

CS 15-281:
AI: Representation and Problem Solving

Tuomas Sandholm and Vincent Conitzer

Jan 13th, 2025

Course team

Professors



Tuomas Sandholm



Vincent Conitzer

TAs



Shruti



Theo



Avi



Ethan



Roy

Outline for today



- Course logistics
- What is AI?
- History of AI
- Current applications of AI
- What is an agent?

Course Information

Website: www.cs.cmu.edu/~15281

Canvas: canvas.cmu.edu



Gradescope: gradescope.com



Communication: plazza.com/cmu/spring2025/15281/home



Please check website for: Prerequisites/Corequisites/Course Scope

Grading policy

- Final scores will be composed of:
 - 15% Midterm 1
 - 15% Midterm 2
 - 30% Final exam
 - 20% Programming homework
 - 10% Written homework
 - 5% Online homework
 - 5% Participation
 - 5% for 80% or more poll participation + recitations attended
 - 0% for 50% or less poll participation + recitations attended
 - Linearly increases in between

Participation Points and Late Days

Participation

- Lecture Polls (on Piazza during lectures)
- Recitation Attendance

Late Days

- Total 6 late days allowed for the semester
 - At most 2 late days for any one homework
 - There will be no questions asked for use of these late days
- No further extensions will be granted unless there are extremely extenuating circumstances

Safety and Wellness

Virtual and in-person office hours!

Lectures are recorded for everyone to use, no questions asked.

Use the late days appropriately.

Contact the TAs ASAP if you think you'll miss more than one class so we can make a plan for how to catch up!

Announcements

Recitation starting this Friday

- Required. Materials are fair game for exams
- Attendance counts towards participation points
- Attend any recitation section for the first 3 weeks, then commit to one

Assignments:

- P0: Python & Autograder Tutorial (out tonight)
 - Required, but worth zero points
 - Due 1/23
- HW1 (online) – out tonight
 - Due 1/23
- Office hours start today (details on course website)
 - Homework questions should generally be directed to TAs
 - Instructors can offer more help in broader and conceptual questions, so their office hours are better used for that

Outline

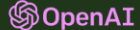
- Course logistics
- ➔ • What is AI?
- History of AI
- Current applications of AI
- What is an agent?



What is AI?

Measurable progress in AI



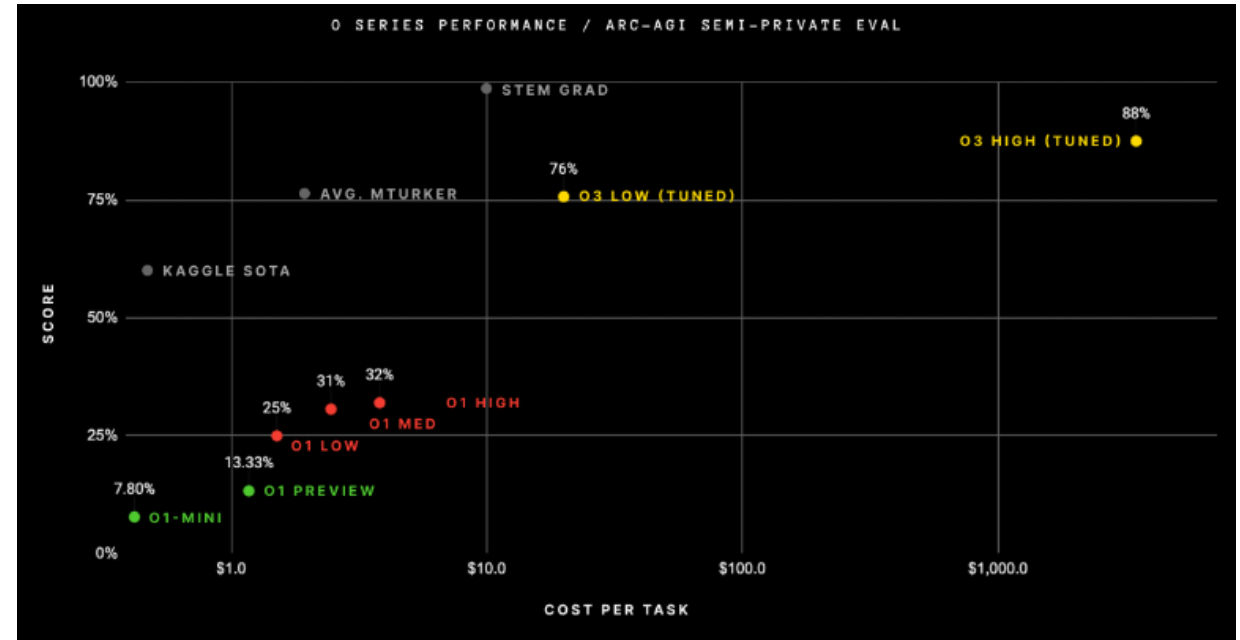
 OpenAI

ChatGPT: Optimizing Language Models for Dialogue

NEWS | 12 January 2023

Abstracts written by ChatGPT fool scientists

Researchers cannot always differentiate between AI-generated and original abstracts.



“Chinese room” argument [Searle 1980]

- **Person** who knows English but not Chinese sits in room
- Receives notes in Chinese
- Has systematic English **rule book** for how to write new Chinese characters based on input Chinese characters, returns his notes
 - Person=CPU, rule book=AI program, really also need lots of paper (storage)
 - Has no understanding of what they mean
 - But from the outside, the room gives perfectly reasonable answers in Chinese!
- Searle’s argument: the room has no intelligence in it!

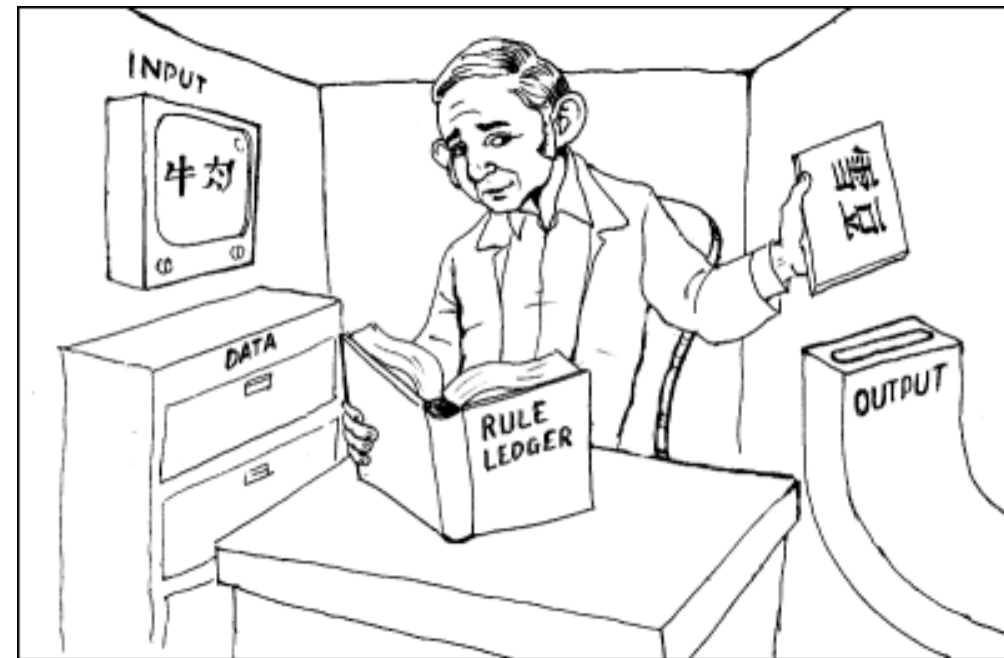
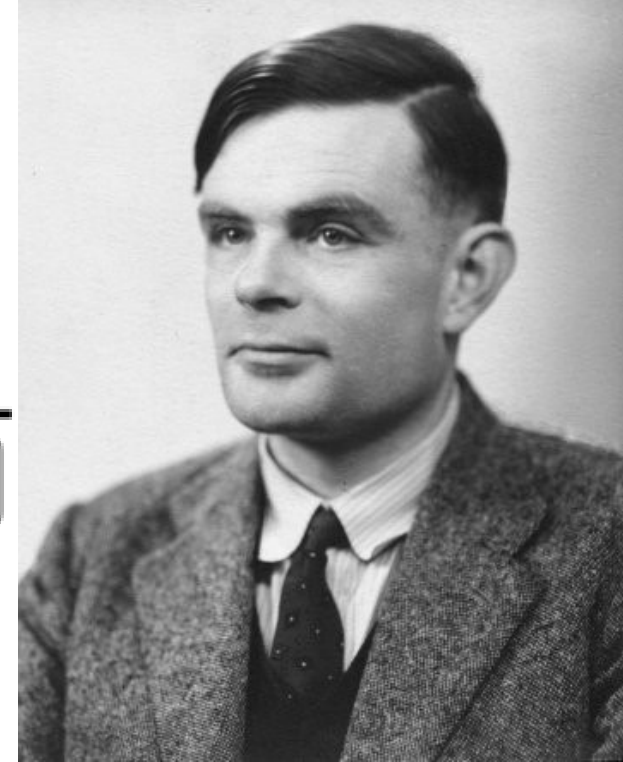
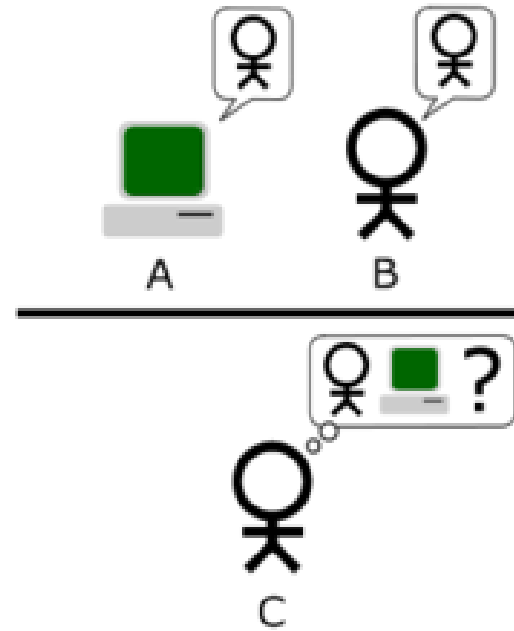


image from <http://www.unc.edu/~prinzh/pictures/c-room.gif>

Turing Test

- In 1950, Turing defined a test of whether a machine could “think”
- “A human judge engages in a natural language conversation with one human and one machine, each of which tries to appear human. If judge can’t tell, machine passes the Turing test”



Some classic definitions

Think like a human

- Cognitive science / neuroscience
- Does intelligence require biology?

Think rationally

- Logic and automated reasoning
- But, not all problems can be solved just by these techniques

Act like a human

- Turing test
- ELISA, Loebner prize
- “What is 1228×5873 ”? ... “I don’t know, I’m just a human”

Act rationally

- Basis for intelligent agents framework
- Unclear if this captures the current scope of AI research

The pragmatist view

“AI is that which appears in academic conferences on AI”

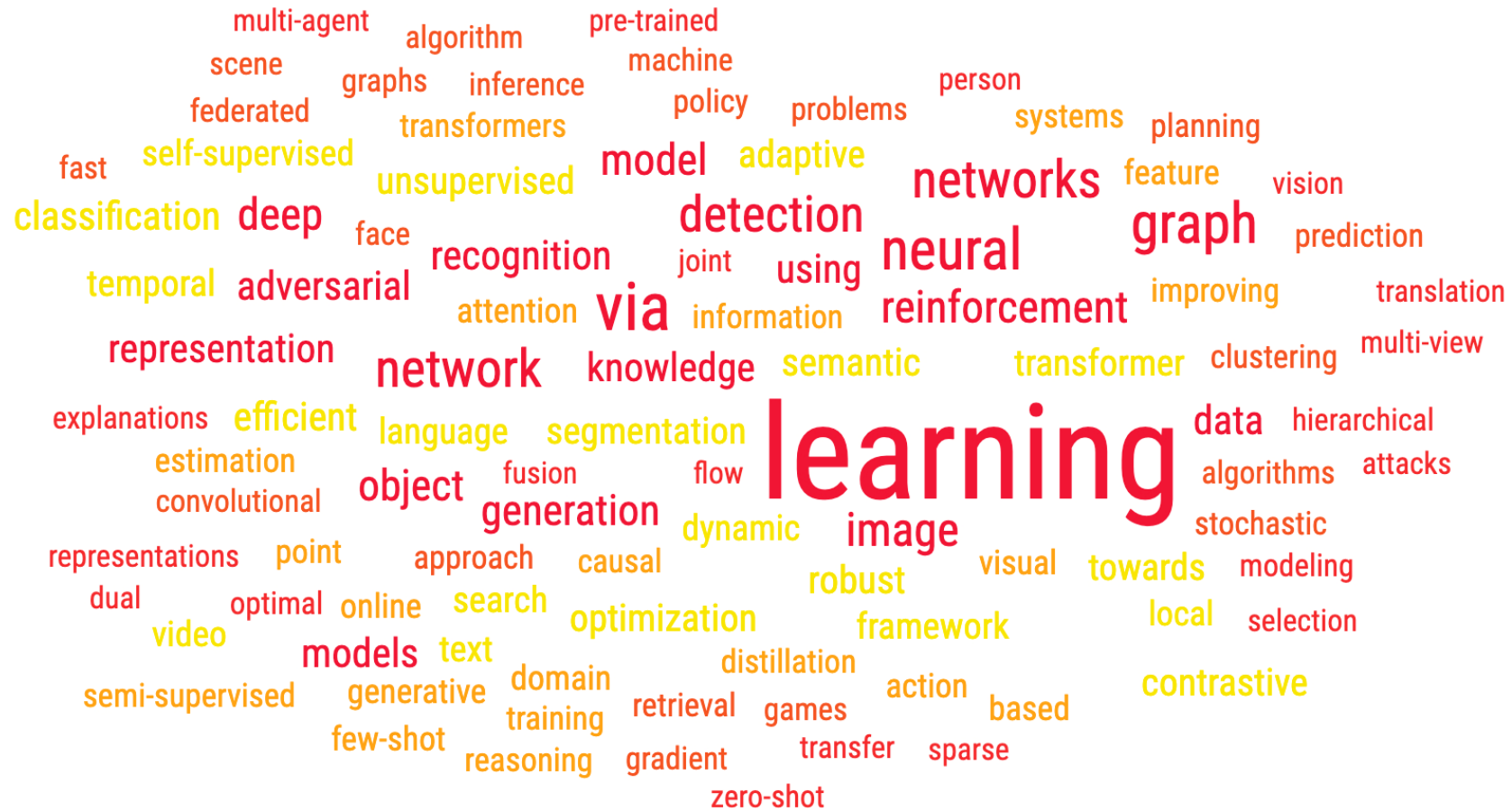
Alternate definition: “AI is that which marketing departments call AI”

Paper titles in AAAI



2000s

Paper titles in AAAI




2020s

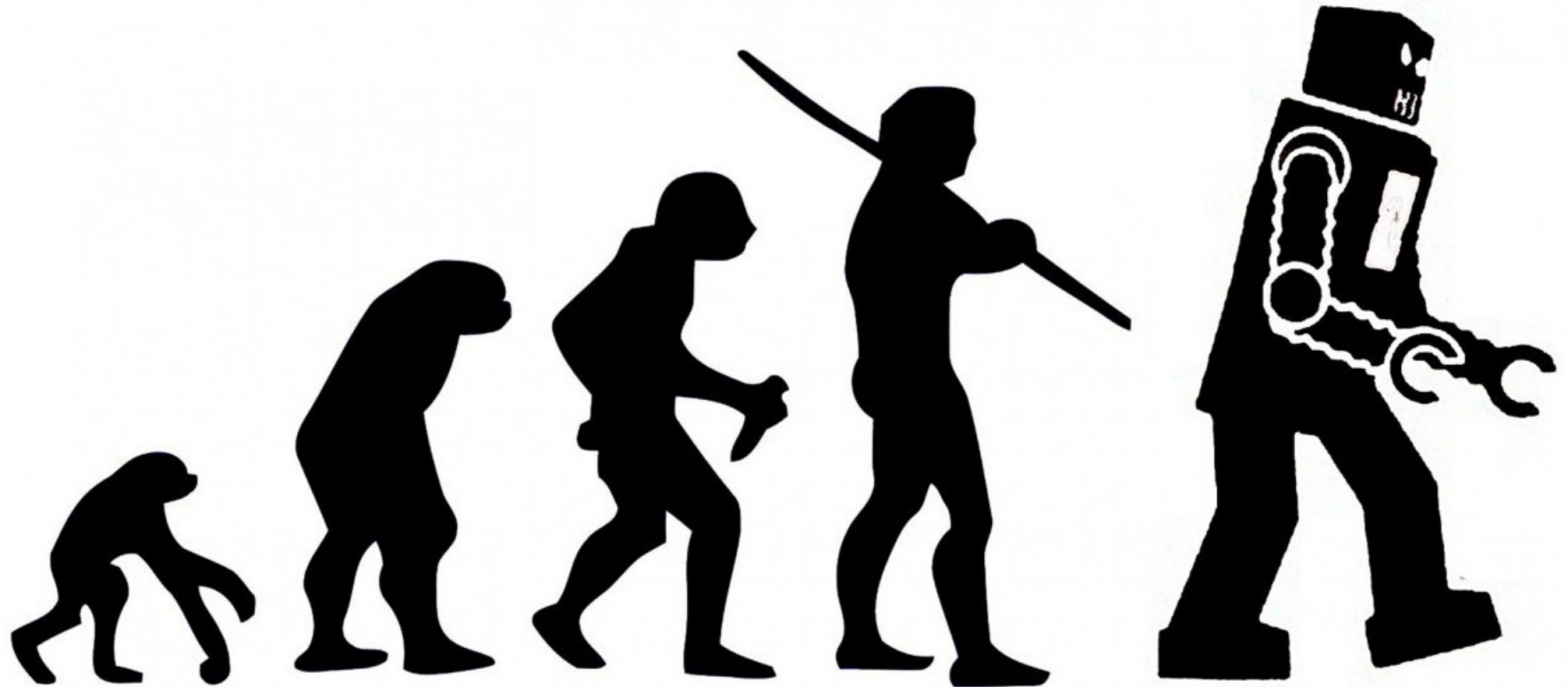
A broader (but vague) definition

Artificial intelligence is the development and study **of computer systems** to address **complex** real-world problems typically associated with some form of intelligence

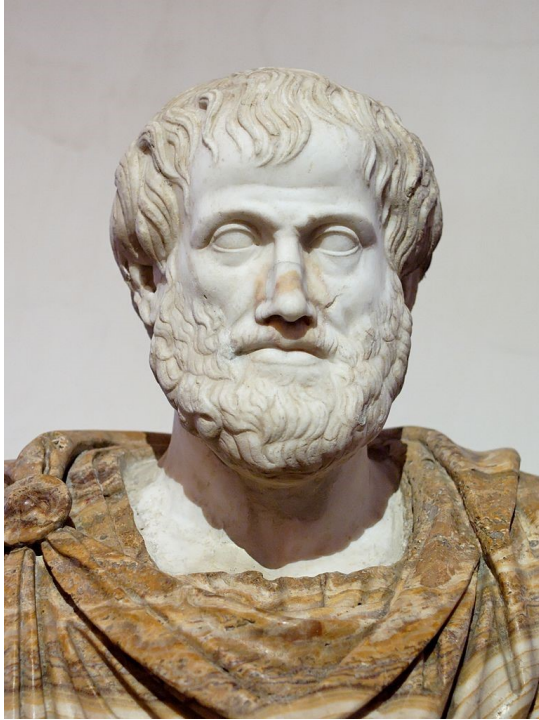
Outline

- Course logistics
- What is AI?
-  • History of AI
- Current applications of AI
- What is an agent?

A brief history of AI



Prehistory (400 B.C -)



Aristotle

Philosophy: mind/body dualism, materialism

Mathematics: logic, probability, decision theory, game theory

Cognitive psychology

Computer engineering

Birth of AI (1943 –1956)

- [1943] **McCulloch and Pitts**: simple neural networks
 - A computational model inspired by the brain

**A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN
NERVOUS ACTIVITY***

- **WARREN S. MCCULLOCH AND WALTER PITTS**
University of Illinois, College of Medicine,
Department of Psychiatry at the Illinois Neuropsychiatric Institute,
University of Chicago, Chicago, U.S.A.

Birth of AI (1943 –1956)

- [1943] McCulloch and Pitts: simple neural networks
- [1950] Turing test



A. M. Turing (1950) Computing Machinery and Intelligence. *Mind* 49: 433-460.

COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

Birth of AI (1943 –1956)

- [1943] McCulloch and Pitts: simple neural networks
- [1950] Turing test
- [1955-56] Newell and Simon: Logic Theorist



First program deliberately engineered to perform automated reasoning; eventually goes onto being called the **first artificial intelligence program**

Uses search + heuristics

Birth of AI (1943 –1956)

- [1943] McCulloch and Pitts: simple neural networks
- [1950] Turing test
- [1955-56] Newell and Simon: Logic Theorist
- [1956] Dartmouth workshop (coined the term **artificial intelligence**)

Birth of AI (1956 Dartmouth workshop)

- 1956: workshop at Dartmouth college; 11 attendees including John McCarthy, Marvin Minsky, Claude Shannon, Allen Newell and Herbert Simon



John McCarthy

Aim for **general principles**: Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it

We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer

Early successes in AI (1950s – 60s)

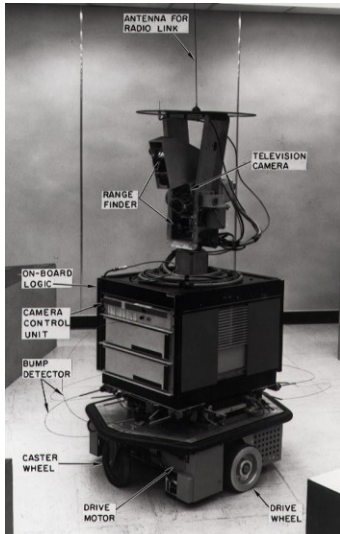


[1952] **Checkers**: Arthur Samuel's program learned weights via self-play and played at a strong amateur level

[1958] **McCarthy LISP**, advice taker, time sharing

[1968-72] **Shakey the robot**

[1971-74] **Blocksworld** planning and reasoning domain



Early success in AI (1950s – 60s)

Overwhelming optimism

Machines will be capable of doing any work a man can do – Herbert Simon [1965]

Within a generation, I am convinced, few compartments of intellect will remain outside the machine's realm—the problems of creating “artificial intelligence” will be substantially solved – Marvin Minsky [1960s]

I visualize a time when we will be to robots what dogs are to humans. And I am rooting for the machines – Claude Shannon

First AI winter (Later 1970s)

AI did not live up to the promise

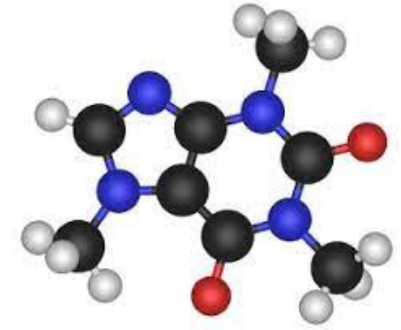
- 1966 ALPAC report cut off funding for machine translation
 - *we will not suddenly or at least quickly attain machine translation*
- 1973 LightHill report
 - In no part of the field have the discoveries made so far produced the major impact that was then promised
- 1970s DARPA cut funding

What went wrong?

- **Limited compute:** search space grows exponentially
- **Limited information** about the complex world
- How to address this? The answer at the time was **knowledge-based systems** or expert systems that encode prior knowledge
 - Moved away from the optimism of generality...

Knowledge based systems (1970s-80s)

- [1971-74] Feigenbaum's **DENRAL** to infer molecular structure from mass spectrometry
- **MYCIN**: diagnose blood infections, recommend antibiotics
- 1981–Japan's “**fifth generation**” computer project, intelligence computers running Prolog
- [1982] **XCON** or **R1 expert system** to configure customer orders; deployed at DEC and saved \$40 million a year



Second AI winter (late 1980s to early 1990s)

- Knowledge based systems also failed to deliver at the time
 - Required **considerable manual effort** to develop and maintain
 - “Knowledge acquisition bottleneck”
 - Deterministic rules could not handle **uncertainty**
- [1987] DARPA cuts AI funding for expert systems
- [1991] Japan’s fifth generation project fails to meet goals

Splintering & changing of AI, and cross-fertilization with other fields (mid 1990s-)

- Many subfields and ideas: machine learning, computer vision, robotics, language processing, multiagent systems, ...
- Ideas from different fields
 - Bayes rule from probability
 - Cross-fertilization between search in AI and integer programming in operations research
 - Game theory from mathematics and economics
 - Stochastic gradient descent from statistics
 - Value iteration from control theory
 - Artificial neural networks from neuroscience
- AI becomes more mathematical
- Statistical rigor starts to be required in experimental results

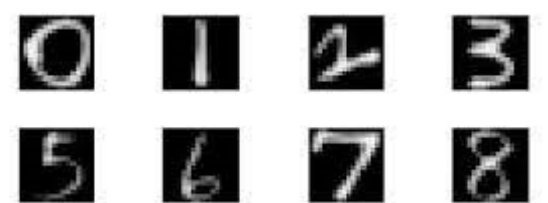
Beyond symbolic AI

- *Symbolic AI*: top-down approach with a vision
- *Neural AI*: bottoms-up approach inspired by the human mind
- Neural AI had its own story with promise, successes, and winters

Phase one of neural AI

- [1943] McCulloch/Pitts model for computation
- [1958] Rosenblatt's perceptron algorithm for binary classification
- [1969] Perceptrons book showed that linear networks could not solve XOR \Leftrightarrow part of the AI winter that killed neural net research

Phase two of neural AI

- [1986] Popularization of the backpropagation algorithm (originally invented by Seppo Linnainmaa in 1972) for training multi-layer networks by **Rumelhardt, Hinton, Williams**
- [1989] **LeCun** applied CNNs (pioneered previously by CMU's Alex Waibel) for handwriting recognition 
- Was still hard to train very deep networks successfully---hence successes limited to simple tasks

Phase three of neural AI

- “AlexNet moment” [2012]: huge gains on a real-world image classification task by successfully training deep networks
 - Large scale data (ImageNet) + GPUs + training heuristics

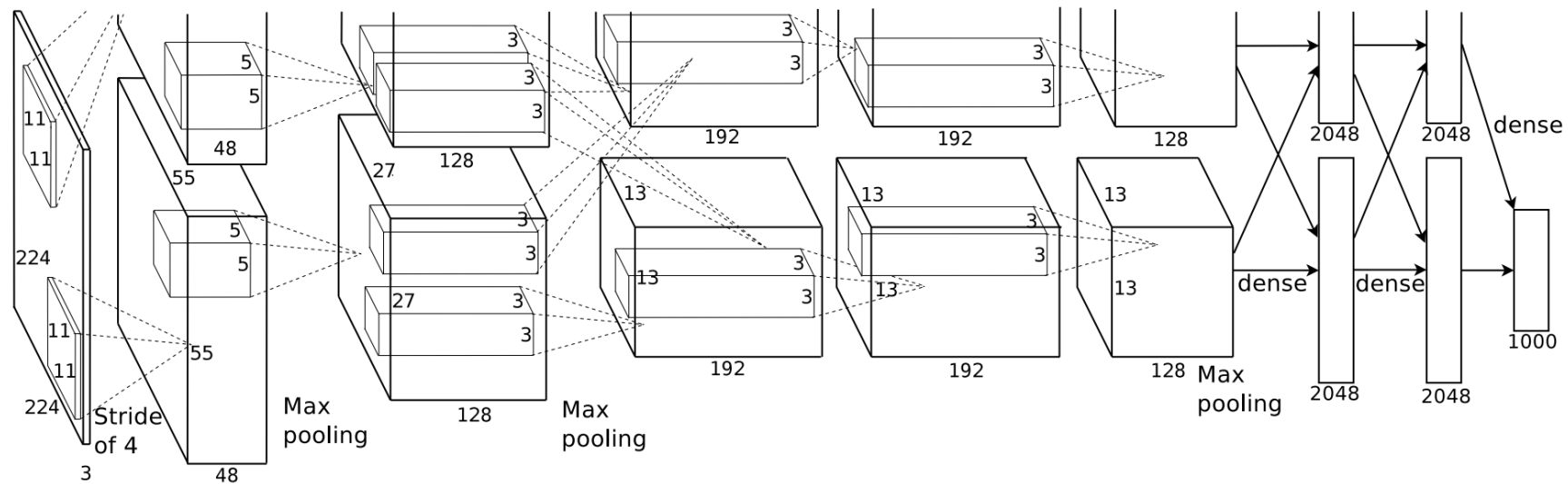


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network’s input is 150,528-dimensional, and the number of neurons in the network’s remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

Phase three of neural AI

- Ever larger amounts of data and compute (both offline and real time)
- [2020] GPT-3 with 175B parameters, trained on about 45TB of text data from different datasets
- [2024] Llama-3.1 was released on July 23, 2024, with three sizes: 8B, 70B, and 405B parameters

The AI renaissance

- [1997] DeepBlue defeats Garry Kasparov



- [1995] NavLab5 automobile drives across country 98% autonomously



- [2005, 2007] Stanford and CMU win DARPA grand challenges in autonomous driving

- [2011] IBM's Watson defeats human Jeopardy opponents




The AI renaissance

- [2016] DeepMind's **AlphaGo** beats top human player Lee Sodol
- [2017] CMU's **Libratus** defeats world's best players at two-player no-limit Texas Hold'em
- [2019] CMU's **Pluribus** defeats world's best players at multi-player no-limit Texas Hold'em
- [2021] DeepMind's **AlphaFold** offers highly accurate protein structure prediction

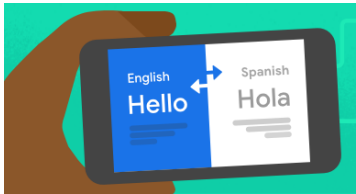


Outline

- Course logistics
- What is AI?
- History of AI
-  • Current applications of AI
- What is an agent?

AI in the real world

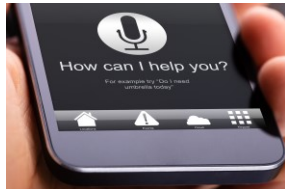
Translation



Spam detection



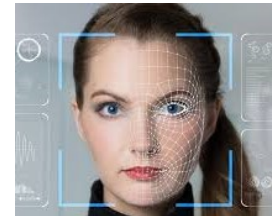
Voice assistant



Stock market



Facial recognition



Hiring systems



Autonomous driving



Large-scale combinatorial multi-attribute sourcing auctions 2001-2010

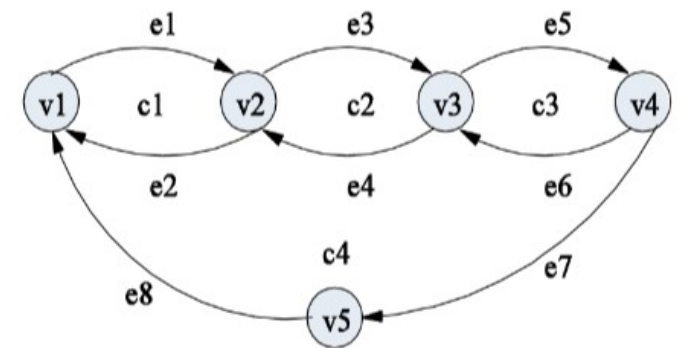
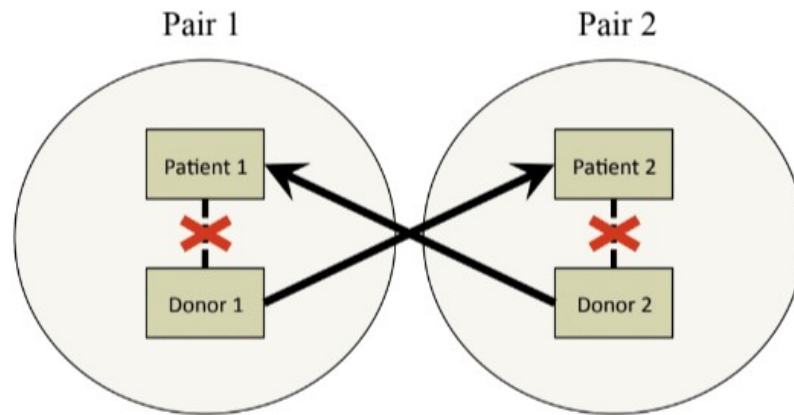
[Sandholm, Handbook of Market Design, Ch. 16, 2013]

- One of the first SaaS analytics companies
- Over 800 auctions, totaling over \$60 Billion
 - The most *expressive* auctions ever conducted
- Created 12.6% savings for buy side
- Suppliers also benefited
- Grew to 130 employees, operations on 4 continents
- Acquired in 2010
- Key AI technologies:
 - Winner determination algorithms
 - Bidding languages
 - Preference elicitation from multiple agents
 - Automated mechanism design

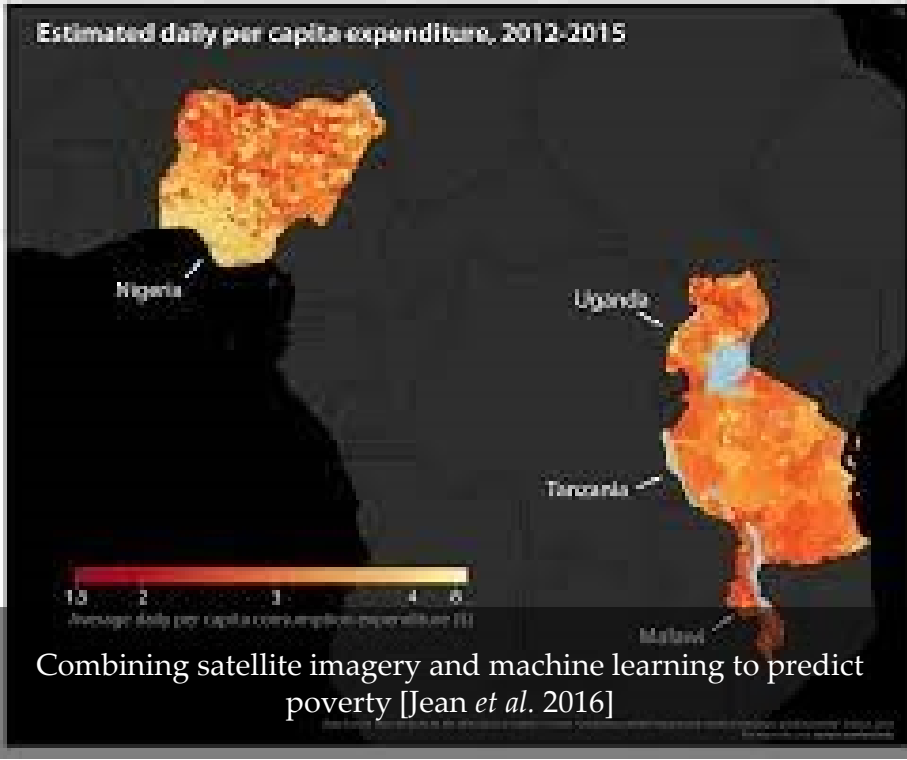


AI from Prof. Sandholm's lab has been running the national kidney exchange for UNOS since 2010.

Kidney exchange



AI and sustainability



ARTIFICIAL INTELLIGENCE

AI Discovers First New Antibiotic in Over 60 Years

A landmark study discovers a new antibiotic using AI deep learning technology.

Posted December 24, 2023 |  Reviewed by Abigail Fagan



KEY POINTS

- Antimicrobial resistance poses a major threat to public health.
- Researchers used AI to discover a new class of antibiotics to treat drug-resistant staph infections.
- The study used graph-based searches for chemical substructure options.

An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy.

"I won, and I didn't break any rules," the artwork's creator says.

Give this article   1.5K

*In Prof. Sandholm's August 2021 IJCAI John McCarthy Award talk, he predicted there will be **human-only categories** in future art competitions*



Jason Allen's A.I.-generated work, "Théâtre D'opéra Spatial," took first place in the digital category at the Colorado State Fair. via Jason Allen



By [Kevin Roose](#)

Sept. 2, 2022

New York Times

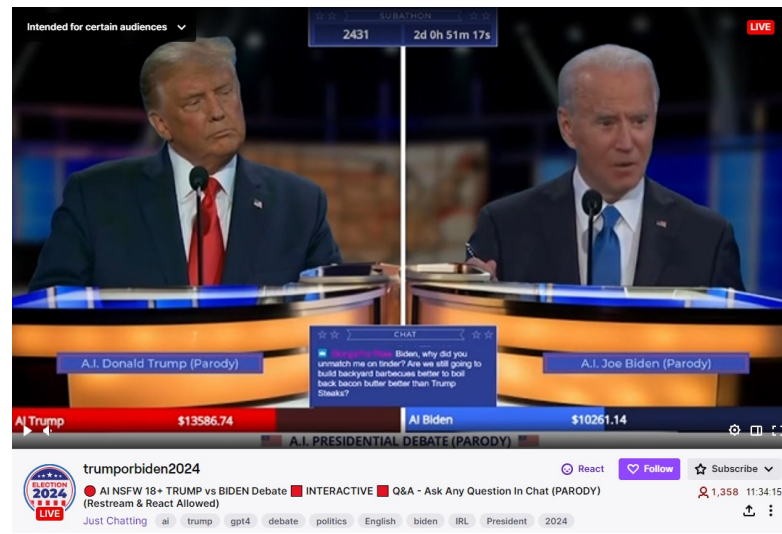


Deepfakes of
Prof. Sandholm
Spring 2023-

Some harmful applications of deepfakes



A recent video from Republican presidential candidate and Florida Gov. Ron DeSantis includes an image with three fake photos of former President Donald Trump and Dr. Anthony Fauci hugging. These three images appear to be AI-generated. DeSantis War Room/Screenshot and annotation by NPR




<https://www.twitch.tv/trumporbiden2024>



An A.I. Hit of Fake 'Drake' and 'The Weeknd' Rattles the Music World

A track like "Heart on My Sleeve," which went viral before being taken down by streaming services this week, may be a novelty for now. But the legal and creative questions it raises are here to stay.

 Give this article



 215



Labels hope that fans will continue to prize the work of artists, including the real Drake, above that of A.I.-generated imitations. Adam Riding for The New York Times



By [Joe Coscarelli](#)

New York Times

Published April 19, 2023 Updated April 24, 2023





THE
BEATLES

NOW AND THEN

MUSIC VIDEO

vevo

Large language models (LLMs) 2018-

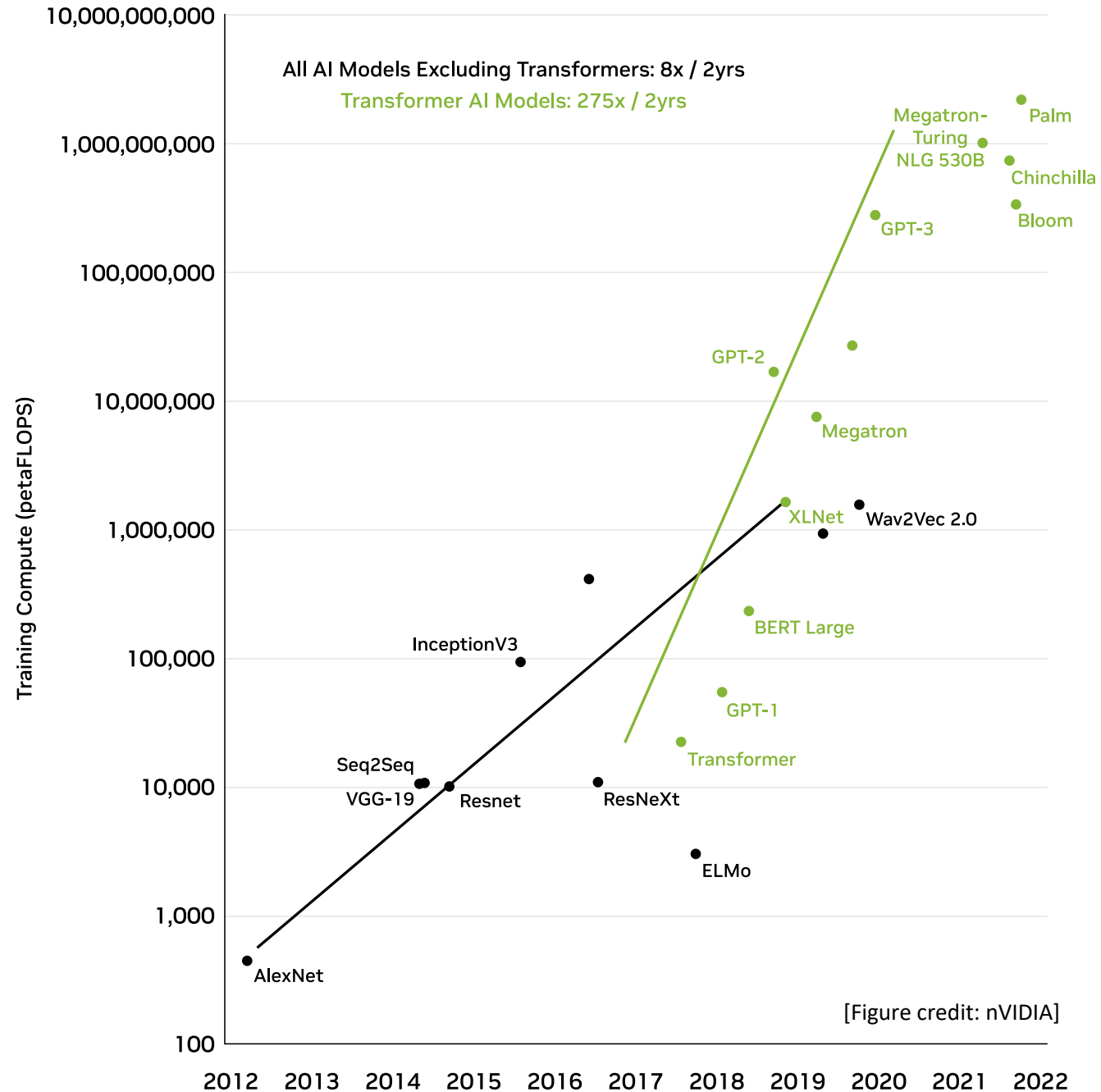
1. Generation (e.g., story writing, marketing content creation)
 - Tim Boucher wrote 97 books in 9 months using ChatGPT from OpenAI and Claude from Anthropic. He used Midjourney to generate images to match the story. Word count 2,000 - 83,000. Price \$1.99 - \$5.99
 - Non-AI-assisted record, by Barbara Cartland, is 191 books per year
2. Summarization (e.g., web search, legal paraphrasing, meeting notes summarization)
 - E.g., Bing co-pilot, Bard
 - In the future also question answering from specific corpora
3. Translation (e.g., between languages, text-to-code, text-to-database)
4. Classification (e.g., toxicity classification, sentiment analysis)
5. Chatbot (e.g., open-domain Q+A, virtual assistants)
 - E.g., Bing co-pilot, Bard

Downsides of LLMs

- Hallucination
 - May be impossible to fully get rid of
 - But code writing & external calls help
 - In important applications, getting sensical answers most of the time isn't good enough (cf. self-driving in 2015)
- Limited reasoning. Thus poor at
 - strategies against adversaries
 - Can't play tic-tac-toe, much less no-limit Texas hold'em poker
 - logic
 - collaborative filtering
 - ...

Downsides of LLMs ...

- Training cost
 - Open AI's GPT-4 took "more than \$100M"
[OpenAI CEO 4/17/2023]
(Microsoft invested \$1B+\$10B in OpenAI)
 - "Really good large language models take \$ Billions to train"
[Amazon CEO 4/13/2023]



Downsides of LLMs ...

- Proliferation of content
- Privacy loss & copyright issues
- Biases & inclusivity
- Explainability
- No “understanding” (?) but LLMs are used as if they understand in consulting, medicine, customer service, etc.

Regulation?

The AI Renaissance

Symbolic AI



Neural AI



tools and ideas
from other fields...

Optimization,
algorithms,
probabilistic inference,
statistics,
game theory, etc.

Outline

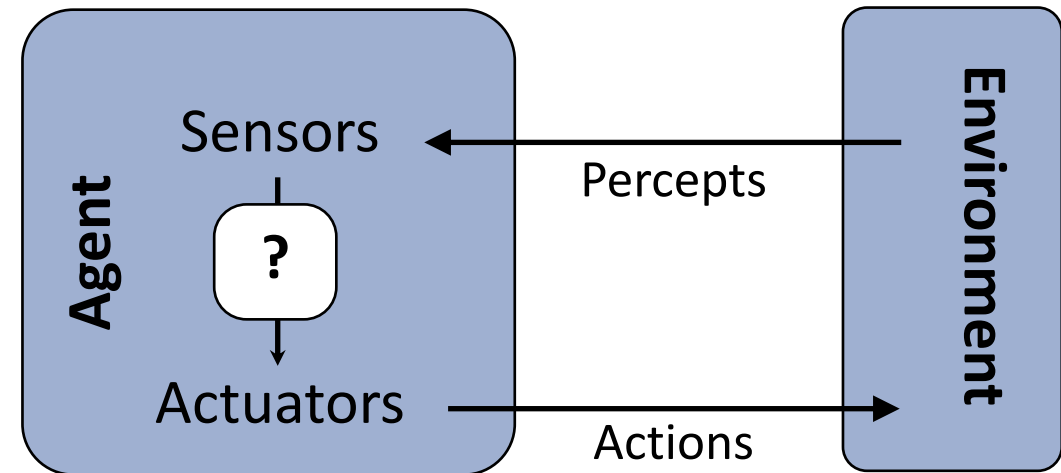
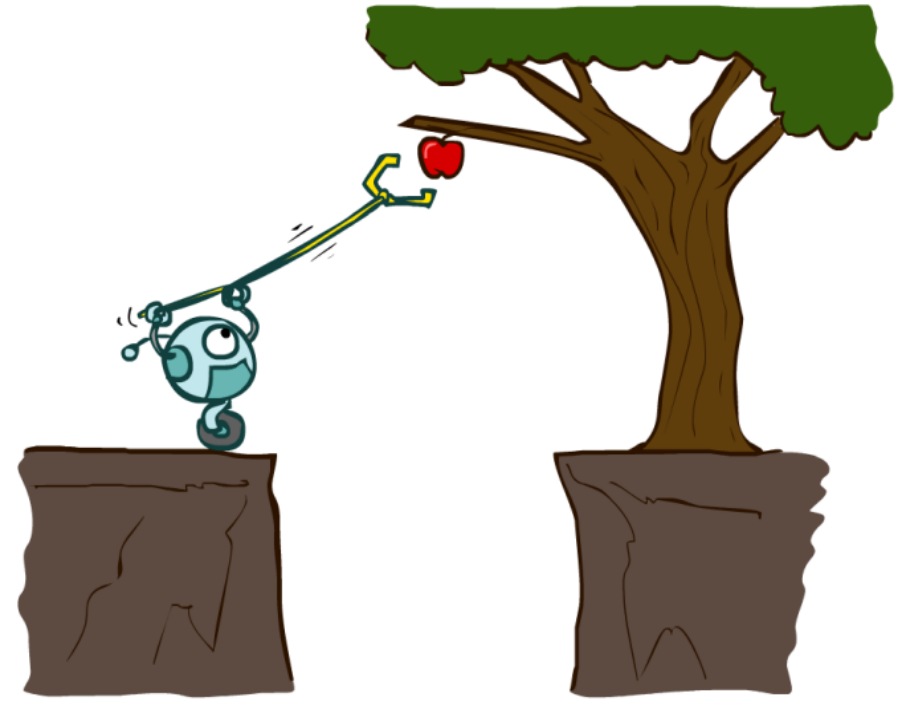
- Course logistics
- What is AI?
- History of AI
- Current applications of AI
- ➔ • What is an agent?

Designing Agents

An **agent** is an entity that *perceives* and *acts*.

Characteristics of the **percepts and state, environment, and action space** dictate techniques for selecting actions

How can we design an AI agent to solve our problems given their task environments?



Rational Decisions

We'll use the term **rational** in a very specific, technical way:

- Rational: maximally achieving pre-defined goals
- Rationality only concerns what decisions are made
(not the thought process behind them)
- Goals are expressed in terms of the **utility** of outcomes
- Being rational means **maximizing your expected utility**

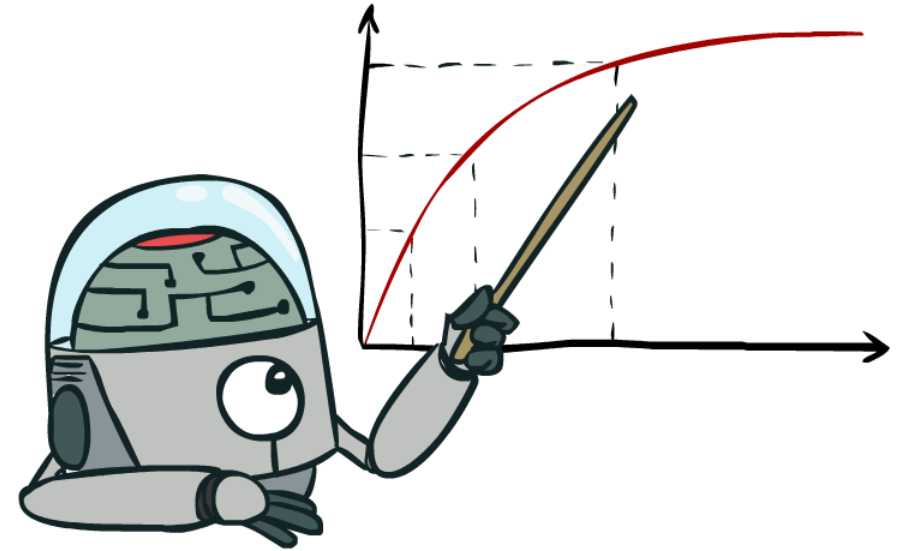
Another title for this course could be:

Introduction to Computational Rationality

Rationality, contd.

What is rational depends on:

- Performance measure
- Agent's prior knowledge of environment
- Actions available to agent
- Percept sequence to date



Being rational means **maximizing your expected utility**

Rational Agents

Are rational agents *omniscient*?

- No – they are limited by the available percepts and state

Are rational agents *clairvoyant*?

- No – they may lack knowledge of the environment dynamics

Do rational agents *explore* and *learn*?

- Yes – in unknown environments these are essential

So, rational agents are not necessarily successful, but they are *autonomous* (i.e., make decisions on their own to achieve their goals)

- Don't have to be used as autonomous but rather as recommenders of actions or in some settings as collaborators

Artificial Intelligence (AI) vs Machine Learning (ML)?

