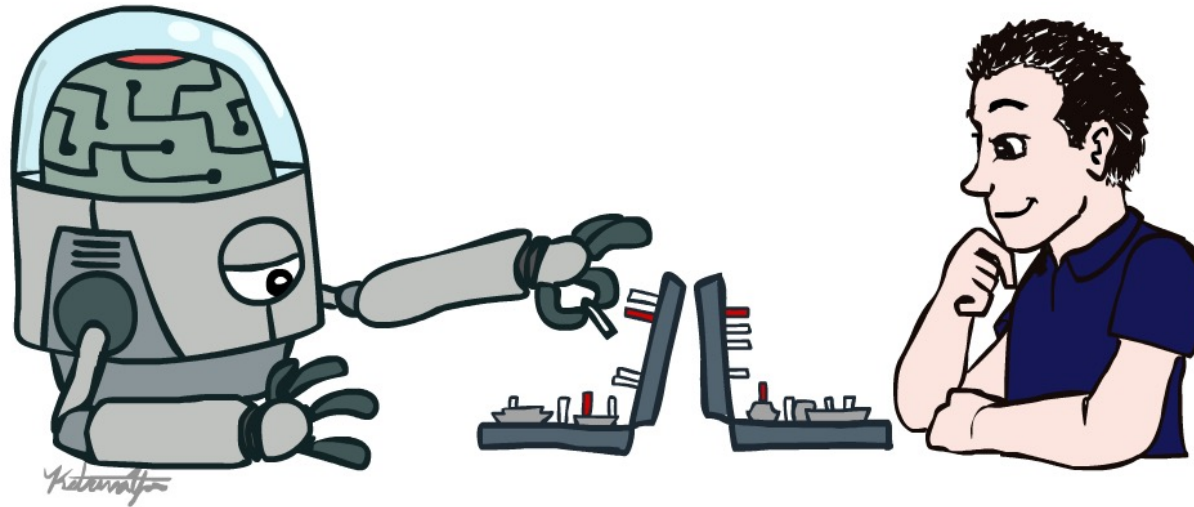


CS 188: Artificial Intelligence

Introduction



Spring 2023

University of California, Berkeley

[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley (ai.berkeley.edu).]

First Half of Today: Intros and Logistics

- Staff introductions: Cam, Michael, and course staff
- Course logistics
 - Lectures, discussions, office hours, and exams
 - Resources and communication platforms
 - Collaboration and academic honesty
 - DSP and extenuating circumstances
 - Stress management and mental health

Staff Introductions: Cameron Allen (he/him)

- **First-year postdoc + lecturer in EECS**
 - AI safety research at Center for Human-Compatible AI
 - First time teaching as co-instructor, so your feedback/advice/complaints are appreciated!
- **Did my PhD in CS at Brown (2023)**
 - Research focus: AI and decision making
- **Experience as TA / Guest Lecturer**
 - Learning and Sequential Decision Making (@Brown)
 - Reintegrating Artificial Intelligence (@Brown)
 - Intro to Artificial Intelligence (@Brown, @Duke)
- **Please call me “Cam” or “Cameron”!**
 - “Professor” is a big promotion I hope to earn someday



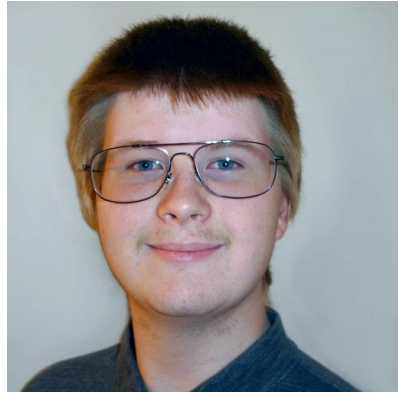
(Actual beard may vary.)

Staff Introductions: Michael Cohen (he/him)

- Doing a Postdoc
 - With Stuart Russell at Center for Human-Compatible AI
- PhD at Oxford
 - With the Future of Humanity Institute
- Research Focus
 - Under what circumstances would artificial agents face an incentive to take over the world?
 - How can we make arbitrarily competent artificial agents that don't take over the world?
 - How can policymakers stop people from making such dangerous AI systems?
- Favorite TV show: Survivor
- You can call me “Michael”!



Our talented course staff!



Evgeny Pobachienko
he/him



Sid Iju
he/him



Xavier Yin
he/him



Alina Trinh
she/her



Marwa Abdulhai
she/her



Cham Yao
he/him



Jerry Sun
he/him

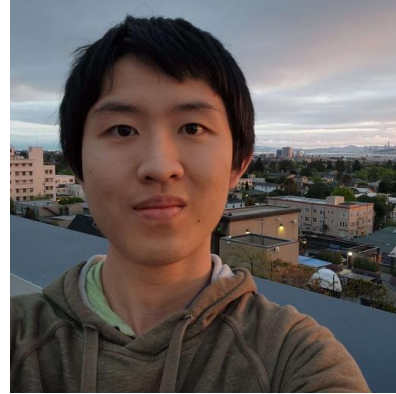


Joshua Liao
he/him

Our talented course staff!



Joy Liu
she/her



Kenny Wang
he/him



Michael Wu
he/him



Matei Gardea
he/him



Pranav Muralikrishnan
he/him



Sashrika Pandey
she/her



Ajay Sridhar
he/him



Aidan Leung
he/him

Our talented course staff!



Allen Cao
he/him



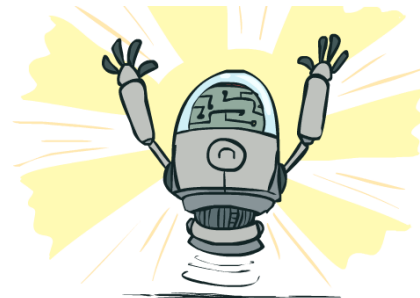
Saathvik Selvan
he/him



Samantha Huang
she/her



Somu Patil
he/him



*might have a few
more people join so
here's a placeholder*

Enrollment

- Course staff does not control enrollment; we have to follow department policy
 - Only CS majors will be able to enroll this spring
 - More details on the course website

Course Structure: Lectures

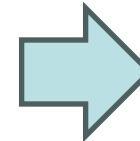
- You are here! (Wheeler 150)
- Tuesday/Thursday, 12:30–2:00 PM PT
- Attendance policy:
 - Attend lecture, in person.
 - We will take attendance. You will get extra credit.
- Lecture recordings:
 - Posted next day on website.
 - Do not count towards attendance.

Course Structure: Attendance

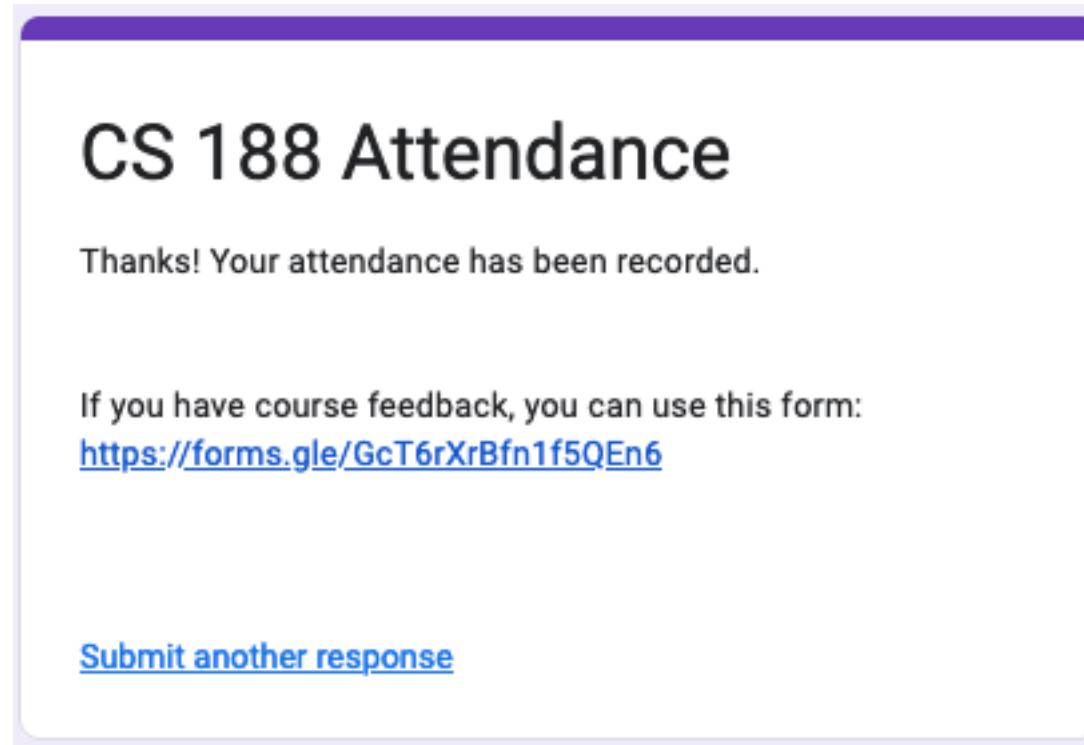
- Lecture code:
 - sp24ai



(Also on website)

A screenshot of the Google sign-in page. At the top is the Google logo, followed by the text "Sign in to continue to Forms". Below this is a text input field labeled "Email or phone" containing a redacted email address "@berkeley.edu". There are links for "Forgot email?", "Create account", and a blue "Next" button.A screenshot of the "CS 188 Attendance" page. The page title is "CS 188 Attendance". Below the title is a user profile section with a redacted email address "@berkeley.edu" and a "Switch account" link. A red asterisk indicates a required question. The "Email" section has a checked checkbox for "Record [redacted]@berkeley.edu as the email to be included with my response". At the bottom, there is a text input field for the attendance code, with "sp24ai" entered.

Course Structure: Feedback



- Includes link to course feedback form.
- Also available on course website.

Course Structure: Discussions

- We offer three types of discussions
 - Regular discussions
 - Exam prep discussions
 - Extended-time discussions
 - We'll try to make recordings, but no promises
- Discussion schedule available on website
 - Discussions start next week (Jan 22)
- You can attend any discussion section you want (no need to enroll in a section)
 - A bit of extra credit available for attendance

Course Structure: Office Hours

- Join in-person or remotely to talk to staff about content, ask questions on assignments, or raise any concerns you have
- Schedule and queue available on website
 - Office hours start next week (Jan 22)

Course Structure: Exams

- Save the dates!
 - Midterm: Tuesday, March 5, 2024, 7–9pm PT
 - Final exam: Thursday, May 9, 2024, 8-11am PT
- If you can't make it:
 - We'll offer remote exams at the listed time – need to fill out a form
 - We'll offer an in-person-only alternate exam right after the listed time
- More logistics closer to the exam

Resources

- Course website: <https://inst.eecs.berkeley.edu/~cs188/sp24/>
 - All resources (slides, notes, recordings, assignments, etc.) posted here
- Ed: Discussion forum
- Staff email for private concerns: cs188@berkeley.edu
 - Making a private post on Ed is easier/faster
- Gradescope: Submit assignments here

Grading Structure

- 6 Projects (25%)
 - Python programming assignments, autograded
 - You can optionally work with a partner
 - Reduced credit for submitting late, unless you have an extension
- 10 Homeworks (20%)
 - Electronic homework: Autograded on Gradescope
 - Written homework: One question per week, graded by TAs on correctness
 - Submit individually (but feel free to discuss with others)
 - No late submissions, unless you have an extension
- Midterm (20%), Final Exam (35%)

Extensions and Accommodations

- We'll drop your lowest homework score
- You have 5 slip days to use across the projects
 - See course policies page for details on how they work
- If you ever need an extension, please request one!
 - We're here to support you, and we understand that life happens.
 - Extension form will be posted on the website

DSP

- Disabled Students' Program (DSP)
 - There's a variety of accommodations UC Berkeley can help us set up for you in this class
 - <https://dsp.berkeley.edu/>
- Are you facing barriers in school due to a disability?
 - Apply to DSP!
 - We maintain proper access controls on this information: Only instructors, course managers, head TAs, and logistics TAs can access any DSP-related info
- Our goal is to teach you the material in our course. The more accessible we can make it, the better.

Collaboration and Academic Dishonesty

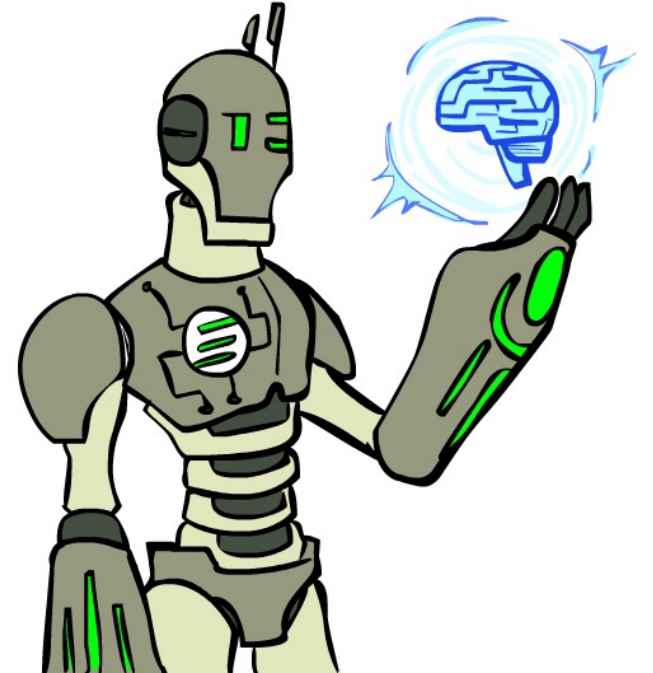
- We're here to help! There are plenty of staff and resources available for you
 - You can always talk to a staff member if you're feeling stressed or tempted to cheat
 - Collaboration on homework is okay, but please cite collaborators
 - Do not post solutions online or share with others!
- Academic dishonesty policies
 - Reported to Center of Student Conduct
 - Negative points on assignments, and/or F in the class

Stress Management and Mental Health

- **Your health is more important than this course**
- If you feel overwhelmed, there are options
 - Academically: Ask on Ed, talk to staff in office hours, set up a meeting with staff to make a plan for your success this semester
 - Non-academic:
 - Counselling and Psychological Services (CAPS) has multiple free, confidential services
 - Casual consultations: <https://uhs.berkeley.edu/counseling/lets-talk>
 - Crisis management: <https://uhs.berkeley.edu/counseling/urgent>
 - Check out UHS's resources: <https://uhs.berkeley.edu/health-topics/mental-health>

Second Half of Today: What is AI?

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



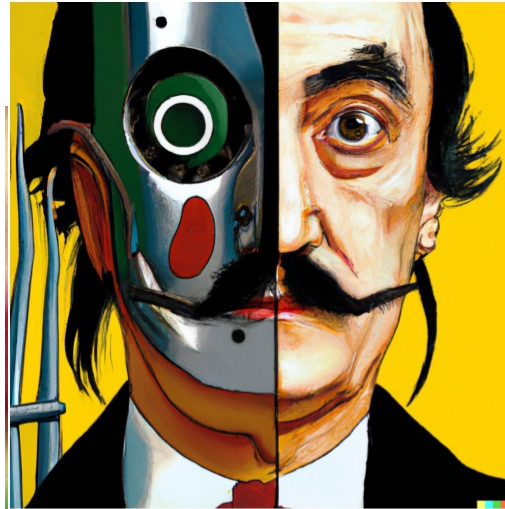
AI is having real-world impact

- Public imagination
 - Text assistants



AI is having real-world impact

- Public imagination
 - Text assistants
 - Image generation



vibrant portrait painting of Salvador Dalí with a robotic half face



a shiba inu wearing a beret and black turtleneck



a close up of a handpalm with leaves growing from it



an espresso machine that makes coffee from human souls, artstation



panda mad scientist mixing sparkling chemicals, artstation

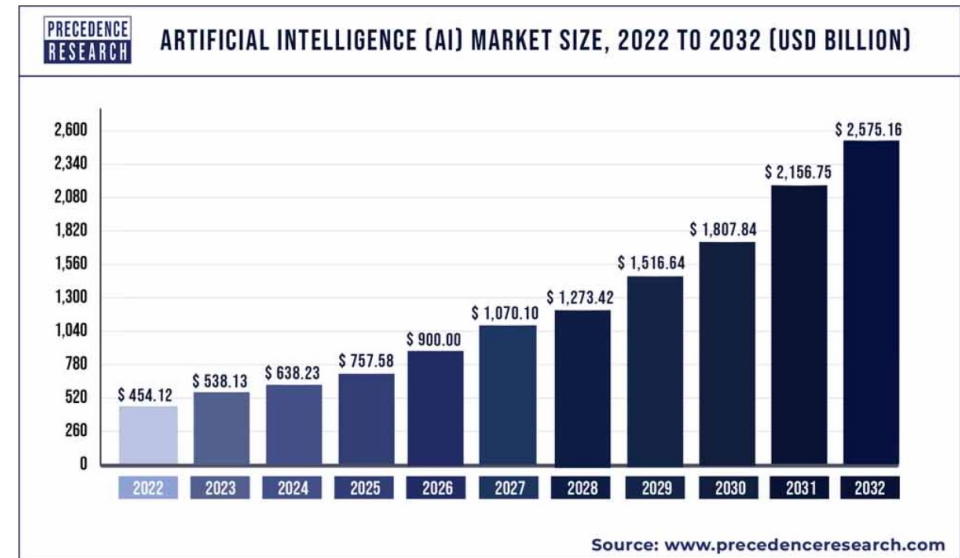


a corgi's head depicted as an explosion of a nebula

AI is having real-world impact

- Public imagination
- Economy
 - 454 billion USD 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



AI is having real-world impact

- Public imagination
- Economy
- Politics



AI is having real-world impact

- Public imagination
- Economy
- Politics



AI is having real-world impact

- Public imagination
- Economy
- Politics



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

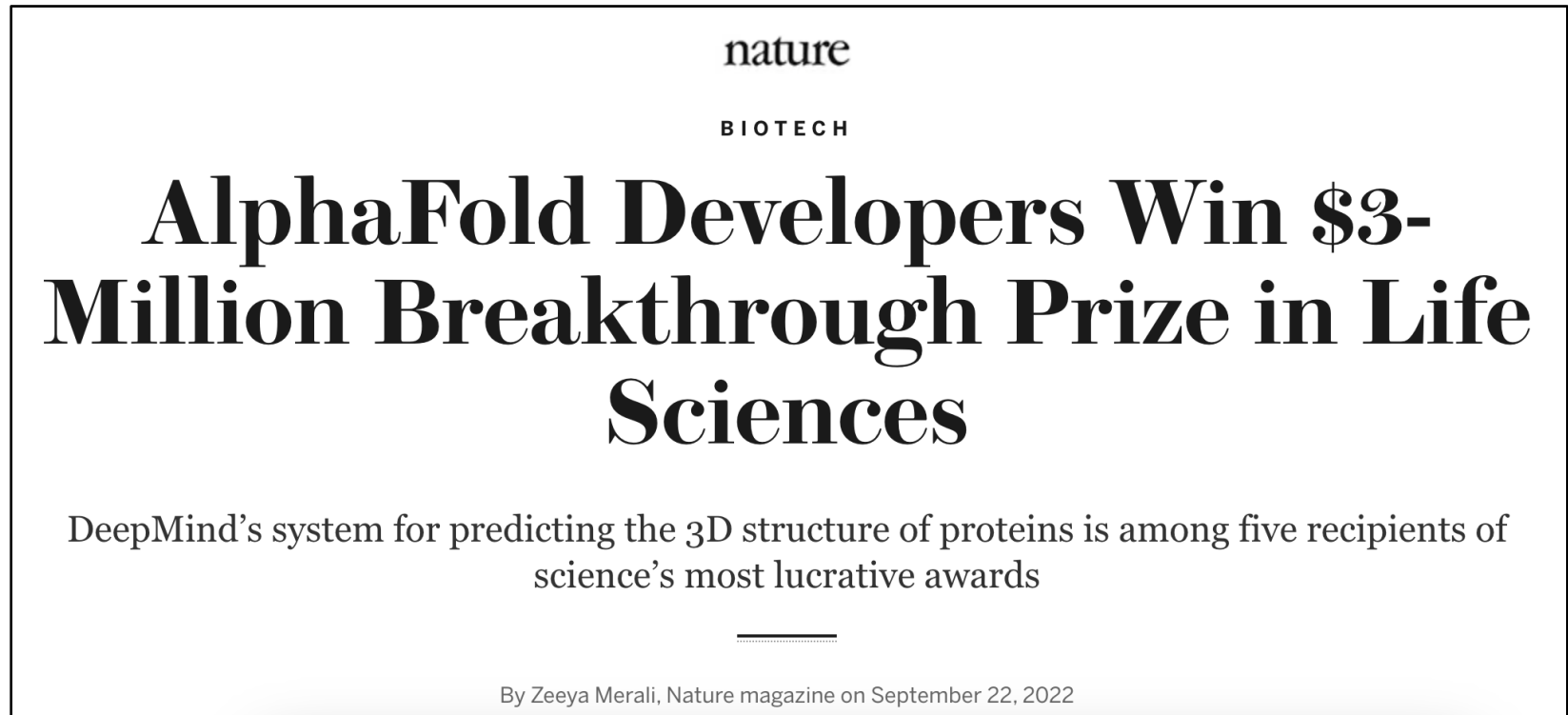
The human labor behind AI chatbots and other smart tools

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, 2022

AI is having real-world impact

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



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

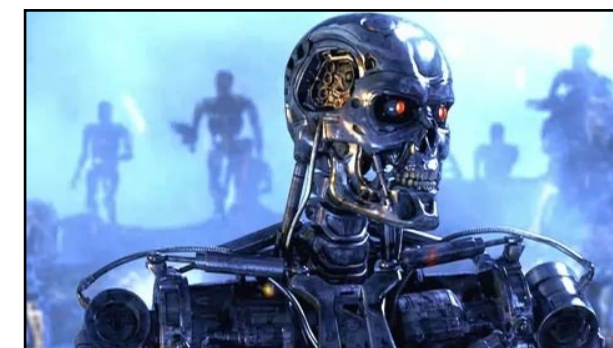
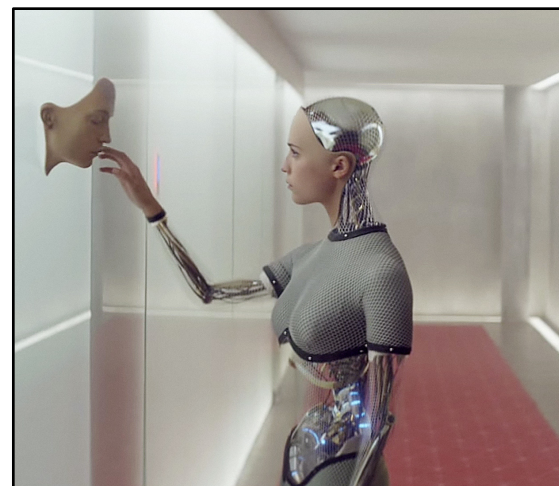
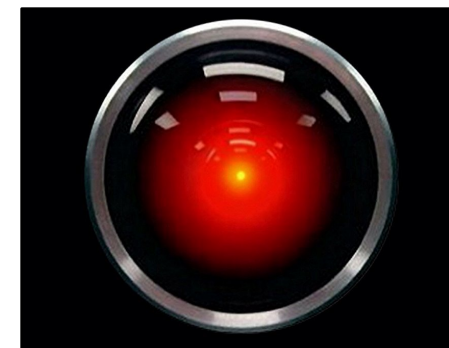
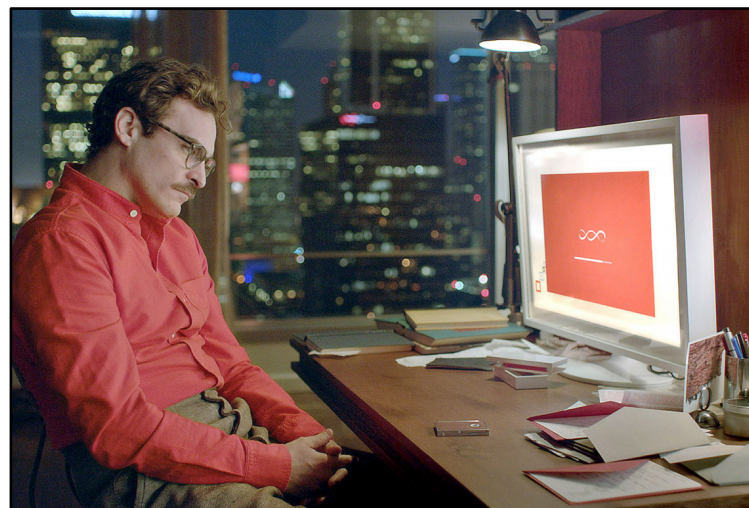
Forbes, 2023

AI is having real-world impact

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

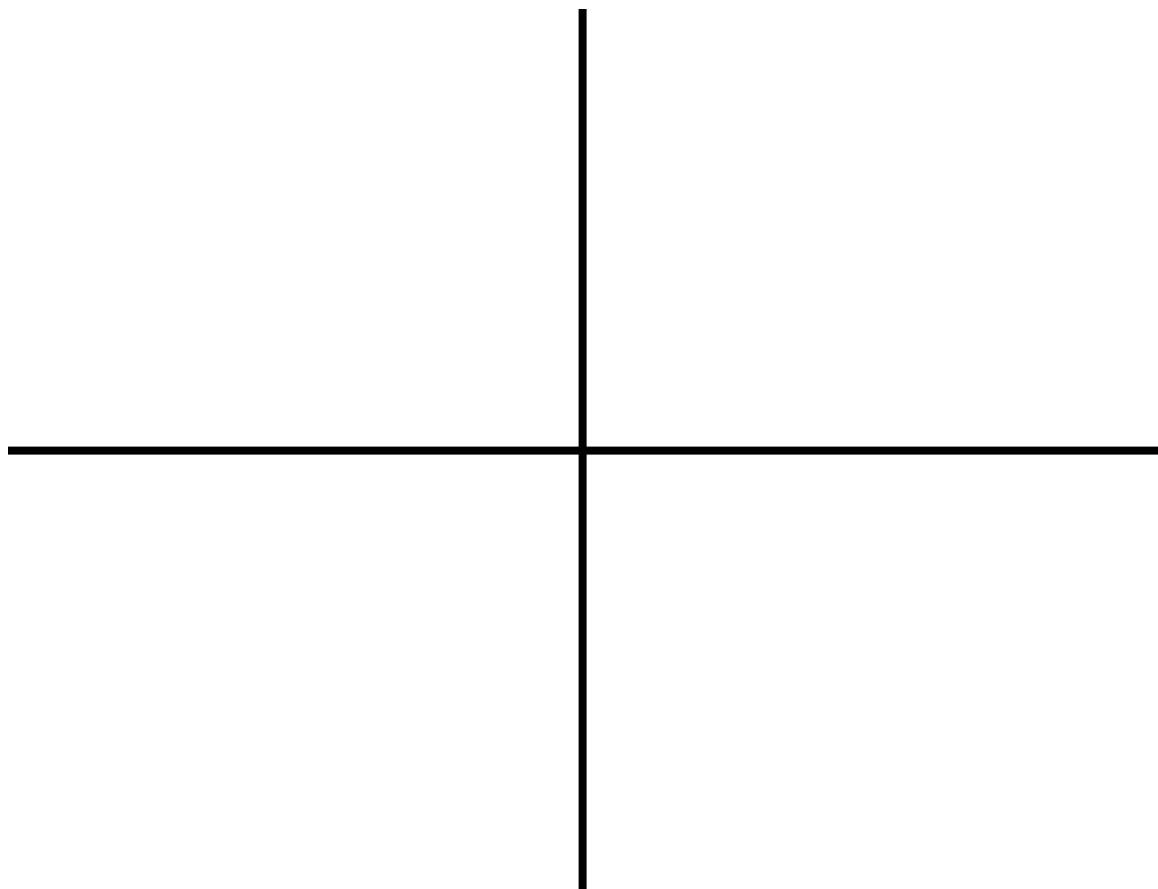
Ok, but what actually is AI???

Science fiction AI?



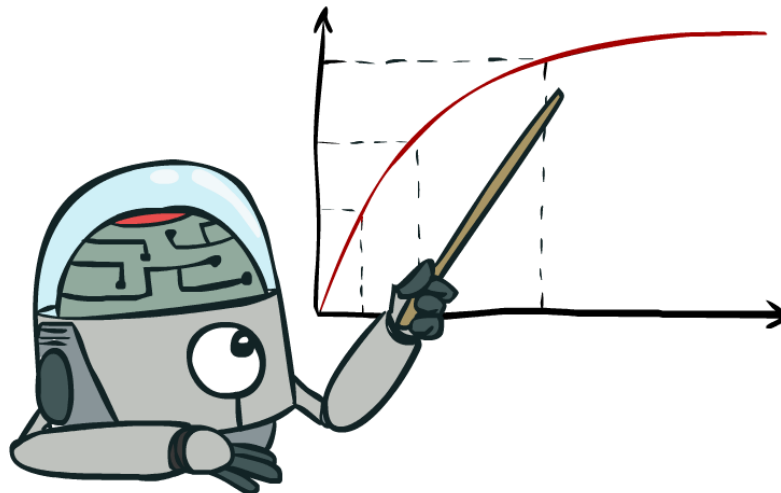
What should we build?

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:

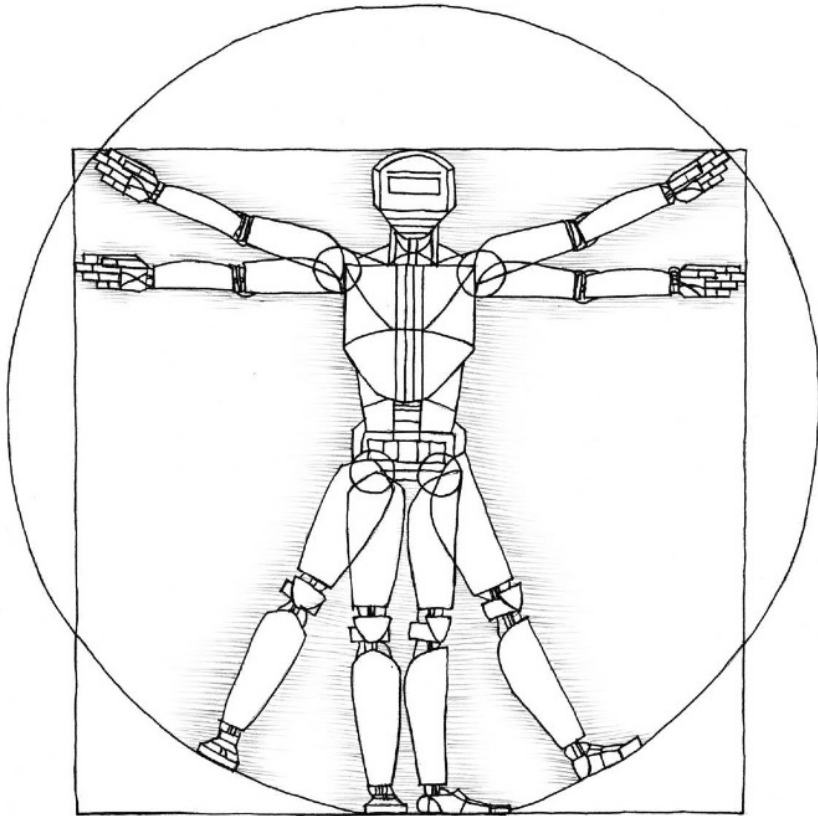
Computational Rationality

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?

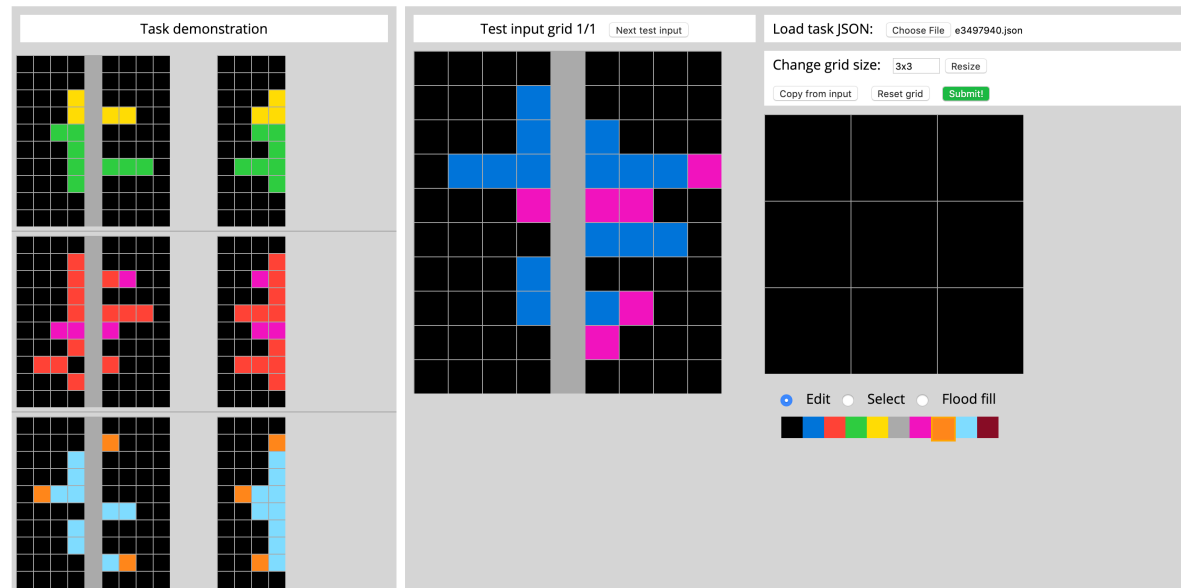
Perspectives on Intelligence

- Embodiment perspective (Rodney Brooks)



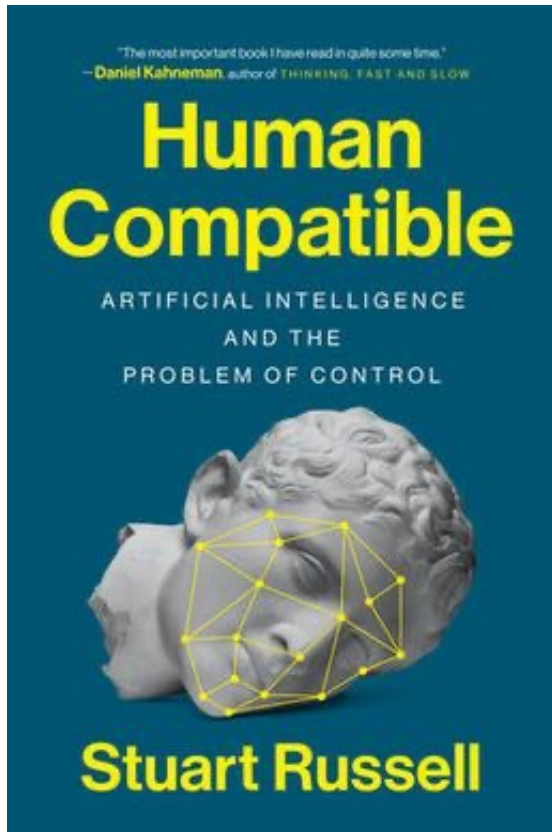
Perspectives on Intelligence

- 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 human utility.
2. Initially uncertain about human preferences.
3. Must learn about preferences from human behavior.

Perspectives on Intelligence

A human being should be able to change a diaper, plan an invasion, butcher a hog, conn a ship, design a building, write a sonnet, 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?

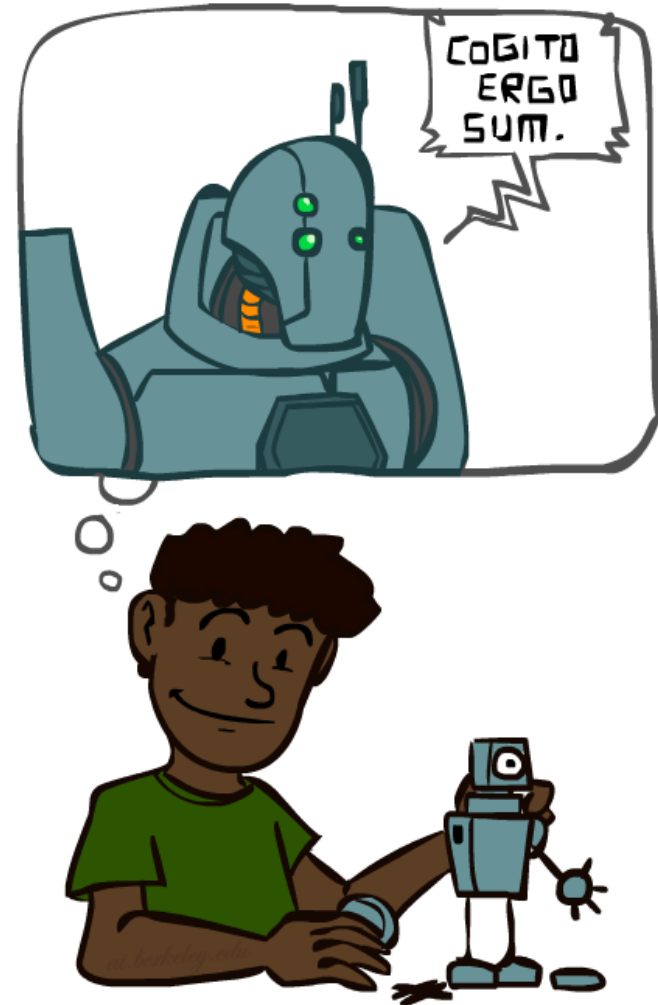
- Brains (human minds) are very good at making rational decisions, but not perfect
- Brains aren't as modular as software, so hard to reverse engineer!
- AI may be better than brains at some tasks
- *“Brains are to intelligence as wings are to flight”*
- We can't yet build AI on the scale of the brain
 - ~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

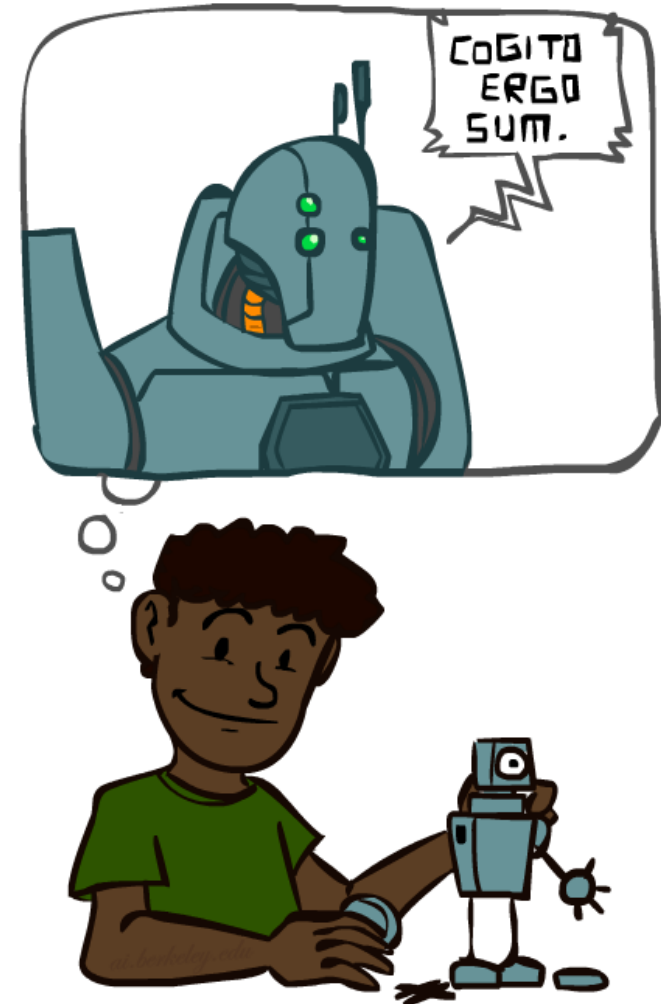
- 1940-1950: Early days: neural and computer science meet
 - 1943: McCulloch & Pitts: Perceptron–boolean circuit model of brain
 - 1950: Turing's "Computing Machinery and Intelligence"
- 1950—70: Excitement! Logic-driven
 - 1950s: Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
 - 1956: Dartmouth meeting: "Artificial Intelligence" adopted

*"We propose that a **2-month, 10-man study of artificial intelligence** be carried out **during the summer of 1956** at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that **every aspect of learning** or **any other feature of intelligence** can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. **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."*



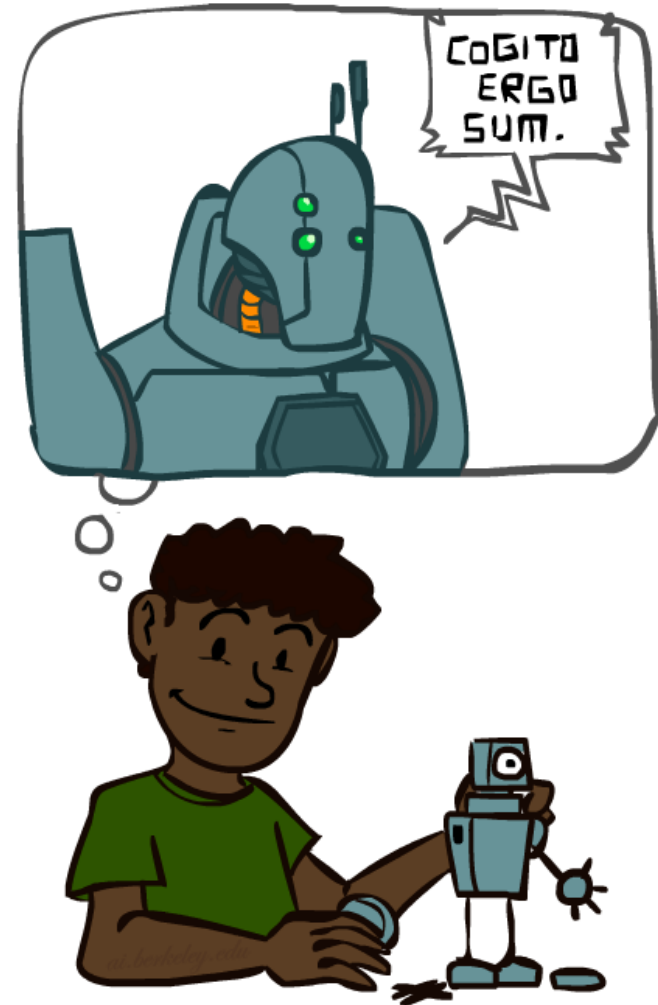
A (Short) History of AI

- **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

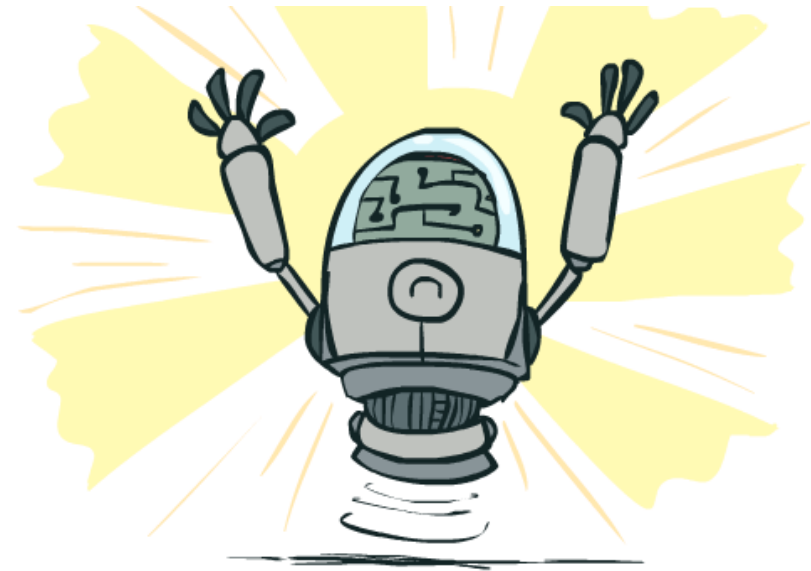
- **2010—2017: Big Data, GPUs, Deep Learning**
 - 2011: Apple releases Siri
 - 2012: AlexNet wins ImageNet competition
 - 2015: DeepMind achieves “human-level” control in Atari games
 - 2016: DeepMind’s AlphaGo defeats Lee Sedol at Go
 - 2016: Google Translate migrates to neural networks
- **2017—: Scaling 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?
- Drive safely along streets of San Francisco?
- Buy a week's worth of groceries on the web?
- Buy a week's worth of groceries at Berkeley Bowl?
- 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?
- Drive safely along streets of San Francisco?
- Buy a week's worth of groceries on the web?
- Buy a week's worth of groceries at Berkeley Bowl?
- 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?
- Drive safely along streets of San Francisco?
- Buy a week's worth of groceries on the web?
- Buy a week's worth of groceries at Berkeley Bowl?
- 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?
- ? Drive safely along streets of San Francisco?
- Buy a week's worth of groceries on the web?
- Buy a week's worth of groceries at Berkeley Bowl?
- 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?

California just opened the floodgates for self-driving cars

After a state regulatory board ruling, San Francisco will have 24/7 robotaxis on its streets. A Washington Post analysis shows how it's a pivotal moment for the industry.

By [Trisha Thadani](#) and [Jeremy B. Merrill](#)

Updated August 10, 2023 at 10:06 p.m. EDT | Published August 10, 2023 at 9:15 p.m. EDT

The Register

California DMV hits brakes on Cruise's SF driverless fleet after series of fender benders

50% chop effective 'immediately' as department investigates traffic and safety issues

[Jude Karabus](#)

Mon 21 Aug 2023 | 14:28 UTC

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?
- ? Drive safely along streets of San Francisco?
- ✓ Buy a week's worth of groceries on the web?
- ✗ Buy a week's worth of groceries at Berkeley Bowl?
- ? 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?
- ? Drive safely along streets of San Francisco?
- ✓ Buy a week's worth of groceries on the web?
- ✗ Buy a week's worth of groceries at Berkeley Bowl?
- ? 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?



You

Please write me a very funny, and extremely short story.

< 6 / 6 >



ChatGPT

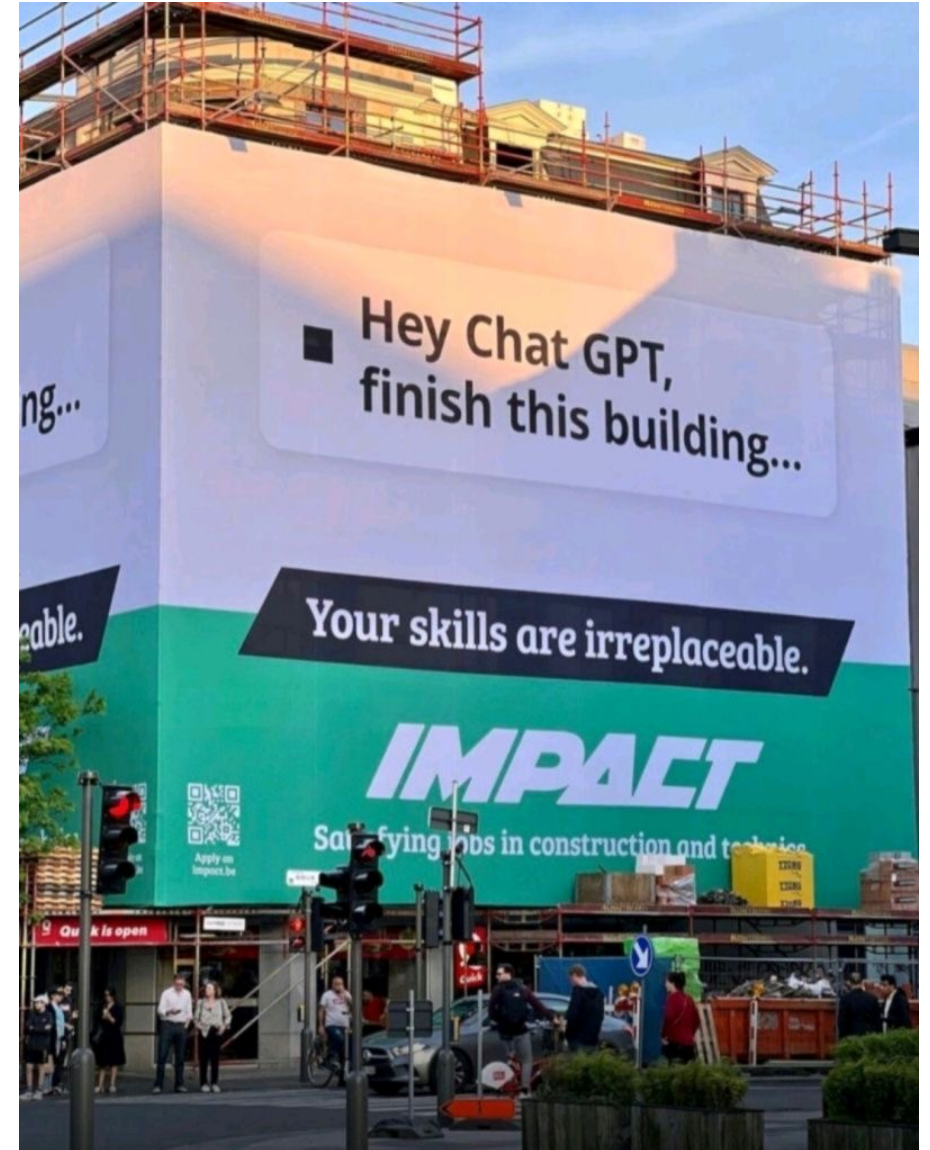
Once there was a chicken who was terribly afraid of crossing roads. One day, it mustered up all its courage, stepped onto the road, and halfway across, it suddenly stopped and said, "Wait, why am I doing this again?"

< 12 / 12 >    

What Can AI Do?

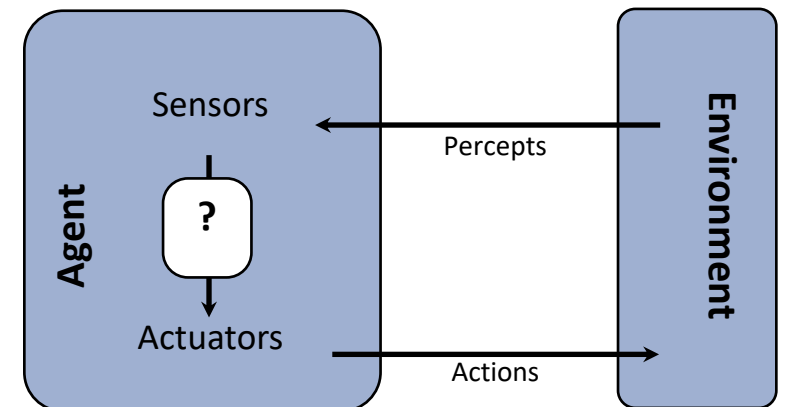
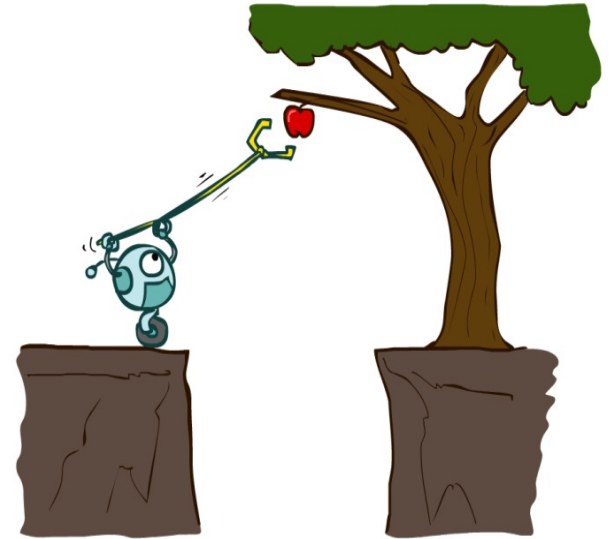
Quiz: Which of the following can be done at present?

- ✓ 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?
- ? Drive safely along streets of San Francisco?
- ✓ Buy a week's worth of groceries on the web?
- ✗ Buy a week's worth of groceries at Berkeley Bowl?
- ? 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?

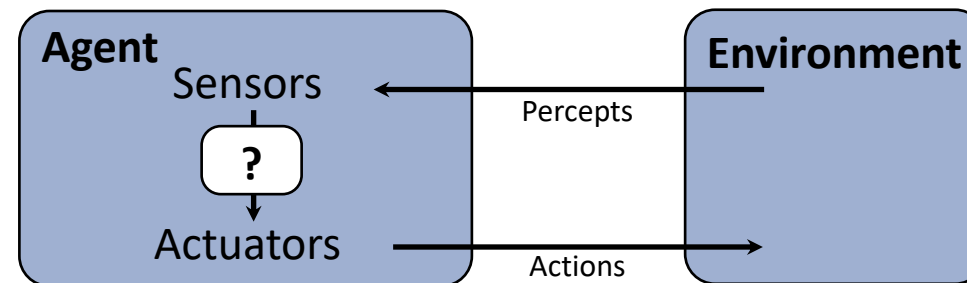
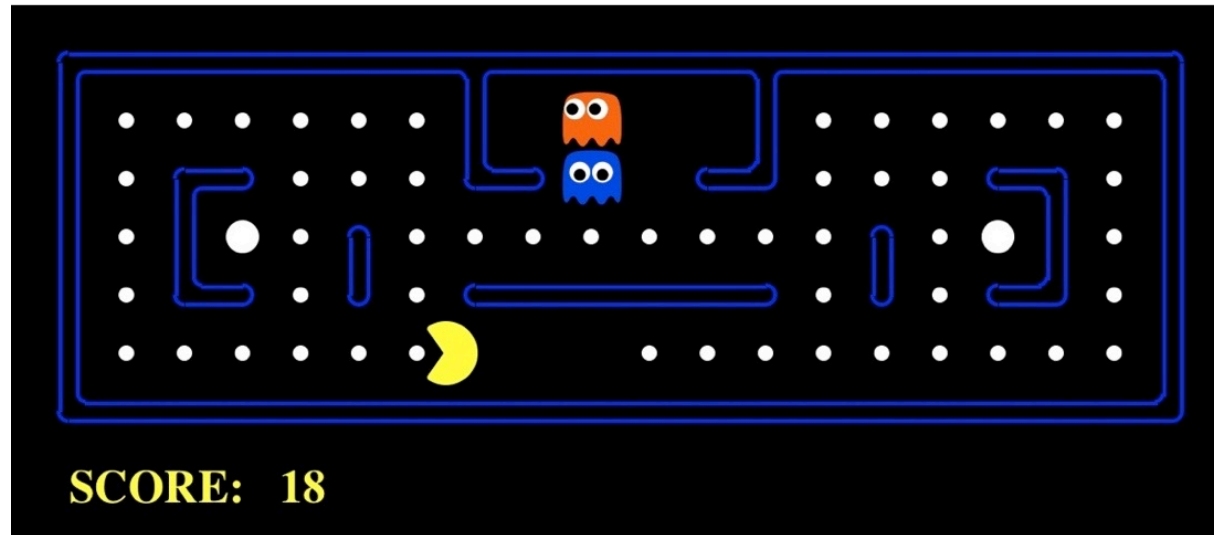


This Course: Designing Rational Agents

- 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





Course Topics

Core Components of Rational Agents:

Search &
Planning

Probability &
Inference

Supervised
Learning

Reinforcement
Learning

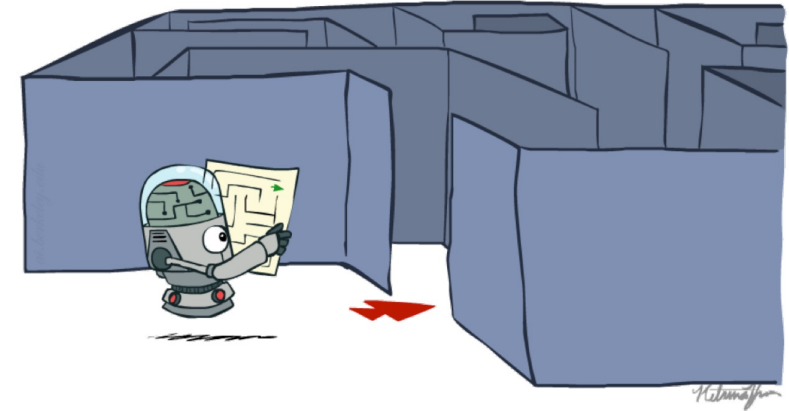
Course Topics

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*?

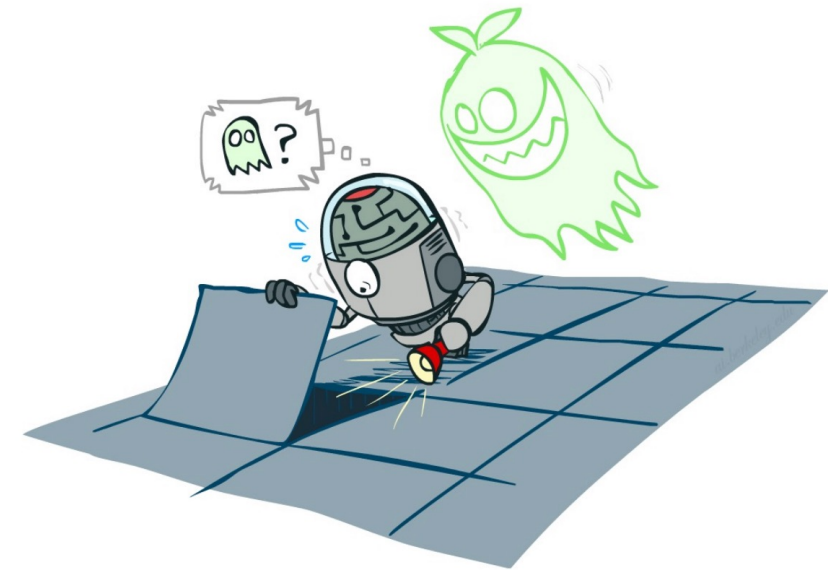
Course Topics

Search &
Planning

Probability &
Inference

Supervised
Learning

Reinforcement
Learning



How can I make sense of *uncertainty*?

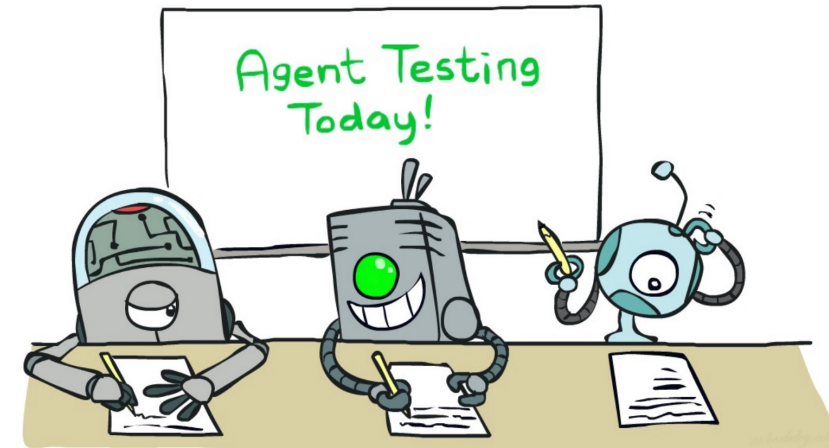
Course Topics

Search &
Planning

Probability &
Inference

Supervised
Learning

Reinforcement
Learning



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

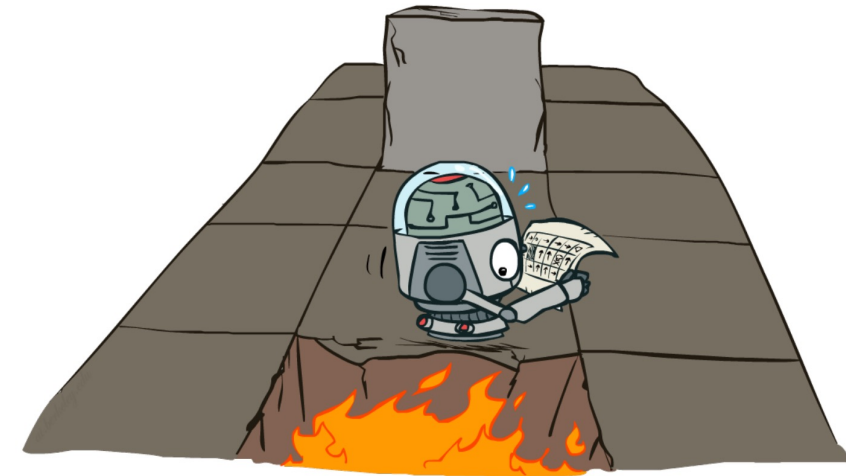
Course Topics

Search &
Planning

Probability &
Inference

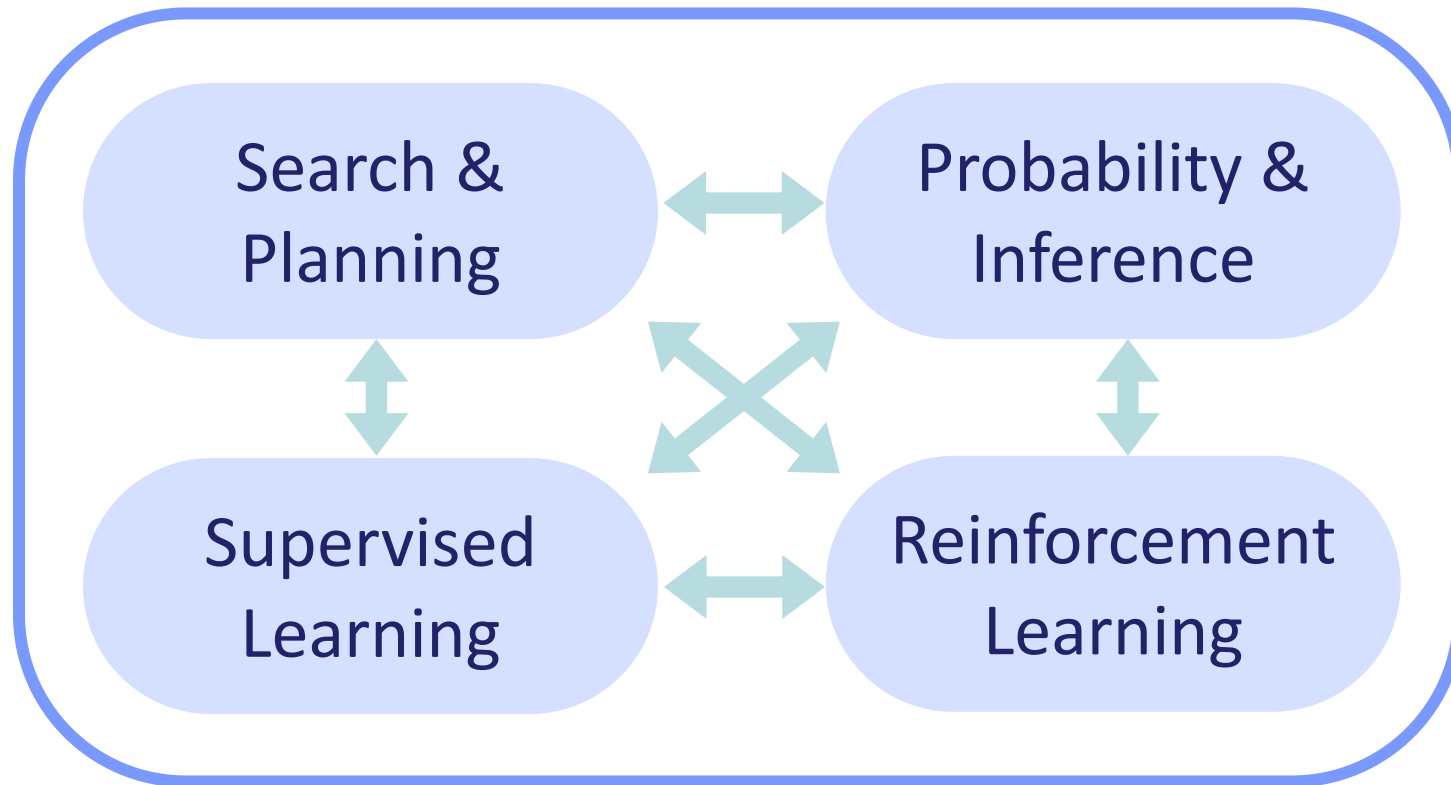
Supervised
Learning

Reinforcement
Learning



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

Course Topics



Course Topics

Search &
Planning

Probability &
Inference

Supervised
Learning

Reinforcement
Learning

Applications

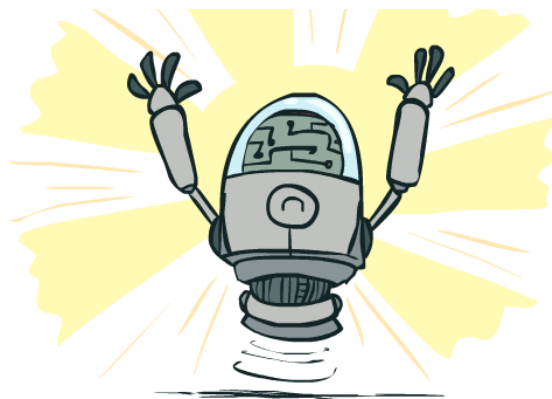
Impact on Sciences, Technology, Society

Should I take CS 188?

- Yes, if you want to know how to design rational agents!
 - CS 188 gives you extra mathematical maturity
 - CS 188 gives you a survey of other non-CS fields that interact with AI (e.g. robotics, cognitive science, economics)
- Disclaimer: If you're interested in making yourself more competitive for AI jobs, CS 189 and CS 182 are better fits.
 - CS 188 will touch on some of the modern tools (like neural networks), but CS 189 and CS 182 cover these in more depth.

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



Next Lecture: Search

