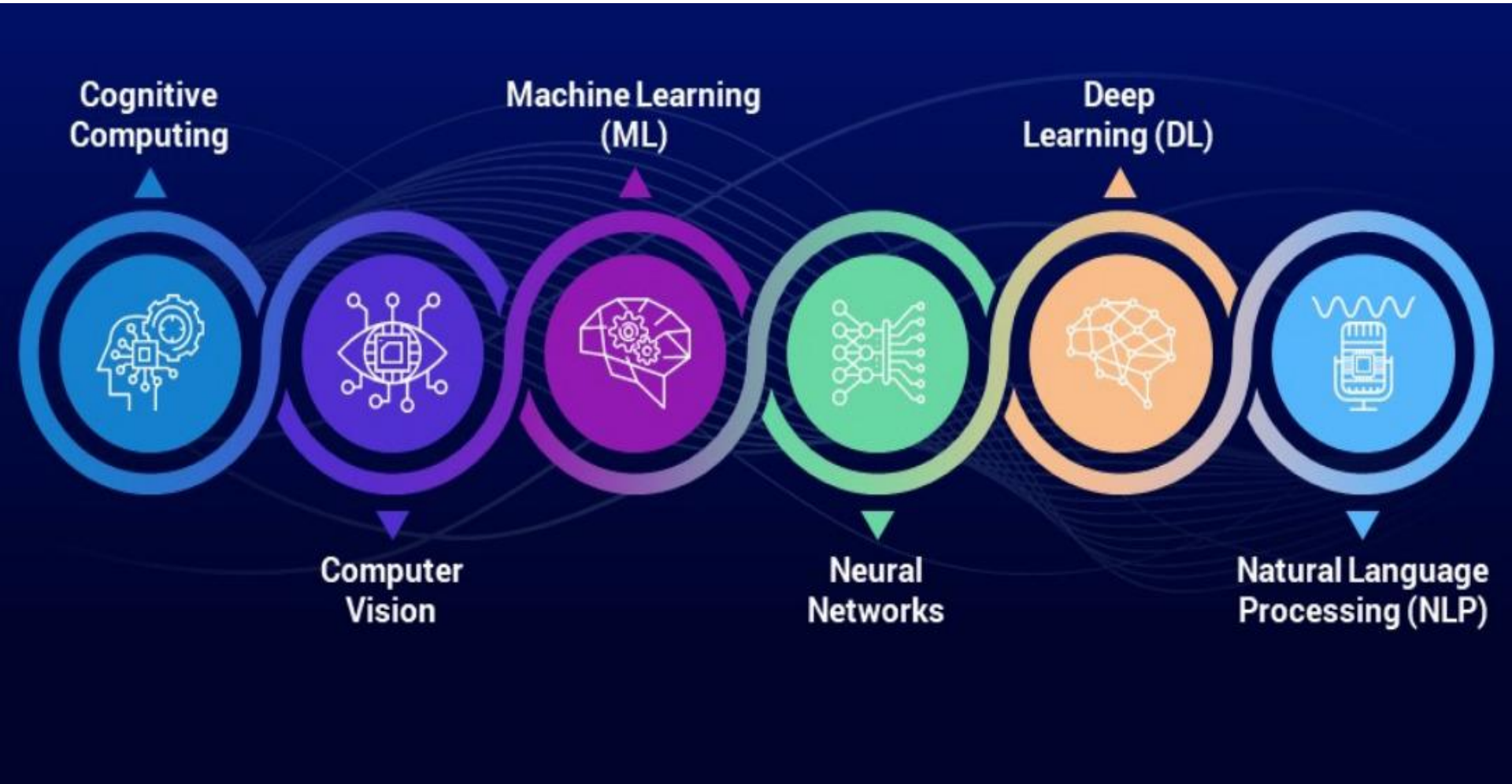
Important

- All rational agents are intelligent agents.
- Not all intelligent agents are rational.

Agent–Environment Interaction

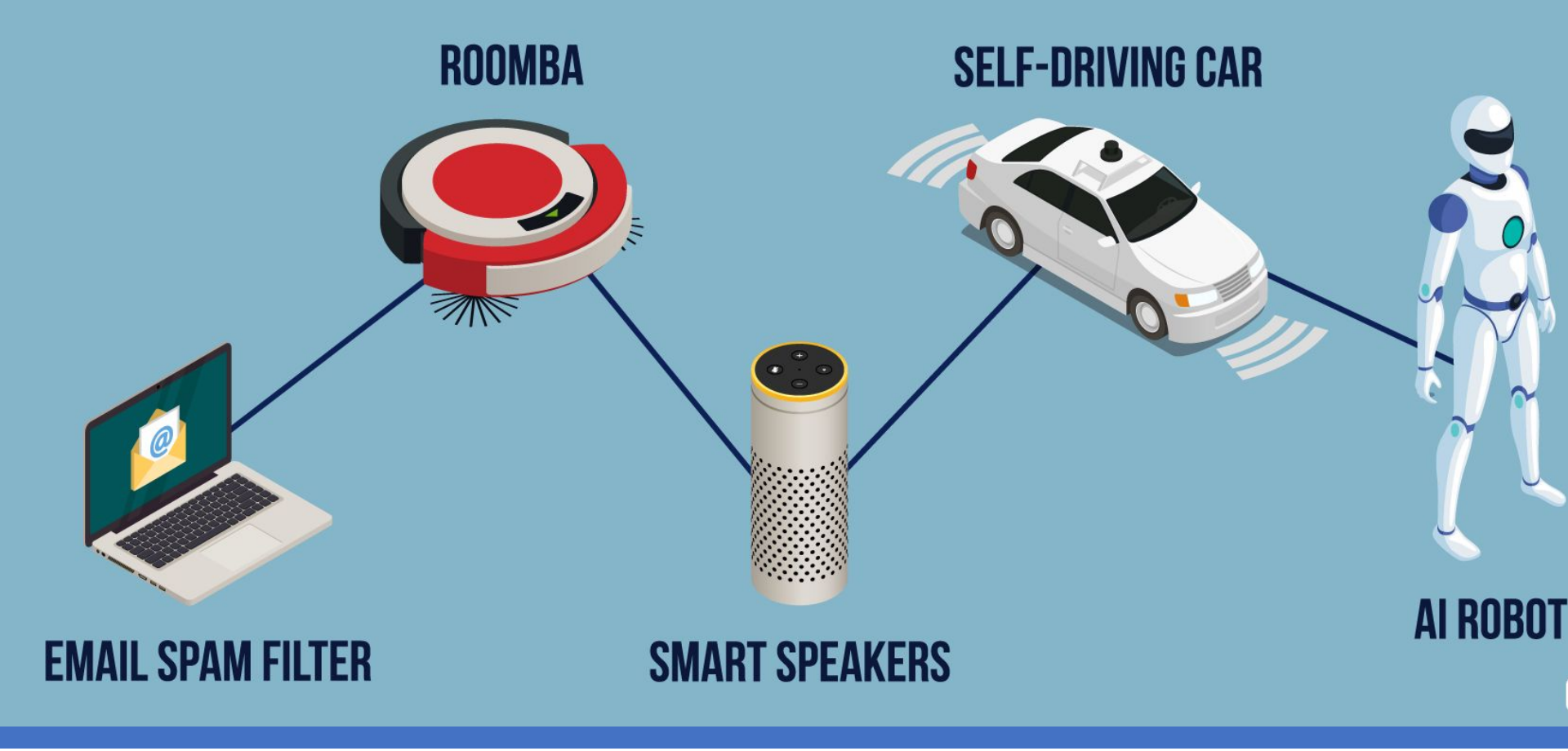
- **Agent**
  - Takes input from environment → chooses actions → affects environment.
- **Percepts**
  - Information sensed (e.g., camera, text input).
- **Actions**
  - Decisions or movements (e.g., answer, steering).
- **Feedback Loop**
  - Environment changes → agent perceives again → repeat.

Core Areas of AI



Core Areas of AI

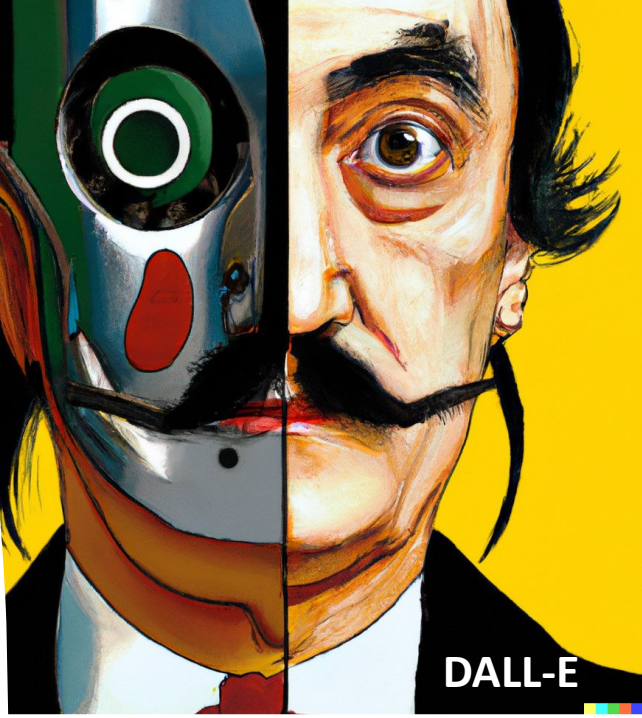
- **Perception**
  - Vision (image & video analysis).
  - Speech recognition.
- **Reasoning**
  - Logic and inference.
  - Planning & decision-making.
  - Knowledge representation.
- **Learning**
  - Machine learning (supervised/unsupervised).
  - Reinforcement learning.
- **Interaction**
  - Natural language processing (NLP)
  - Robotics & human-AI interaction.

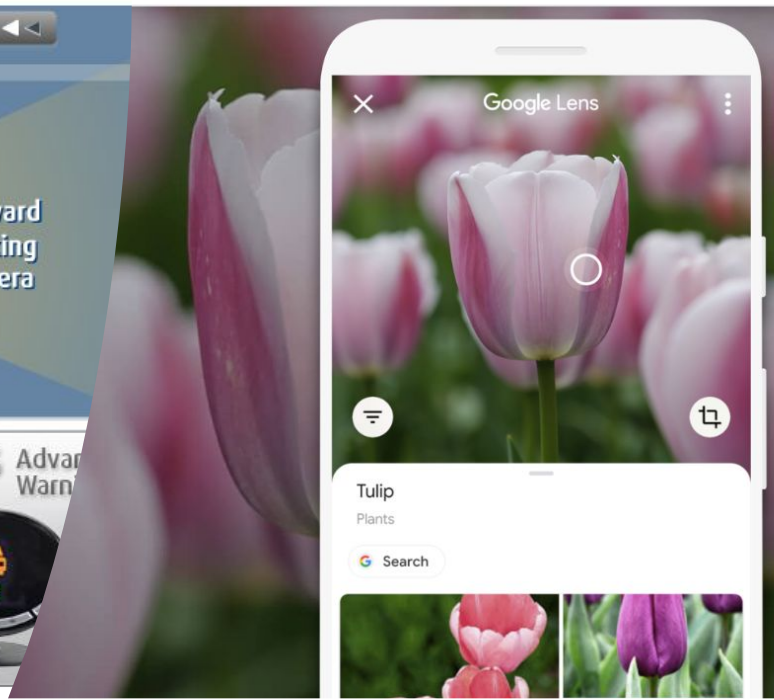


Applications of AI Today

Vision & Perception

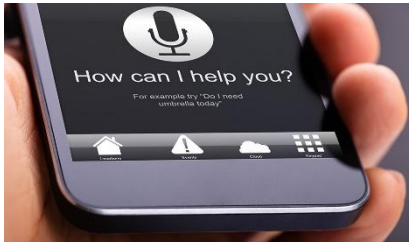
- **OCR** – Read license plates, scanned documents.
- **Face Detection** – Smartphones, security & surveillance.
- **Image Recognition** – Medical imaging, autonomous vehicles.
- **Vehicle Safety Systems** – Collision detection, driver assistance.
- **Visual Search** – Search products/images by picture.
- **Image Generation** – Generative AI (art, design, creativity).

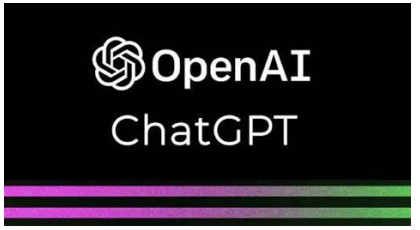


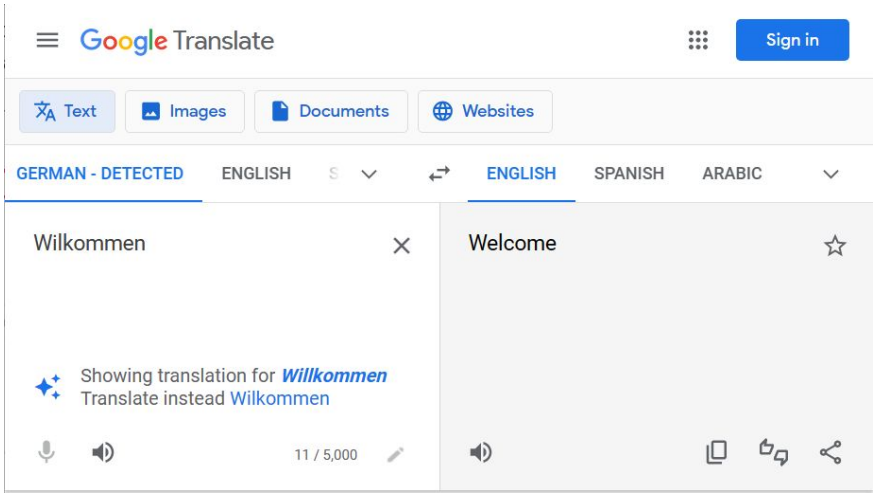


Natural Language Processing

- **Text-to-Speech (TTS)** – Generate natural-sounding voices.
- **Speech-to-Text (ASR)** – Detect and transcribe voice commands.
- **Machine Translation** – Translate between languages (Google Translate, DeepL).
- **Text Generation** – Q/A systems, chatbots, Large Language Models (LLMs).





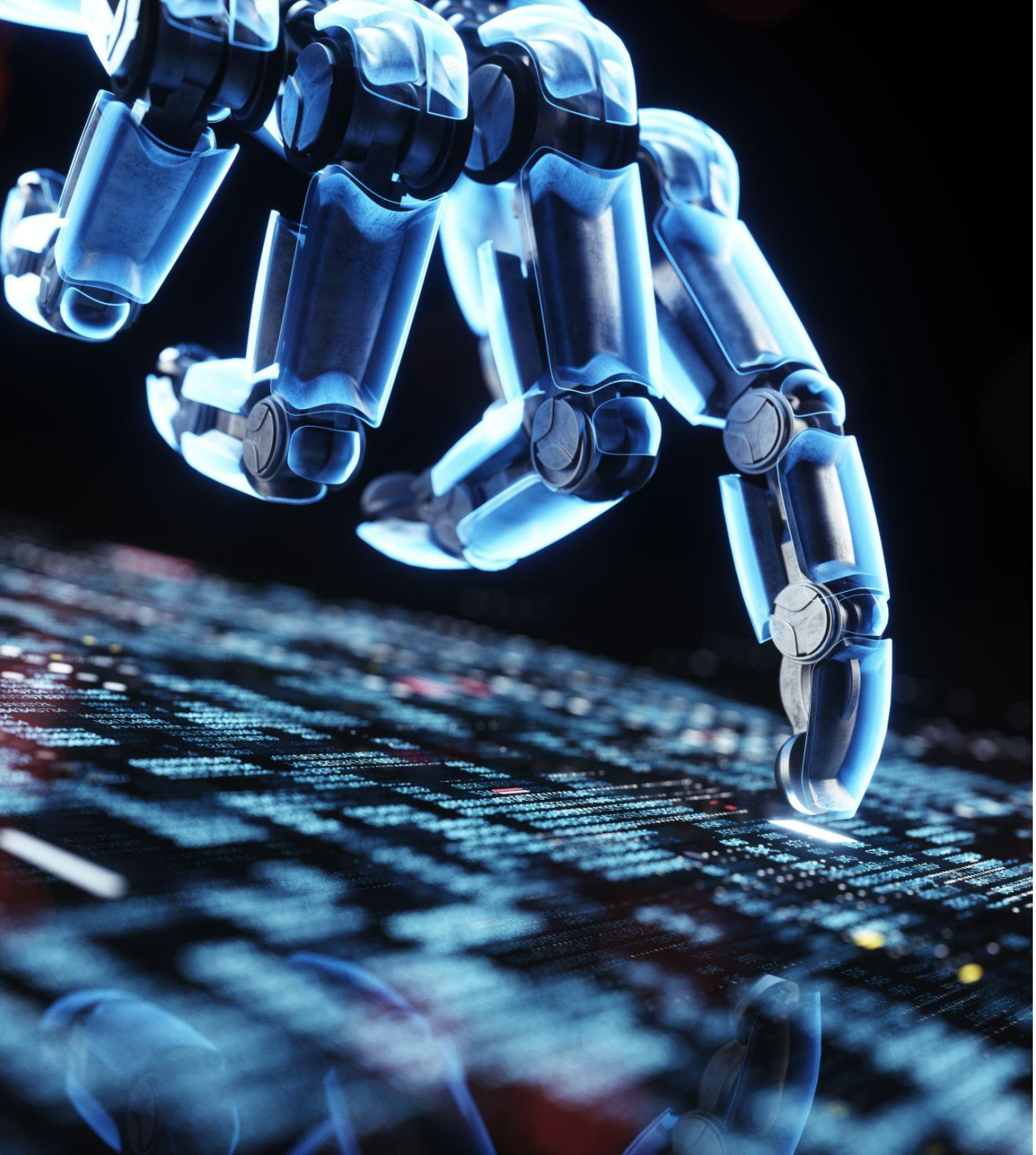


Robotics

- **Exploration** – Mars rovers.
- **Autonomous Vehicles** – Self-driving cars (Google, Tesla), DARPA Grand Challenge.
- **Aerial Robotics** – Autonomous helicopters and drones.
- **Sports & Competitions** – Robot soccer, RoboCup.
- **Personal Robotics** – Service robots, assistants, humanoid robots.
- **Robotic Companions** – Robotic pets, home assistants.

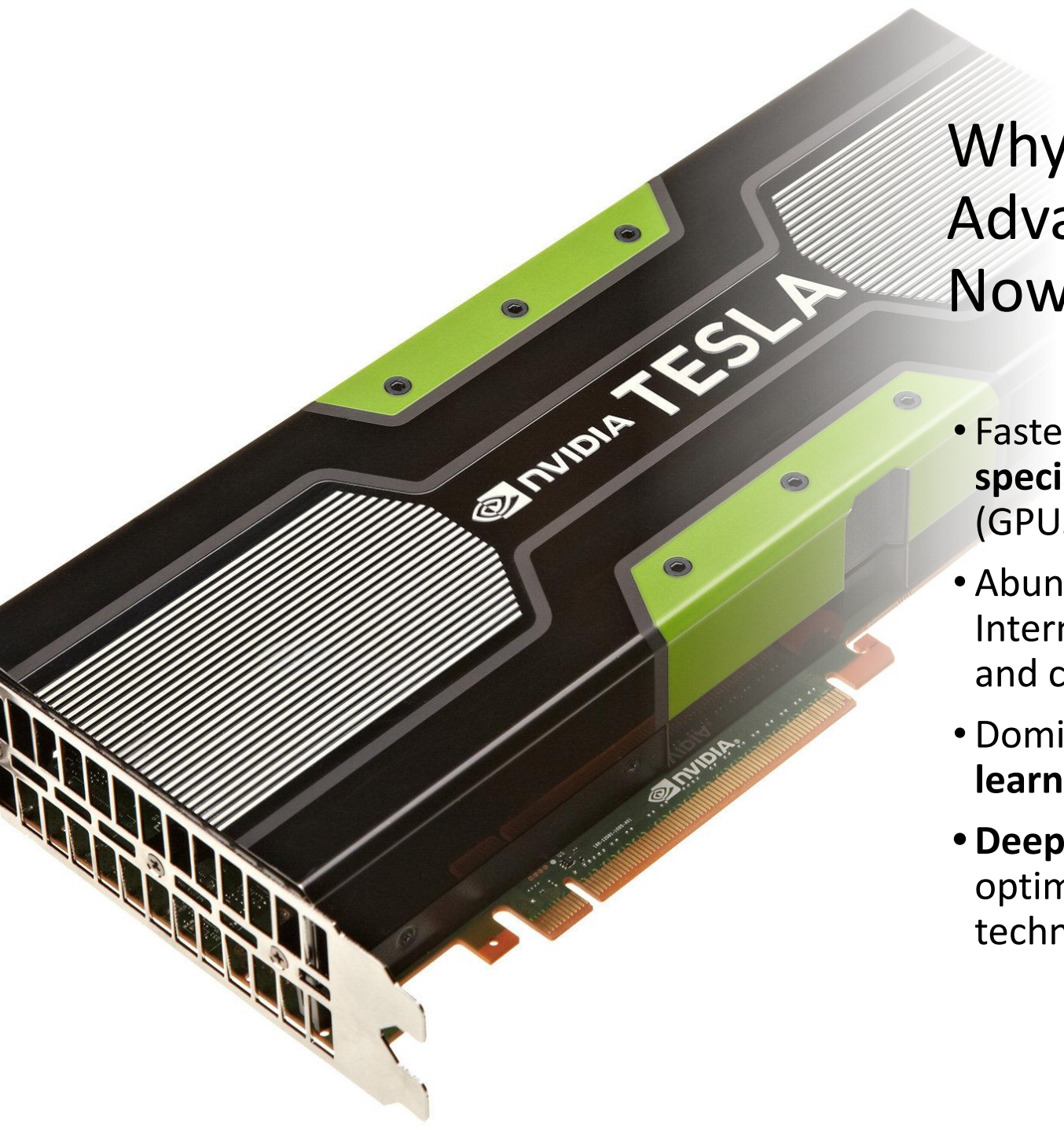
Industry

- **Healthcare**
  - Medical diagnosis & decision support.
  - Drug discovery & personalized medicine.
- **Finance**
  - Fraud detection & risk analysis.
  - Algorithmic trading & credit scoring.
- **Education**
  - Personalized learning platforms.
  - Intelligent tutoring systems.



Challenges and Limitations

- **Bounded Rationality** limits of knowledge & computation.
- **Moravec's Paradox** easy for humans, hard for AI.
- **The AI Effect** once AI works, it's no longer "intelligence."
- **Ethics & Safety** bias, misuse, regulation.



Why is AI Advancing Now?

- Faster computers and **specialized hardware** (GPUs).
- Abundance of **data** (the Internet, text, sensors) and cloud storage
- Dominance of **machine learning** methods
- **Deep learning** and new optimization techniques.

“Moravec’s Paradox”

Hans Moravec (1988)

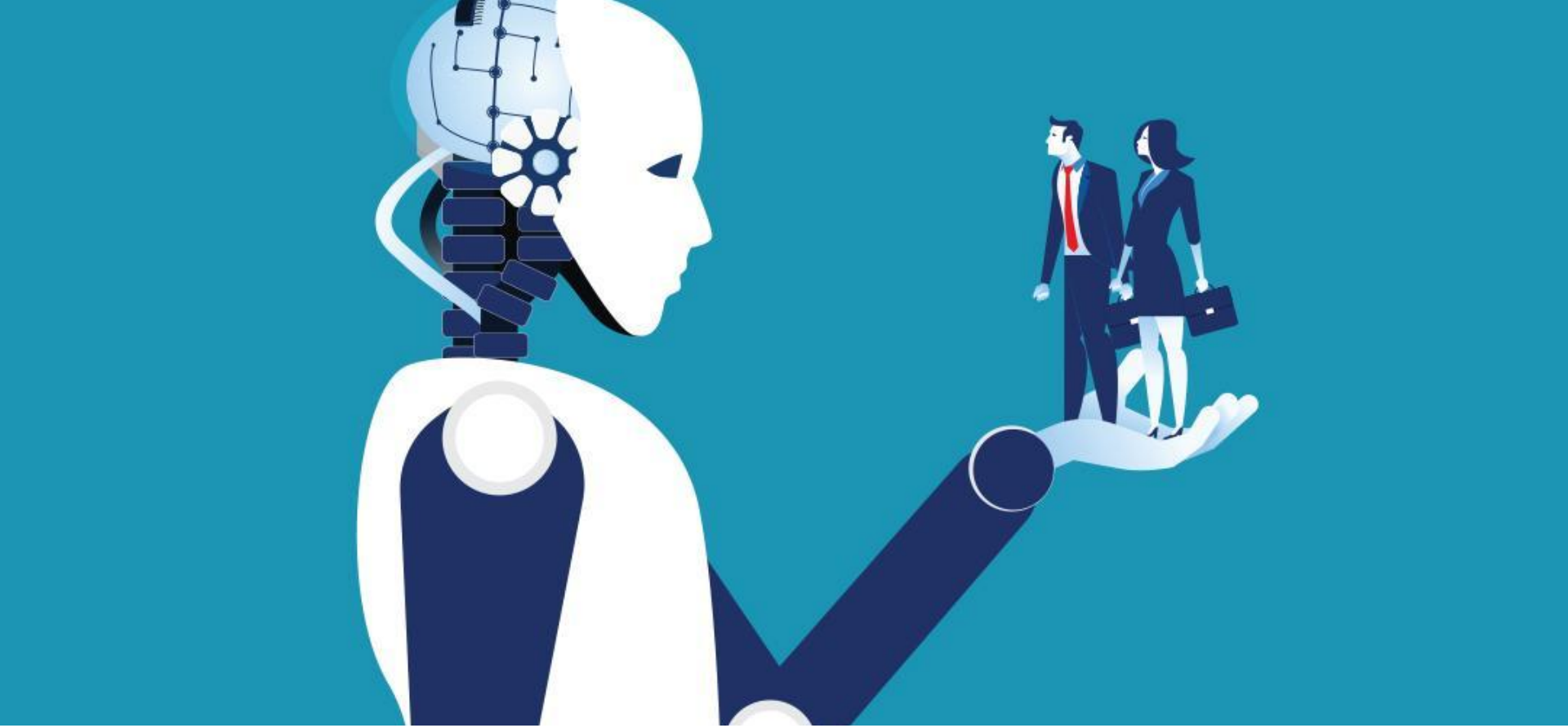
“It is easy to make computers exhibit adult-level performance in logic or games, but hard to give them the skills of a one-year-old in perception and mobility.”

- **Example**
  - Teenager → learns driving in a few hours.
  - AI → decades of work, still no *fully* autonomous cars.

https://www.newsweek.com/googles-new-two-legged-robot-future-warfare-429831

The AI Effect

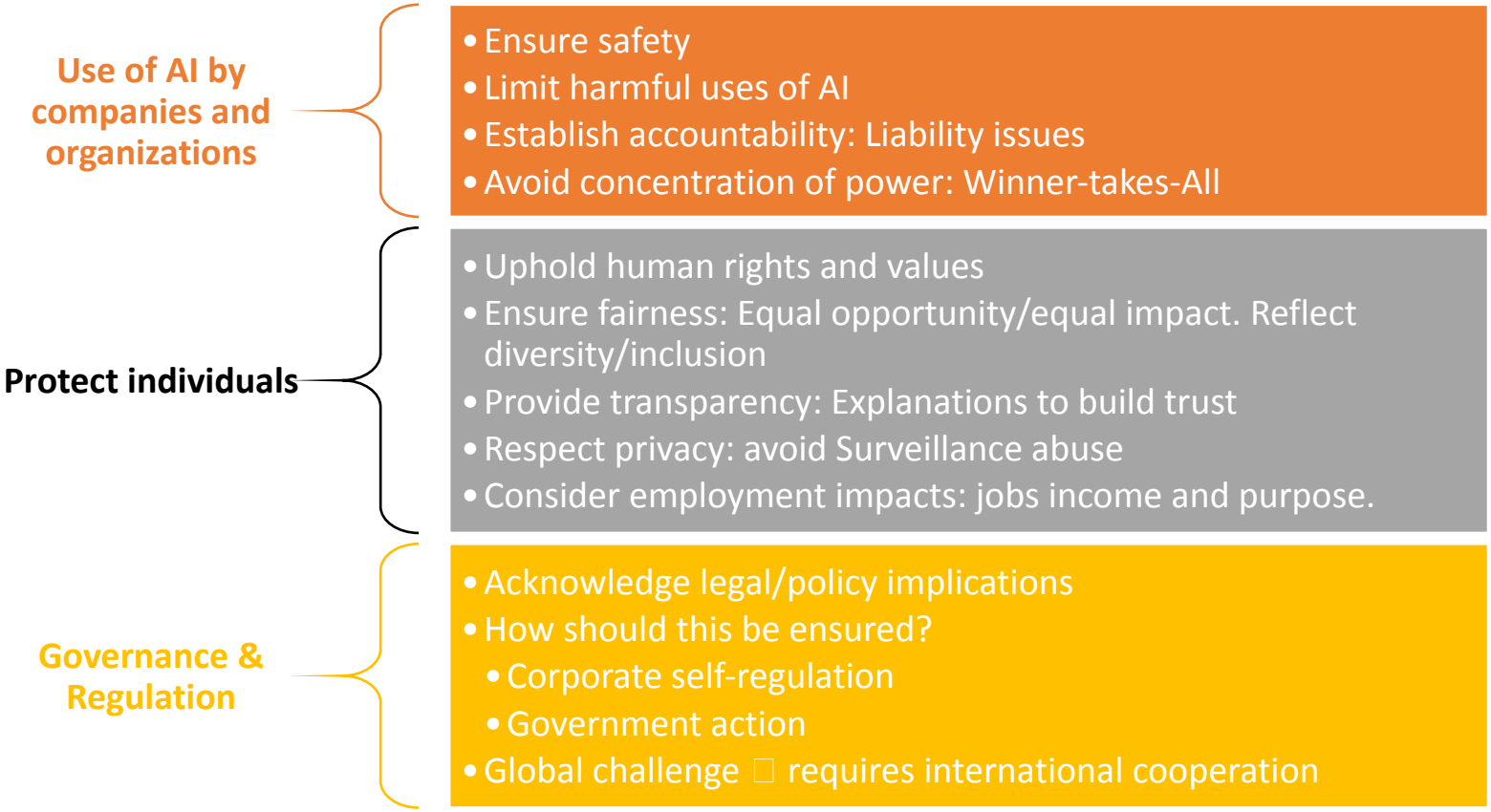
- Once a task is solved, it's no longer considered "real AI."
- **Examples**
  - **Calculators** □ arithmetic no longer "intelligence."
  - **Chess** □ computers now super-human.
  - **Machine learning** □ once "uniquely human," now routine.
  - Even **art** can be generated by AI.



AI Ethics & Safety

A new Frontier for Fairness and Freedom

AI Ethics & Principles



AI Safety in Practice

Goal: “Prevent accidents, misuse, or other harmful consequences of AI.”

Methods

Rigorous AI Testing

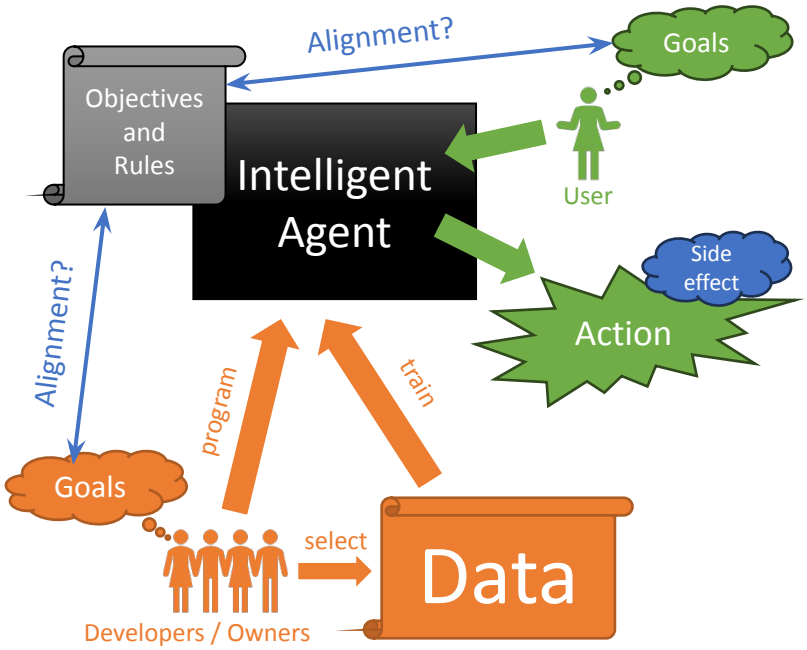
Continuous monitoring of deployed Systems

Adversarial robustness □ defend against malicious inputs

Intelligent Agents and Safety Challenges

Agents are "optimizers!" □ may misalign with human values

- **Goal/reward alignment** issues
- **Reward hacking** □ unintended side effects
- **Instrumental convergence** □ All agents pursue sub goals like:
  - Acquiring more resources.
  - Preserving themselves.
- Need to embed **social norms** and **ethical constraints**.



ator 3: Rise of the Machines. Warner Bros.

Summary

- We explored the **foundations of AI**:
  - What AI is and how it is defined.
  - Intelligence & intelligent behavior.
  - Goals, types, and approaches to AI.
  - Core areas and real-world applications.
  - Rational & intelligent agents.
  - Challenges, ethics, and limitations.

- *AI is not just about building smart machines — it's about understanding intelligence itself and applying it responsibly.*