Introduction to AI

Jacopo Mauro

Slide Based on Slides of the Artificial Intelligence: A Modern Approach book

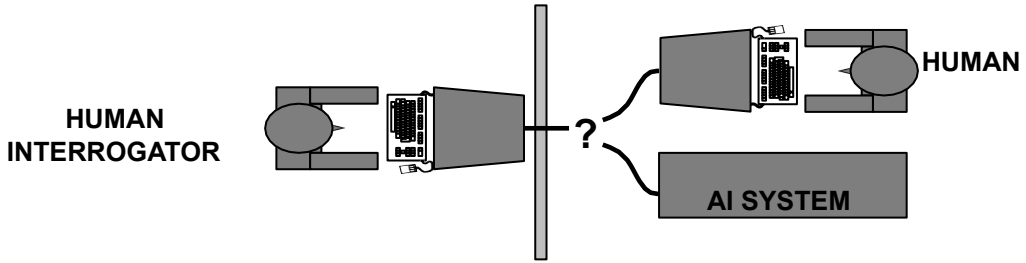
What is AI?

No universal accepted definition of what AI means

Different schools of thought:

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

Acting humanly: The Turing test



Turing (1950) “Computing machinery and intelligence”:

- “Can machines think?” → “Can machines behave intelligently?”

Operational test for intelligent behavior: the Imitation Game

Problem: Turing test is not reproducible or amenable to mathematical analysis

Thinking humanly: Cognitive Science

Cognitive Science. Interdisciplinary study of the mind and its processes, including how humans think, learn, perceive, remember, and interact with the world

Requires scientific theories of internal activities of the brain

- How to validate? Requires
  - Predicting and testing behavior of human subjects (top-down)
  - Direct identification from neurological data (bottom-up)

Both approaches (roughly, Cognitive Science and Cognitive neuroscience) are now distinct from AI

Thinking rationally: Laws of Thought

Normative (or prescriptive) rather than descriptive

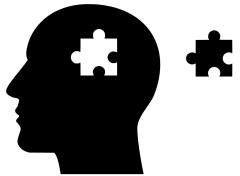
Several Greek schools developed various forms of logic:

- notation and rules of derivation for thoughts;

Direct line through mathematics and philosophy to modern AI

Problems:

- Not all intelligent behavior is mediated by logical deliberation
  - reactive agents → frog capturing a fly, remove hand from stove



Acting rationally

Rational behavior: doing the right thing

The right thing: that which is expected to maximize goal achievement, given the available information

Doesn’t necessarily involve thinking—e.g., blinking reflex—but thinking should be in the service of rational action

Rational agents

This course is about designing rational agents

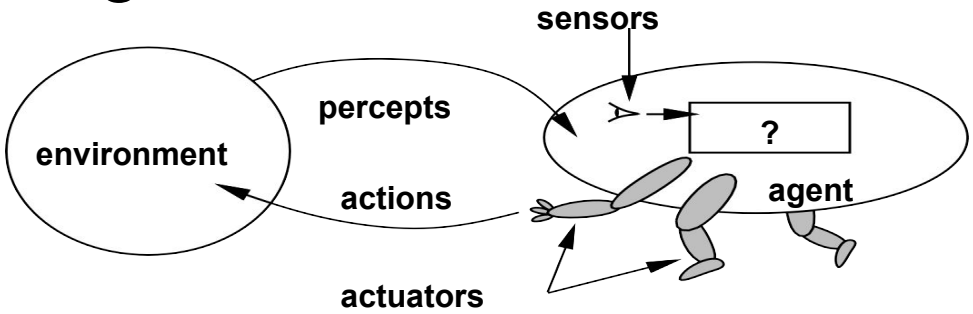
An agent is an entity that perceives and acts

For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance

Difficulty: computational limitations make perfect rationality unachievable

→ design best program for given machine resources

Agents and environments



Agents include humans, robots, softbots, thermostats, etc.

An agent can be anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators

The agent function maps from percept histories to actions:

$$f : P* \rightarrow A$$

The agent program runs on the physical architecture to produce f

History of AI

1940s–1950s: Early Foundations

- Birth of AI as a concept; Alan Turing proposes the "Turing Test."
- Development of the first electronic computers enabling complex calculations.

1956: The Birth of AI

- The term "Artificial Intelligence" coined at the Dartmouth Conference (John McCarthy)
- Early AI programs developed (e.g., Logic Theorist, which mimicked human reasoning).

1960s–1970s: Growth & Challenges

- Progress in symbolic AI and natural language processing.
- Introduction of neural networks (Perceptron by Frank Rosenblatt).
- Funding cuts during "AI winters" due to unmet expectations (complexity problems)

History of AI

1980s: Revival Through Expert Systems

- Development of rule-based systems to solve domain-specific problems.
- Renewed interest and funding in AI applications for business.

1990s–2000s: Machine Learning Era

- Shift towards data-driven approaches and statistical learning.
- Breakthroughs like IBM's Deep Blue defeating world chess champion Garry Kasparov (1997).

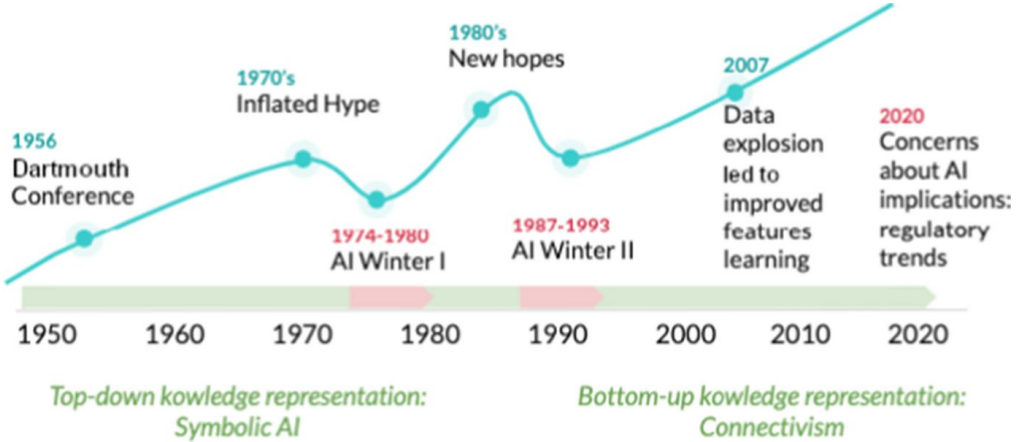
2010s: Deep Learning Revolution

- Advancements in neural networks and computational power.
- Success of AI in image recognition, speech processing, and games (e.g., AlphaGo defeating Lee Sedol in 2016).
- Rise of autonomous systems and virtual assistants (e.g., Siri, Alexa).

History of AI

2020s–Present: AI Everywhere

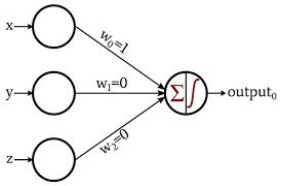
- Widespread application of AI in healthcare, finance, robotics, and more.
- Emergence of generative AI (e.g., ChatGPT) transforming content creation.
- Growing focus on ethics, fairness, and explainability in AI development.



Symbolic AI (Classical)

- Foundation: Logic, rules, and symbolic representations.
- Era: Dominated early AI research (1950s–1980s).
- Key Figures: John McCarthy, Marvin Minsky, Herbert Simon.
- Strengths: Transparent reasoning, excels in structured tasks (e.g., theorem proving).
- Weaknesses: Struggles with perception, requires manual knowledge engineering.

Neuro AI (Connectionist)

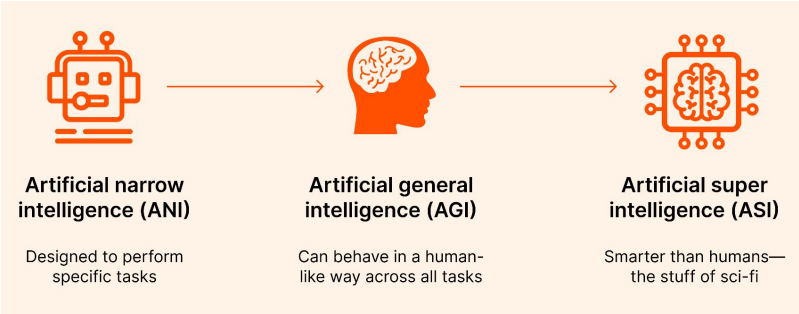


- Foundation: Inspired by biological neural networks.
- Era: Gained prominence from the 1980s, revolutionized AI in the 2010s.
- Key Figures: Geoffrey Hinton, Yann LeCun, Yoshua Bengio.
- Strengths: Excels in perception tasks (e.g., image recognition), adapts to unstructured data.
- Weaknesses: Black-box models, needs large data and high computational power.

AGI

Artificial General Intelligence (AGI)

something capable of understanding, learning, and applying knowledge across a wide range of tasks at a level comparable to or exceeding human intelligence



AGI

Key Characteristics of AGI:

- Generalization: Can transfer knowledge from one domain to another, adapting to new tasks without needing extensive retraining.
- Autonomy: Functions independently, making decisions and solving problems without predefined rules or heavy supervision.
- Human-Like Cognition: Capable of reasoning, problem-solving, and understanding language, emotions, and context at a human level.
- Learning Efficiency: Learns new skills or knowledge with minimal input, akin to how humans learn.

AI & Philosophy

Philosopher John Searle (1980):

- **weak AI**: idea that machines could act as if they were intelligent
- **strong AI**: assertion that machines that do so are actually consciously thinking (not just simulating thinking)

Some philosophers claim that a machine that acts intelligently would not be actually thinking, but would be only a simulation of thinking

Good Old-Fashioned AI

Good Old-Fashioned AI (GOFAI)

- Corresponds to the simplest logical agent design (details during the course)
- Strength: Formalizes intelligence with necessary and sufficient logical rules
- Challenge:
  - Qualification Problem → Capturing all contingencies for appropriate behavior is complex and often impractical/impossible
  - Turing's "argument from informality of behavior" says that human behavior is far too complex to be captured by any formal set of rules
  - Dreyfus's Argument → Logical inference engines lack the context and adaptability of situated agents.
  - Embodied Cognition Approach
    - Intelligence is not disembodied but situated in a physical body interacting with an environment
    - Cognitive processes are shaped by physical experiences and sensory input.

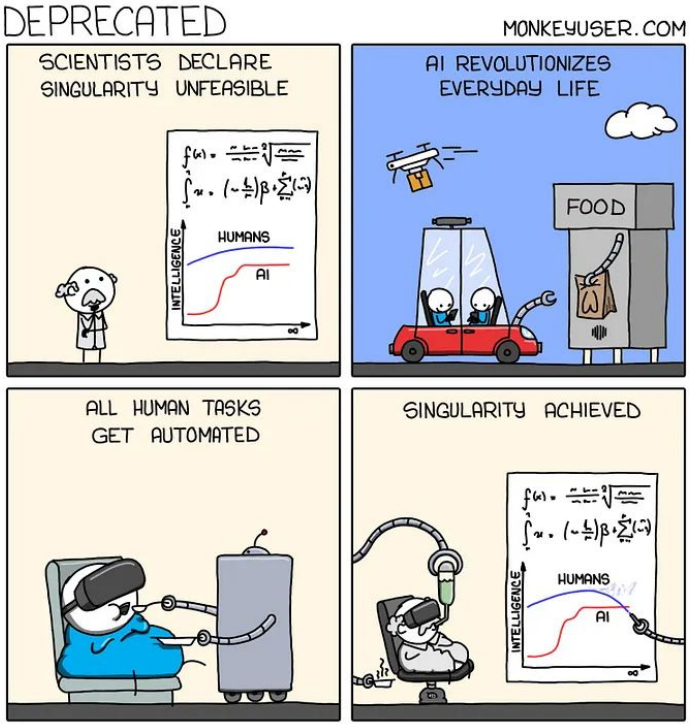
Disability Argument

The “argument from disability” makes the claim that “a machine can never do X.

- Turing’s lists of X: be kind, resourceful, beautiful, friendly, have initiative, have a sense of humor, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream, make someone fall in love with it, …, be the subject of its own thought, have as much diversity of behavior as man, do something really new.
- Some of these are rather easy to be replicated by AI. However some are not possible

Overall, programs exceed human performance in some tasks and lag behind on others.

Humans as comparison?



Mathematical Objection

AI systems, based on computational models, face inherent mathematical limitations.

Rooted in foundational results of mathematics and computation:

- Gödel's Incompleteness Theorems
- Turing’s Halting Problem

Criticism of the Objection

- Human Cognition Comparison: Humans are also limited by these principles; yet, this does not undermine human intelligence.
- Practical AI: AI focuses on approximation and probabilistic reasoning, avoiding reliance on strict formal proofs.

Can Machines Really Think?

The Polite Convention

- Turing's Argument: Assume both humans and machines think unless proven otherwise.
- Searle's Rejection: Thinking requires understanding, not just processing rules.

The Chinese Room Thought Experiment

- Imagine a human, who understands only English, inside a room with:
  - A rulebook (written in English) for manipulating symbols.
  - Stacks of paper with symbols resembling Chinese characters
- Symbols are passed into the room from outside.
- The human follows the rulebook, matches symbols, and produces a response.
- Responses are sent back outside, appearing as coherent Chinese communication.

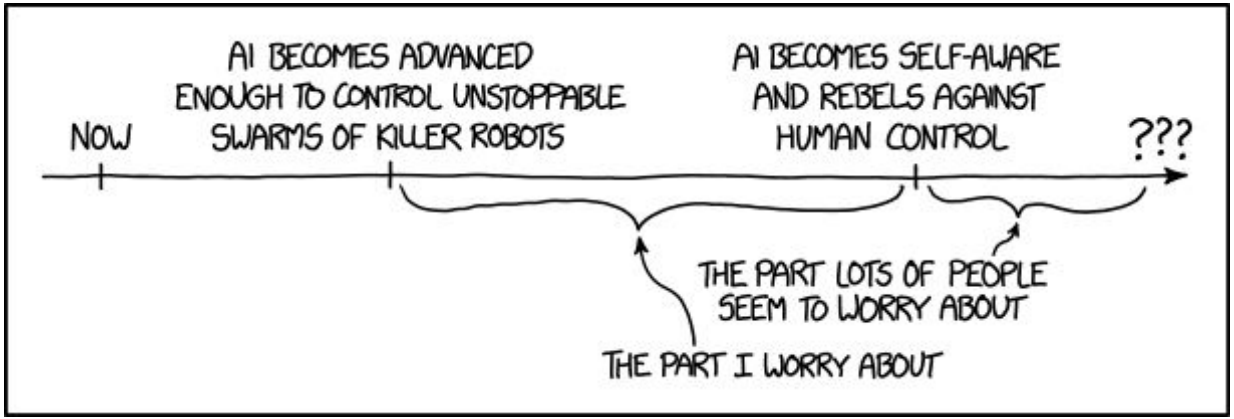
Computers work similarly by manipulating symbols without true understanding → they lack genuine cognition

Homework

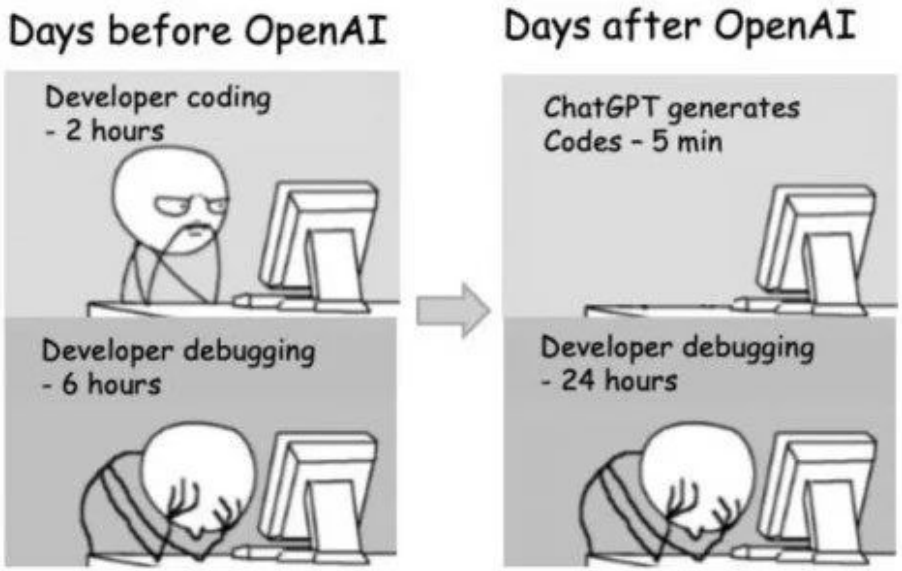
Read Chapter 1

Read Section 28.1-2 (until the Ethics of AI - covered in AI502: Ethics and Privacy)

Memes


NOW
AI BECOMES ADVANCED ENOUGH TO CONTROL UNSTOPPABLE SWARMS OF KILLER ROBOTS
AI BECOMES SELF-AWARE AND REBELS AGAINST HUMAN CONTROL
???
THE PART LOTS OF PEOPLE SEEM TO WORRY ABOUT
THE PART I WORRY ABOUT

Memes


Days before OpenAI
Developer coding
- 2 hours
Developer debugging
- 6 hours
Days after OpenAI
ChatGPT generates
Codes - 5 min
Developer debugging
- 24 hours

Memes


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

Chihuahua or Muffin?