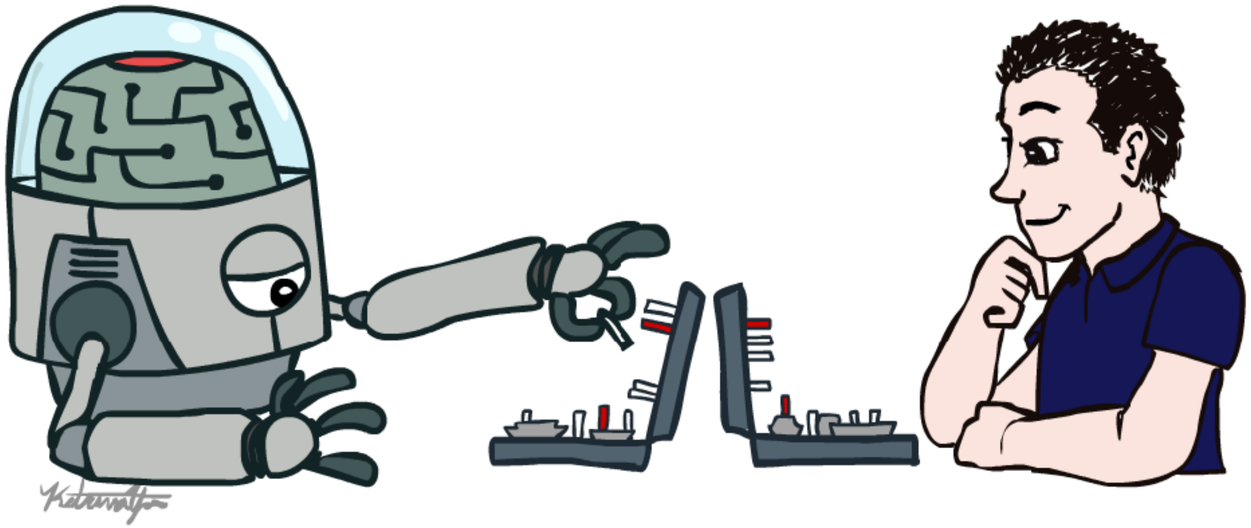
CS383: Artificial Intelligence

Introduction



Prof. Scott Niekum

University of Massachusetts Amherst

[Based on slides created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley.

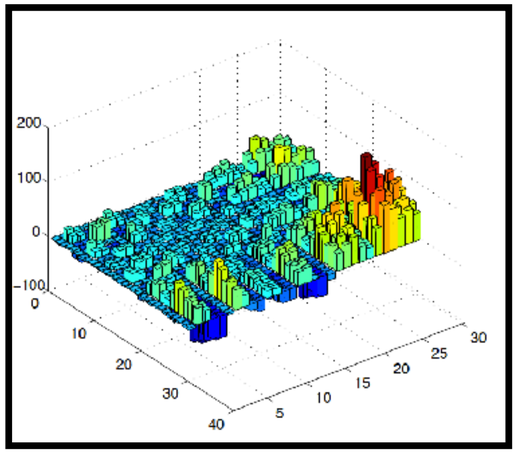
All materials available at http://ai.berkeley.edu.]

A bit about me

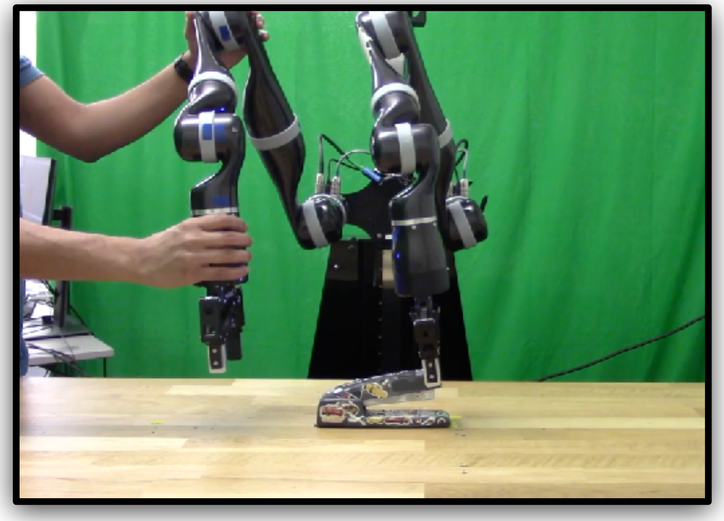
AI Safety

SCALAR LAB

Safe, Confident, and Aligned Learning + Robotics



Reinforcement learning



Robotic manipulation and learning from demonstration

Course Information

- Communication:
  - Announcements on webpage/Piazza
  - Grades on Gradescope
  - Piazza for discussion

- Assignments:
  - Gradescope for interactive homework (unlimited submissions!)
  - Autograded programming projects (submit via Gradescope)
  - Make sure you have access to a system where you can run Python

Class website:

https://people.cs.umass.edu/~sniekum/classes/383-F23/desc.php

(Link available on class Moodle page)

TAs:

Devdhar Patel (devdharpatel@cs.umass.edu)
Erfan Entezami (eentezami@umass.edu)
John Raisbeck (jraisbeck@umass.edu)

TA Office Hours:

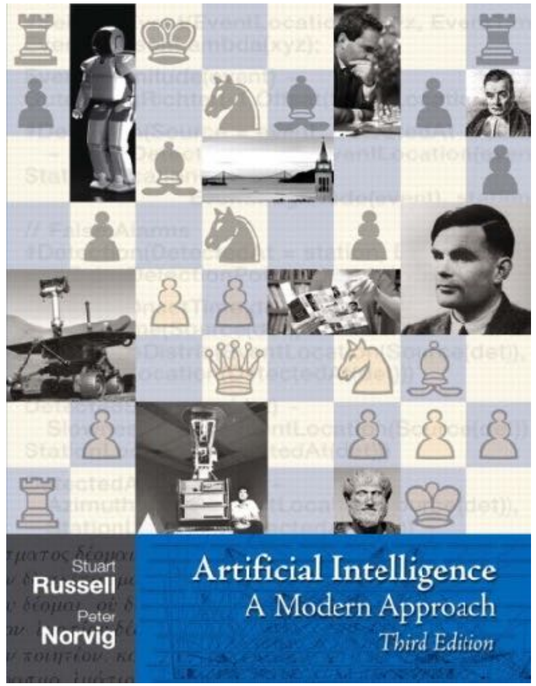
Mon: LGRT 222 4pm-6pm
Tues: LGRT 220 9am-11am
Thurs: CS 207 2pm-4pm
And as-needed

Workload

- There will be a lot of math (and programming)
- Reading assignments
- 8 homework assignments:
  - ~2 weeks for each, but sometimes overlapping
  - Online, autograded, solve and submit alone
  - Can be turned in up to 5 days late for -20%
- 5 programming projects
  - Python, groups of 1 or 2 (except Project 0)
  - ~2 weeks for each, non-overlapping
  - Can be turned late until last day of classes for -20%
- One midterm, one final

Textbook

Russell & Norvig, AI: A Modern Approach, 3rd Ed.



- I'll also post class slides

Warning: Not a course textbook, so our presentation does not necessarily follow the presentation in the book.

Homework Exercises

- Online on Gradescope
- Autograded text boxes / multiple choice
- Try as many times as you want!
- Goal: self-assess and prepare for tests
- Can discuss at high-level, but work alone
- No spoilers on Piazza discussions!

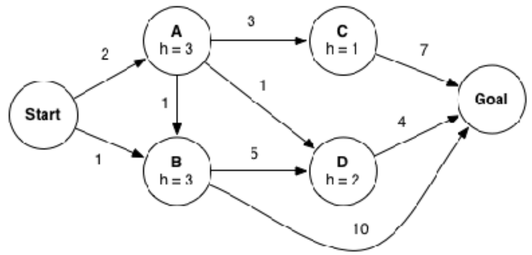
hw1_search_q4_a*_graph_search VIEW UNIT IN STUDIO

Bookmark this page

Q4: A* Graph Search

8.0 points possible (graded)

Consider A* *graph* search on the graph below. Arcs are labeled with action costs and states are labeled with heuristic values. Assume that ties are broken alphabetically (so a partial plan S->X->A would be expanded before S->X->B and S->A->Z would be expanded before S->B->A.

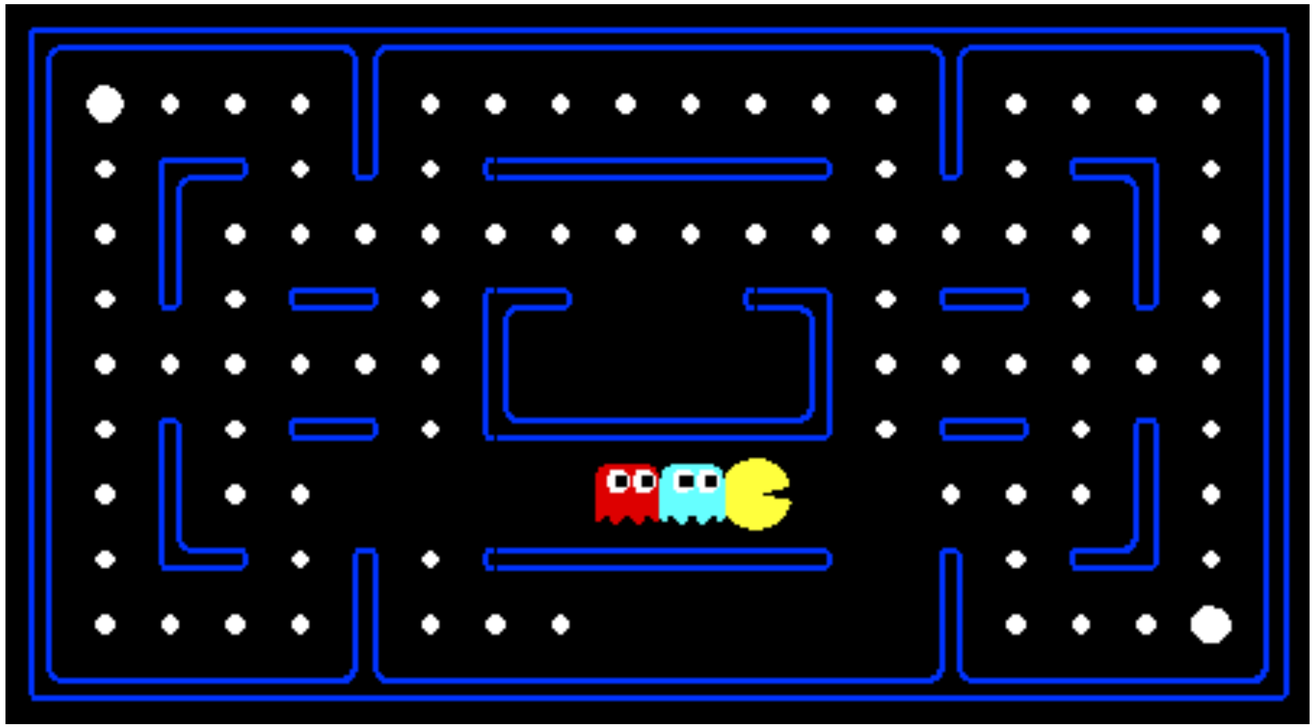


In what order are states expanded by A* graph search? You may find it helpful to execute the search on scratch paper.

- Start, A, B, C, D, Goal
- Start, A, C, Goal
- Start, B, A, D, C, Goal
- Start, A, D, Goal
- Start, A, B, Goal
- Start, B, A, D, B, C, Goal

Programming Assignments

Pacman domain



Projects include:

- path planning and search
- multi-agent game trees
- reinforcement learning
- state estimation

Highly suggested: Pair programming
(switch “driver” and “observer” roles often)

Midterm and Final

- Midterm will cover roughly half the class material
- Final will be comprehensive
- Midterm in-class, Final during finals week
- Very similar to Gradescope homework questions

Attendance

- There will be daily in class quizzes given via iClicker
- Not graded — only participation is counted
- You must get an iClicker Model 2 and register it on Moodle
- Instructions: https://www.umass.edu/it/audience-response-system

Grading

Plus/minus grading:

93-100: A
90-93: A-
87-90: B+
83-87: B
80-83: B-
77-80: C+
73-77: C
70-73: C-
67-70: D+
63-67: D
60-63: D-
<60: F

Grades will be weighted as follows:

- In-class quiz participation (10%)
- Gradescope Exercises (20%)
- Programming Assignments (30%)
- Midterm (15%)
- Final (25%)

Academic Honesty

READ THE STATEMENT IN THE SYLLABUS

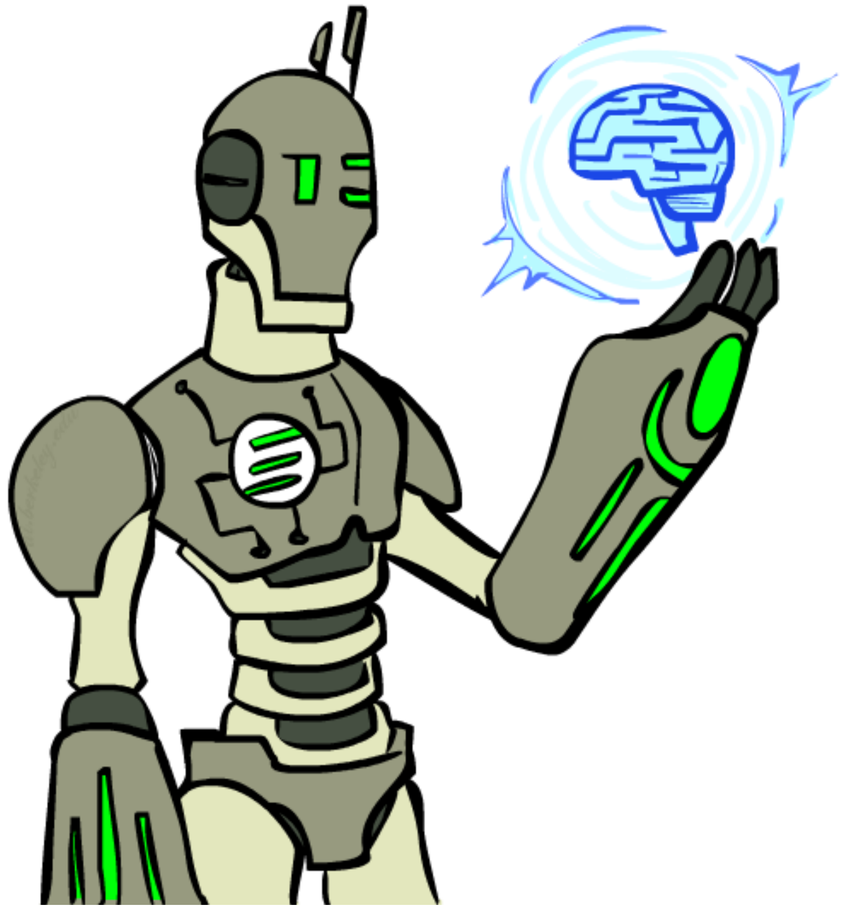
- Discuss concepts, but don’t share solutions or written work with other students
- Don’t look for answers / code online or elsewhere
- Automated tools will be used to discover cheating
- If unsure, check departmental guidelines or ask — ignorance is not an excuse
- We will pursue the harshest penalties available, so please don’t cheat!
- To be clear: you will fail the class automatically and be reported to the university

Important This Week

- Important this week:
  - **Make sure** that you can get into Gradescope — create an account and use the course code that you were emailed (and which is also posted on Piazza).
  - **Be sure** that you have a computer where you can run Python
  - **P0: Python tutorial** is out (due 9/11 at 11:59 pm via Gradescope)
- Also important:
  - **If you are wait-listed,** you might or might not get in depending on how many students drop. Be patient if possible — many students often drop early in the course.
  - **Office Hours** begin Monday

Today

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



A definition for AI

A definition for AI

"Artificial Intelligence (AI) is a science and a set of computational technologies that are inspired by — but typically operate quite differently from — the ways people use their nervous systems and bodies to sense, learn, reason, and take action."

Philosophical questions

- AI is one of the great intellectual adventures of the 20th and 21st centuries.

  - What is intelligence?
  - Can a computer have a mind?
  - Can we build a mind?
  - Can trying to build one teach us what a mind is?

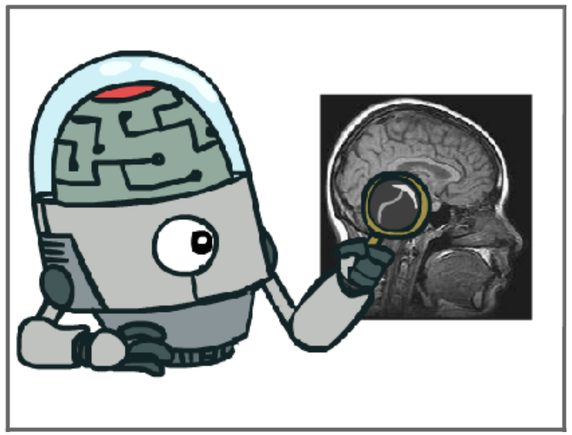
What is AI?

The science of making machines that:

What is AI?

The science of making machines that:

Think like people

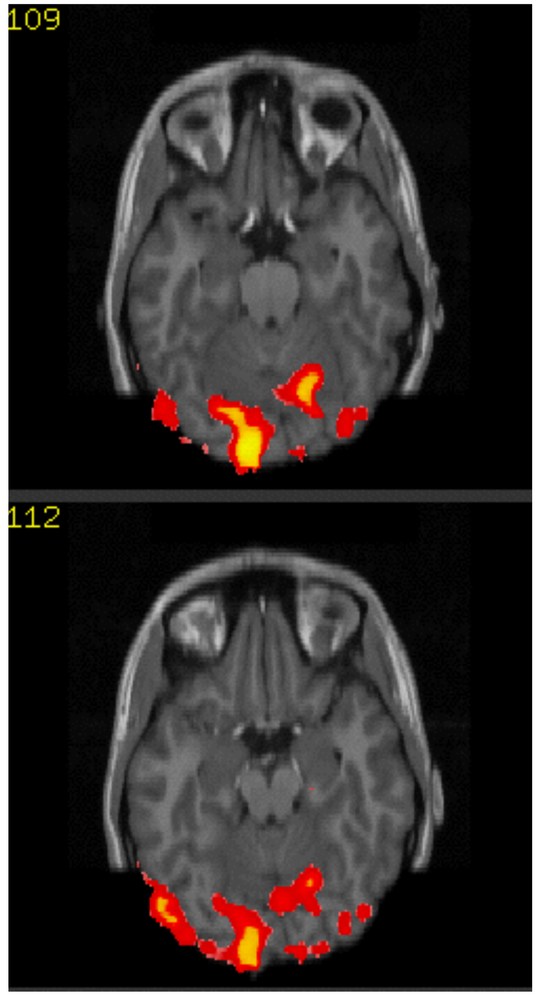


Thinking Like Humans?

- The cognitive science approach:
  - 1960s ``cognitive revolution'': information-processing psychology replaced prevailing orthodoxy of behaviorism (reflexive behaviors, classical conditioning, etc.)

- Scientific theories of internal activities of the brain
  - What level of abstraction? “Knowledge'' or “circuits”?
  - **Cognitive science:** Predicting and testing behavior of human subjects (top-down)
  - **Cognitive neuroscience:** Direct identification from neurological data (bottom-up)
  - Both approaches now distinct from AI
  - *The available theories do not yet come close to explaining human intelligence and thinking*

- Even if possible, what is the utility of having AI think like humans?

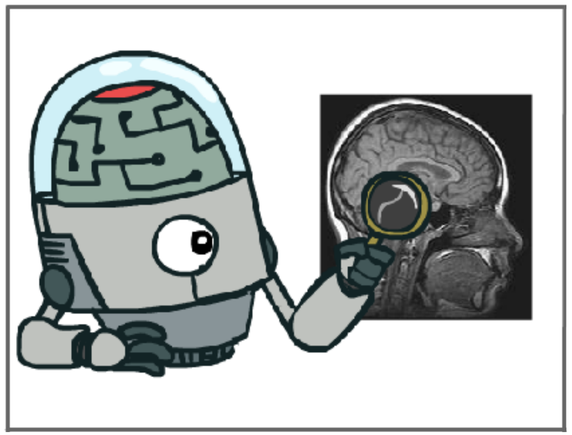


Images from Oxford fMRI center

What is AI?

The science of making machines that:

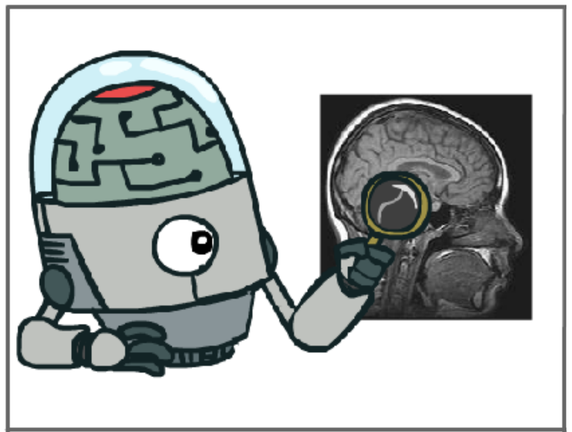
Think like people



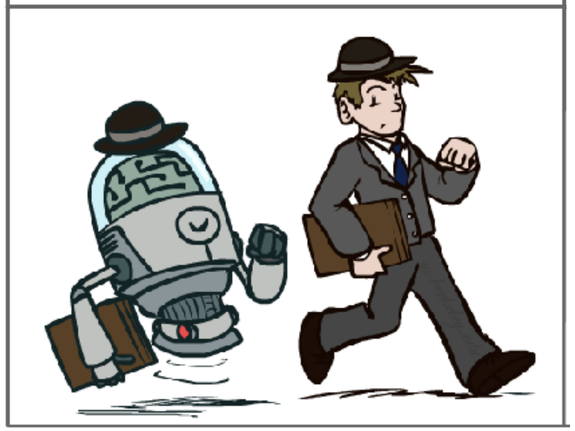
What is AI?

The science of making machines that:

Think like people

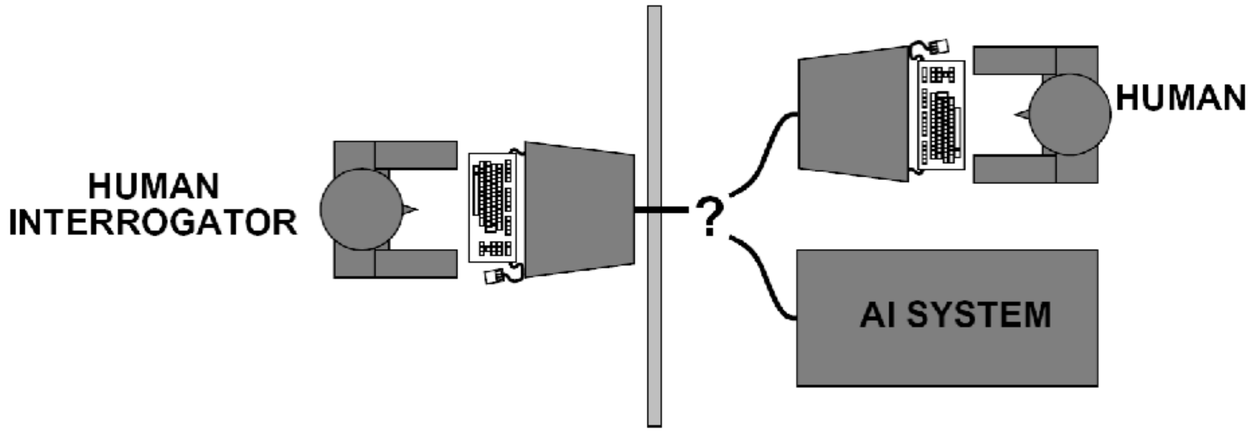


Act like people



Acting Like Humans?

- Turing (1950) “Computing machinery and intelligence”
  - “Can machines think?” → “Can machines behave intelligently?”
  - Operational test for intelligent behavior: the *Imitation Game*

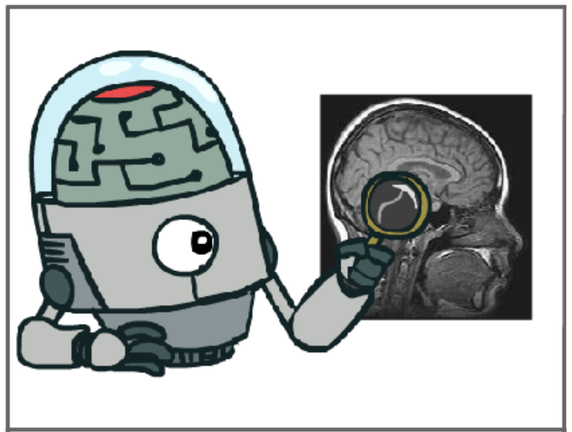


  - Predicted by 2000, a 30% chance of fooling a lay person for 5 minutes
  - Anticipated all major arguments against AI in following 50 years
  - Suggested major components of AI: knowledge, reasoning, language understanding, learning
- Problem: Does the Turing test really measure what we want?

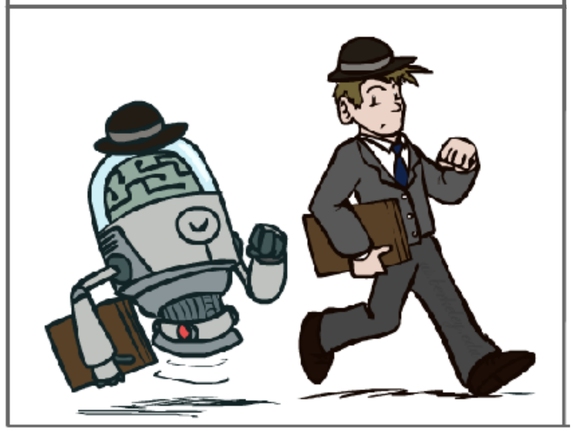
What is AI?

The science of making machines that:

Think like people



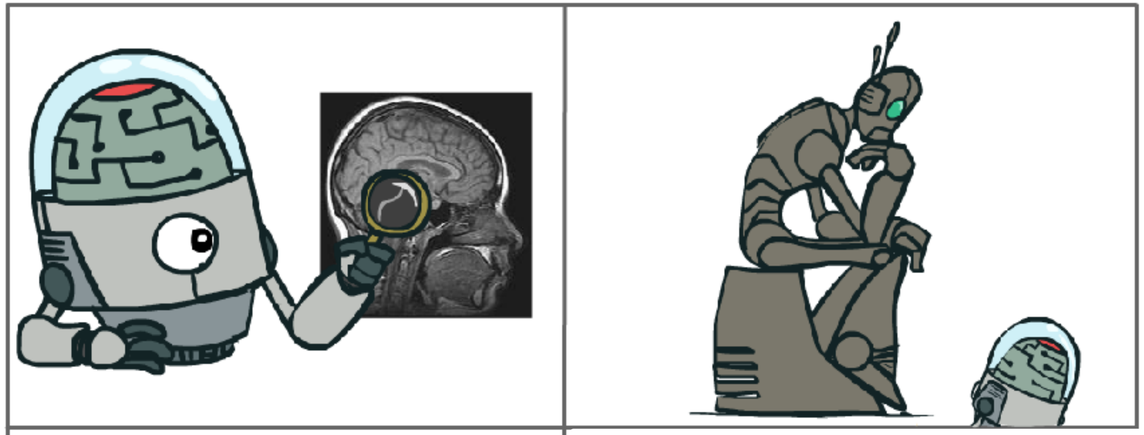
Act like people



What is AI?

The science of making machines that:

Think like people



Think rationally

Act like people

Thinking Rationally?

- The “Laws of Thought” approach
  - What does it mean to “think rationally”?
  - Normative / prescriptive rather than descriptive

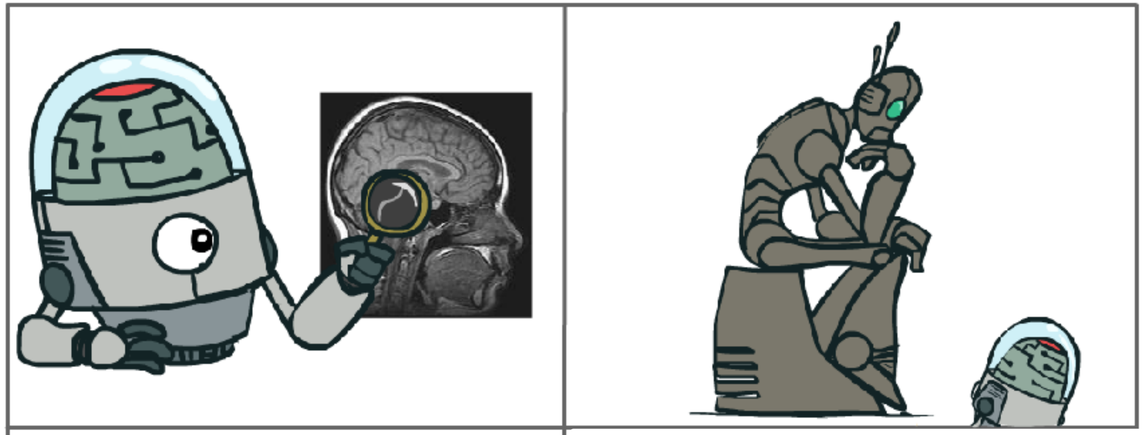
- Logicist tradition:
  - Logic: notation and rules of derivation for thoughts
  - Aristotle: what are correct arguments/thought processes?
  - Direct line through mathematics, philosophy, to modern AI

- Problems:
  - Not all intelligent behavior is mediated by logical deliberation
  - What is the purpose of thinking? What thoughts should I (bother to) have?
  - Logical systems tend to do the wrong thing in the presence of **uncertainty**
  - Why should we care about thought at all, when **action** is what matters?

What is AI?

The science of making machines that:

Think like people



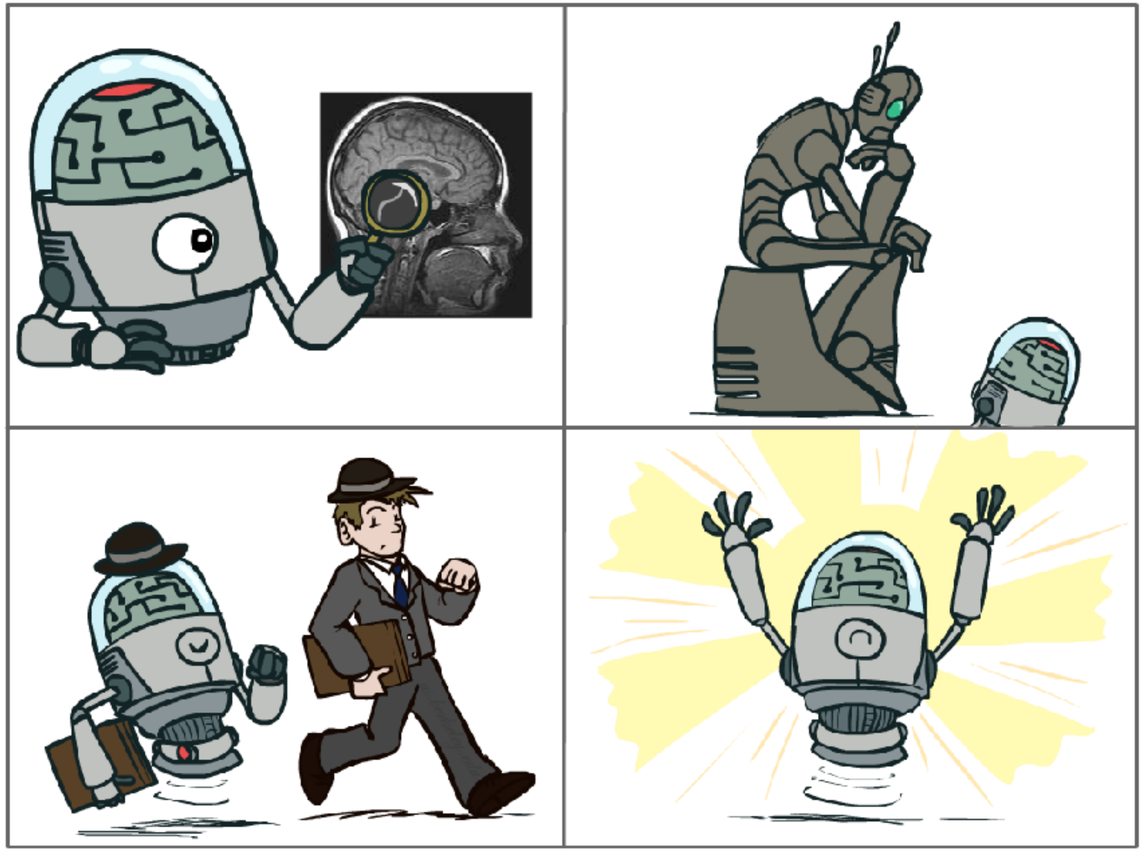
Think rationally

Act like people

What is AI?

The science of making machines that:

Think like people



Think rationally

Act like people

Act rationally

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
  - Thinking can be in the service of rational action
  - Entirely dependent on goals!
  - Irrational ≠ insane, irrationality is sub-optimal action
  - Rational ≠ successful

- Our focus here: rational agents
  - Systems which make the best possible decisions given goals, evidence, and constraints
  - In the real world, usually lots of uncertainty
    - … and lots of complexity
  - Usually, we’re just approximating rationality

Rational Decisions

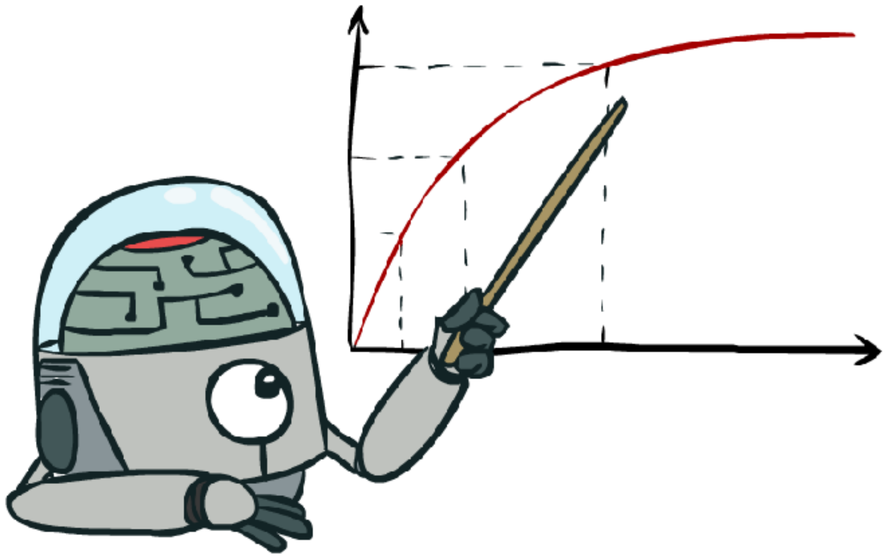
We'll use the term **rational** in a very specific, technical way:

- 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**

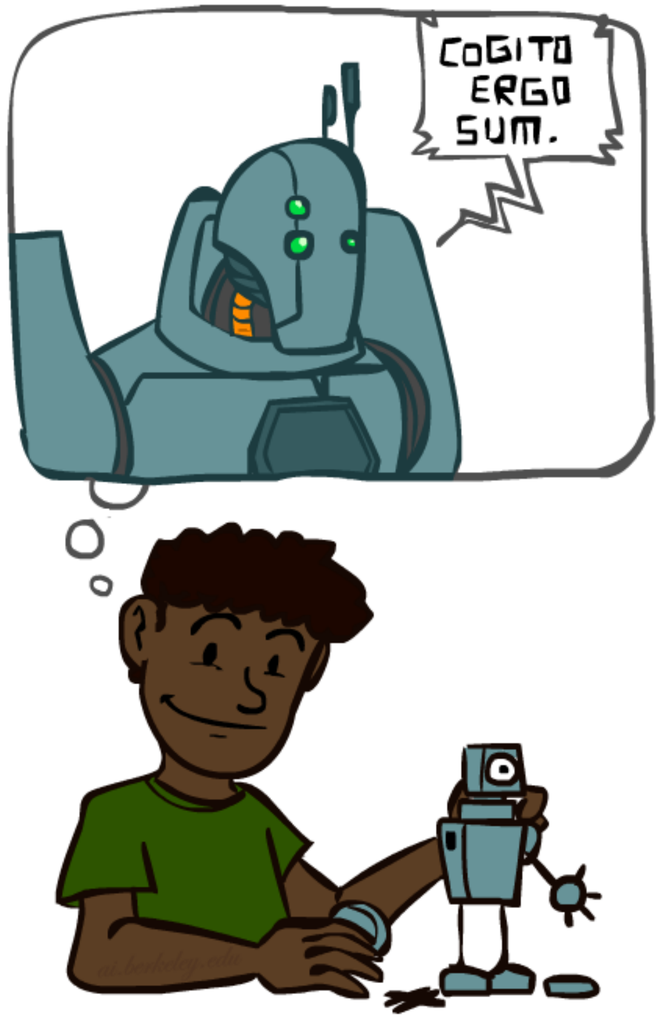
A better title for this course would be:

Computational Rationality

Maximize Your Expected Utility

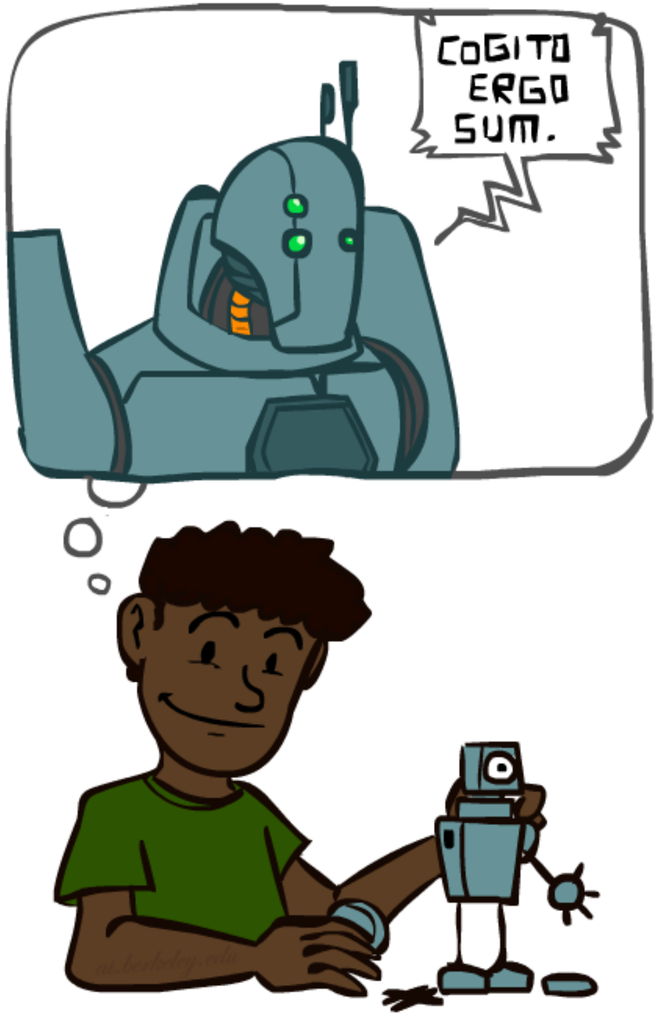


A (Short) History of AI

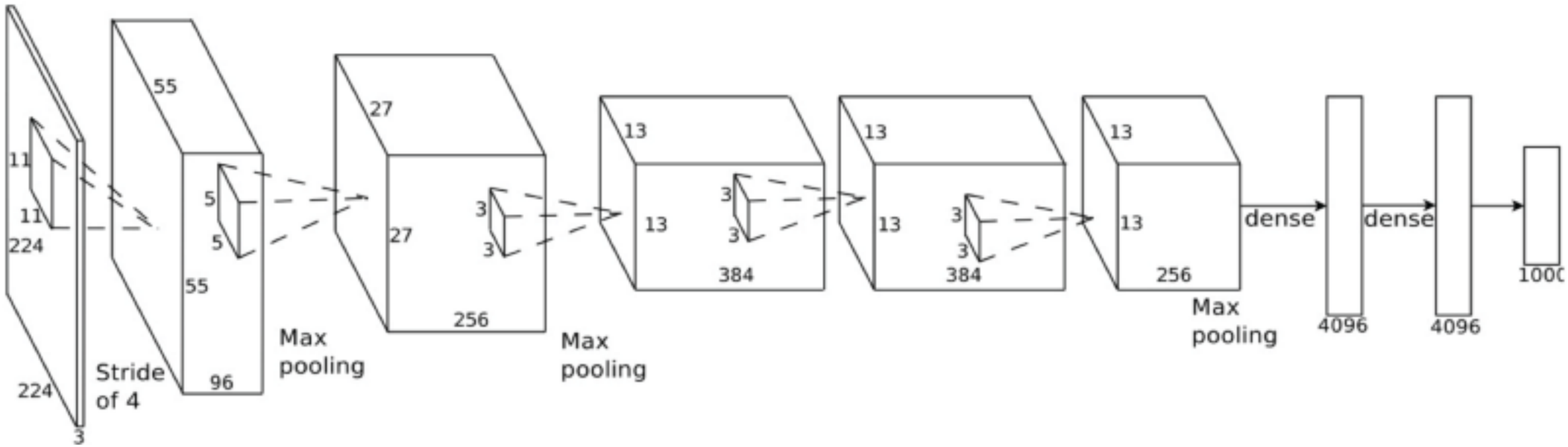


A (Short) History of AI

- 1940-1950: Early days
  - 1943: McCulloch & Pitts: Boolean circuit model of brain
  - 1950: Turing's “Computing Machinery and Intelligence”
- 1950—70: Excitement: Look, Ma, no hands!
  - 1950s: Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
  - 1956: Dartmouth meeting: “Artificial Intelligence” adopted
  - 1965: Robinson's complete algorithm for logical reasoning
- 1970—90: Knowledge-based approaches
  - 1969—79: Early development of knowledge-based systems
  - 1980—88: Expert systems industry booms
  - 1988—93: Expert systems industry busts: “AI Winter”
- 1990—: Statistical approaches
  - Resurgence of probability, focus on uncertainty
  - General increase in technical depth
  - Agents and learning systems… “AI Spring”?
- 2000—: Where are we now?



Deep learning



Natural Language

- Speech technologies (e.g. Siri)
  - Automatic speech recognition (ASR)
  - Text-to-speech synthesis (TTS)

- Language processing technologies
  - **ChatGPT**
  - Machine translation

"Il est impossible aux journalistes de rentrer dans les régions tibétaines"

Bruno Philip, correspondant du "Monde" en Chine, estime que les journalistes de l'AFP qui ont été expulsés de la province tibétaine du Qinghai "n'étaient pas dans l'illégalité".

Les faits Le dalaï-lama dénonce l'"enfer" imposé au Tibet depuis sa fuite, en 1959

Vidéo Anniversaire de la rébellion

"It is impossible for journalists to enter Tibetan areas"

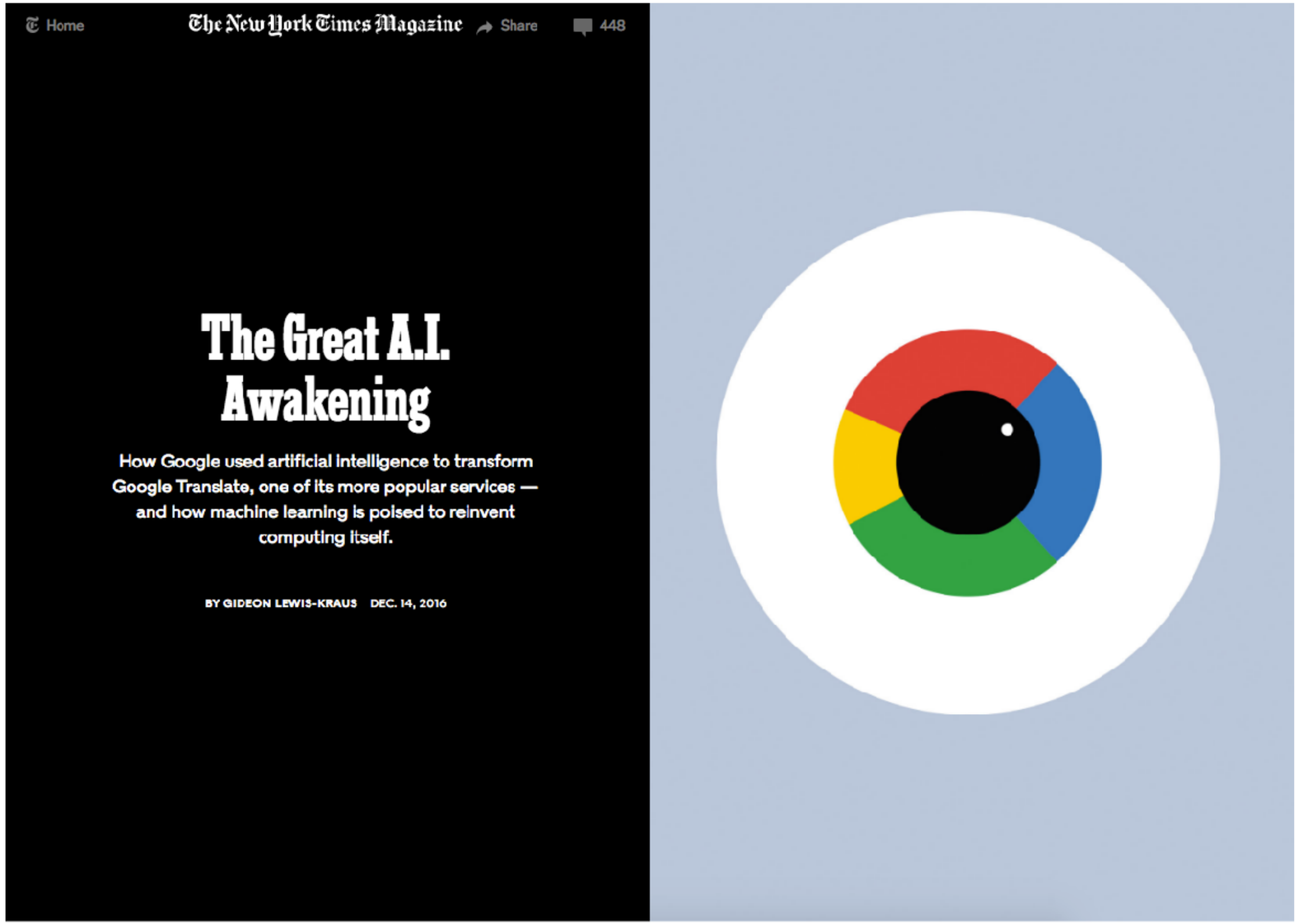
Philip Bruno, correspondent for "World" in China, said that journalists of the AFP who have been deported from the Tibetan province of Qinghai "were not illegal."

Facts The Dalai Lama denounces the "hell" imposed since he fled Tibet in 1959

Video Anniversary of the Tibetan rebellion: China on guard

  - Web search
  - Text classification, spam filtering, etc…

Natural Language



Perception + Natural Language

What vegetable is on the plate?
Neural Net: broccoli
Ground Truth: broccoli

What color are the shoes on the person's feet ?
Neural Net: brown
Ground Truth: brown

How many school busses are there?
Neural Net: 2
Ground Truth: 2

What sport is this?
Neural Net: baseball
Ground Truth: baseball

What is on top of the refrigerator?
Neural Net: magnets
Ground Truth: cereal

What uniform is she wearing?
Neural Net: shorts
Ground Truth: girl scout

What is the table number?
Neural Net: 4
Ground Truth:40

What are people sitting under in the back?
Neural Net: bench
Ground Truth: tent

Face generation

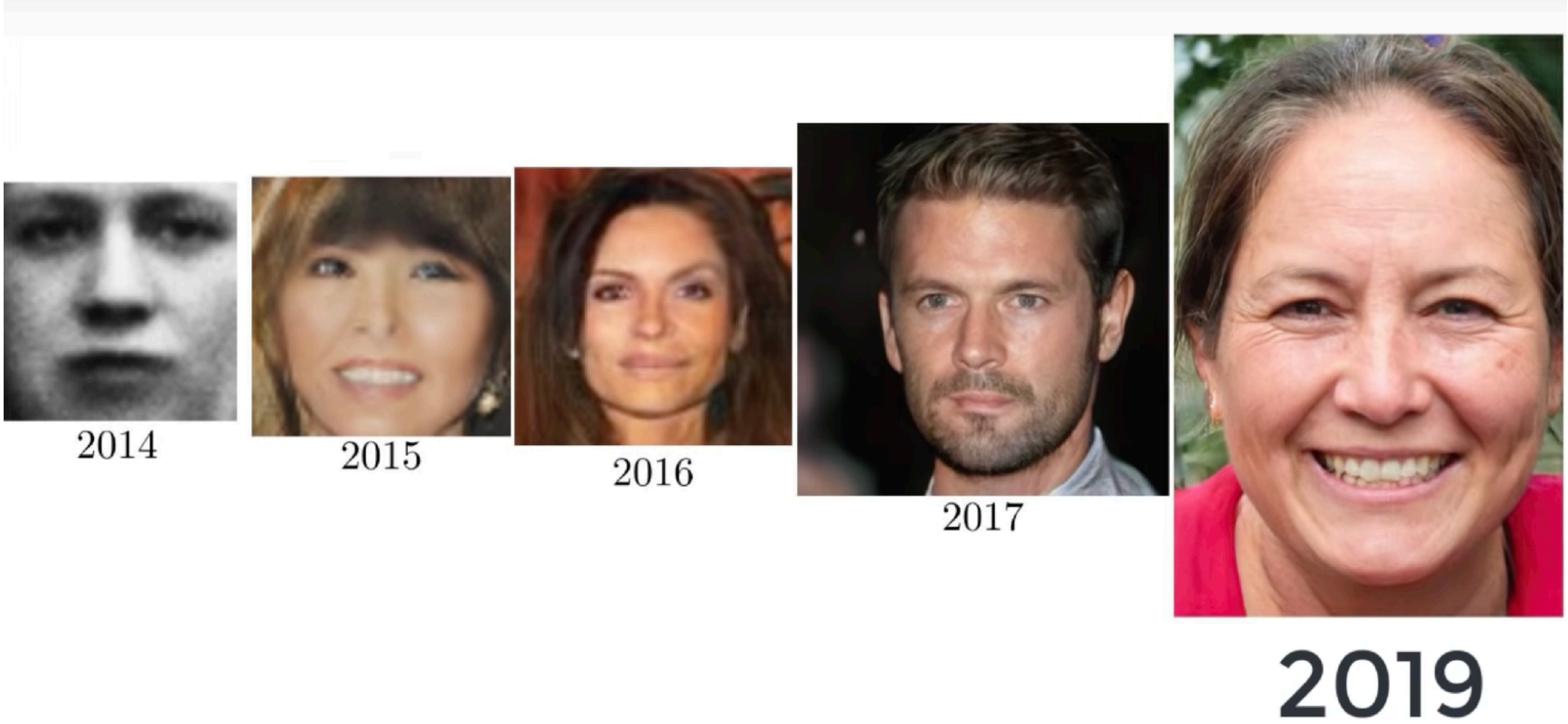


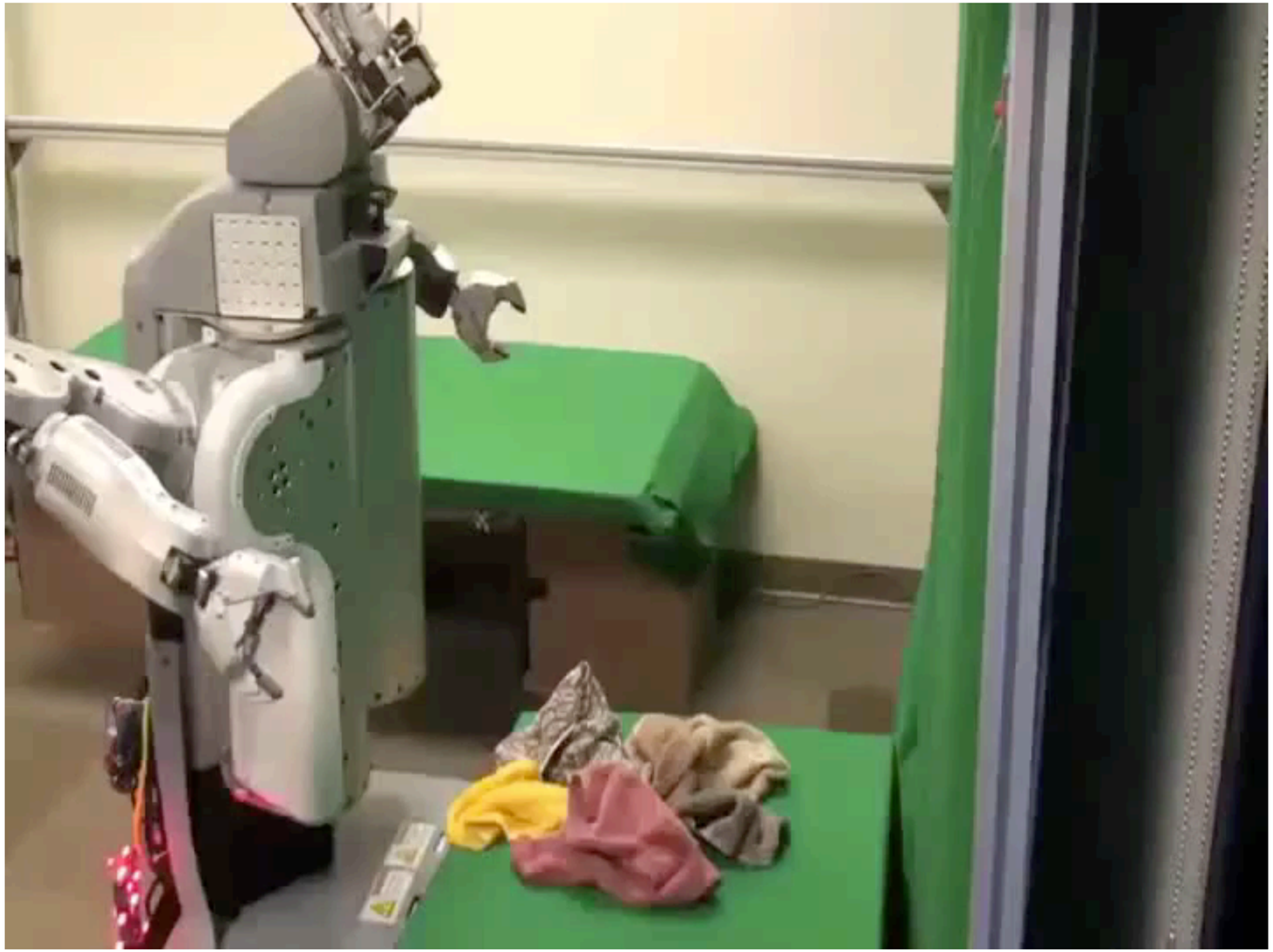
Image and video generation (2023)

Perception, Natural Language, Generative AI

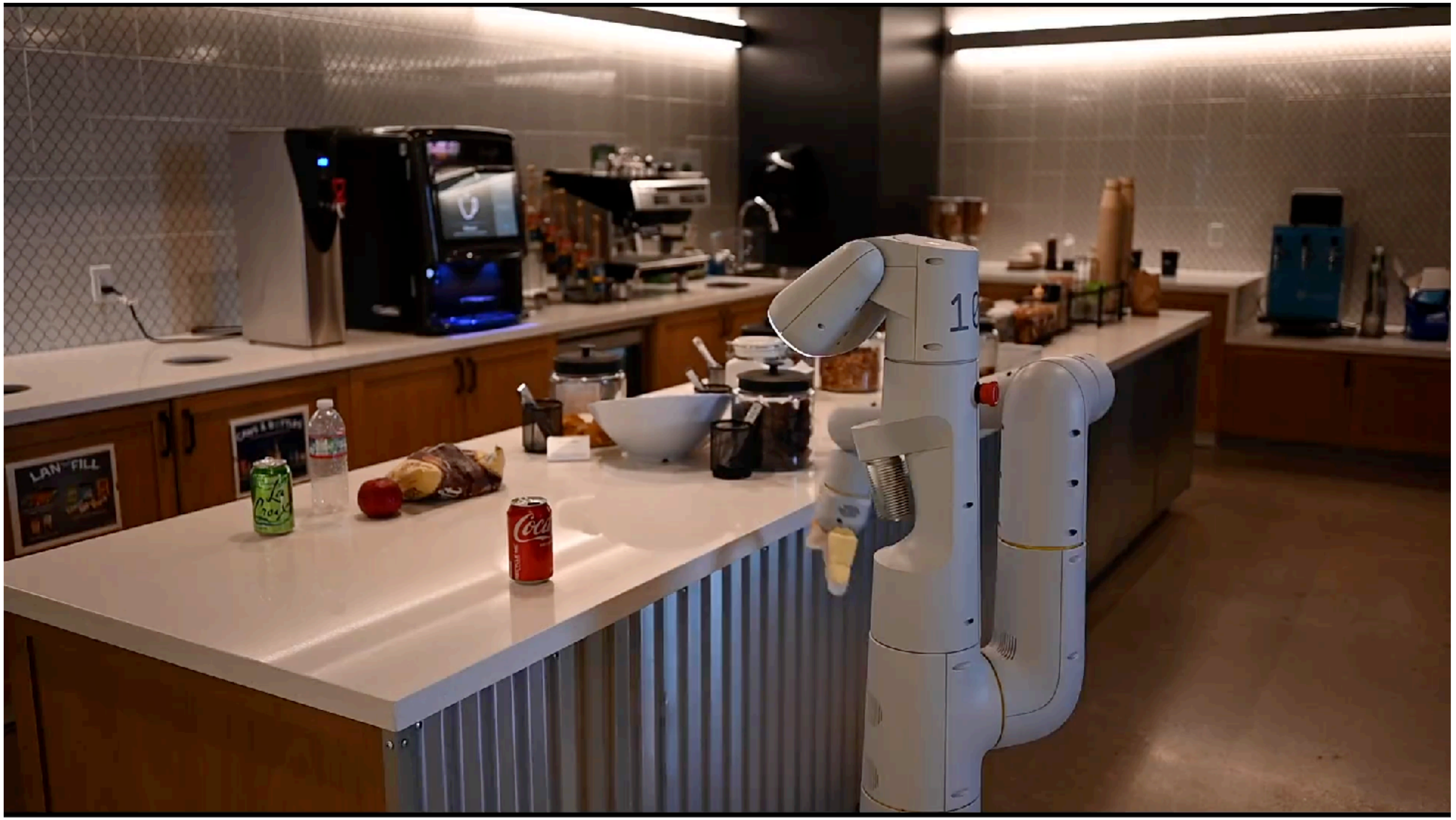
We won’t discuss NLP, perception, and generative AI directly, but we will cover the foundations that are useful for understanding them:

- Bayes nets
- Supervised learning
- Deep learning

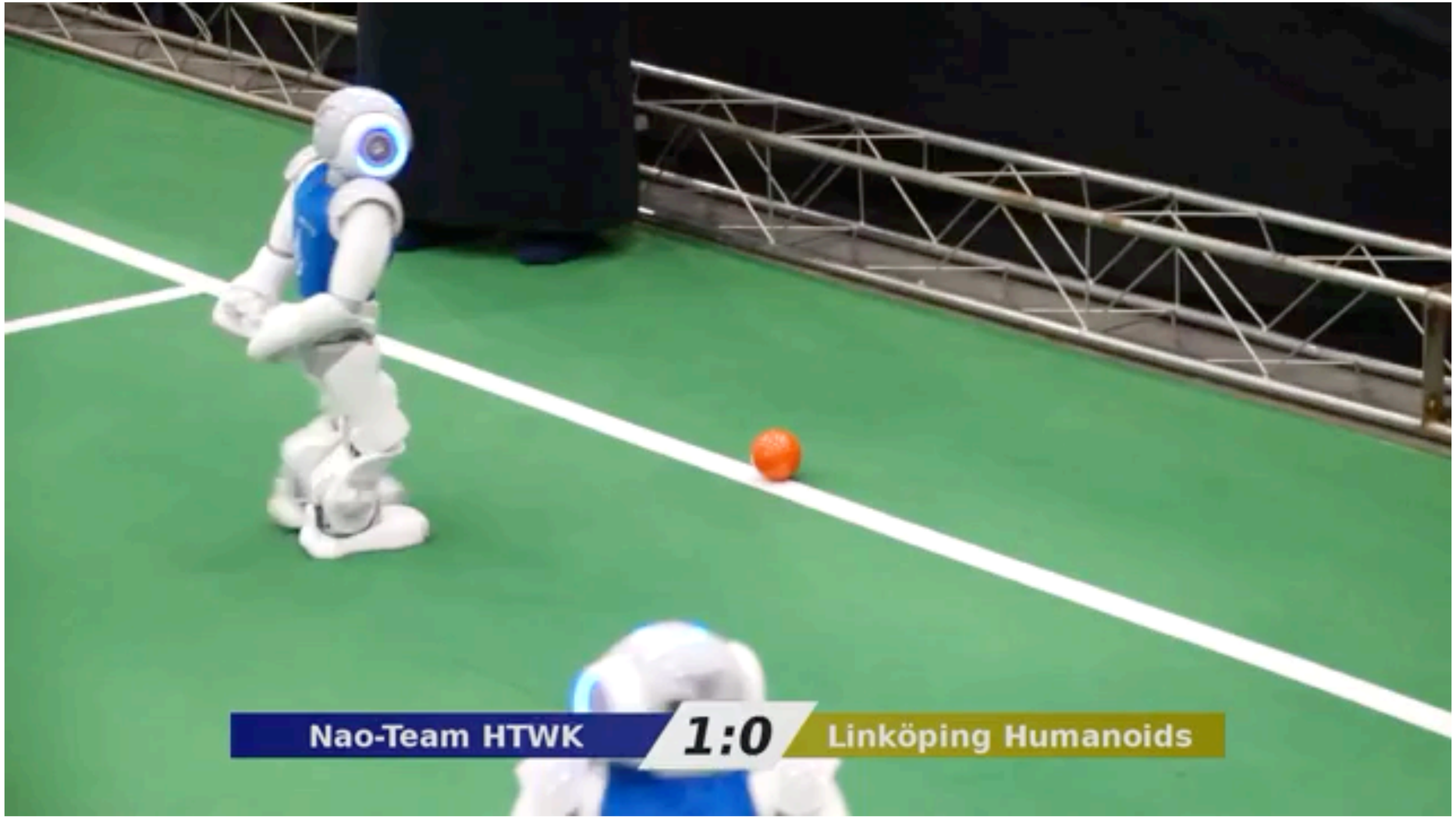
Robot Laundry (2010)



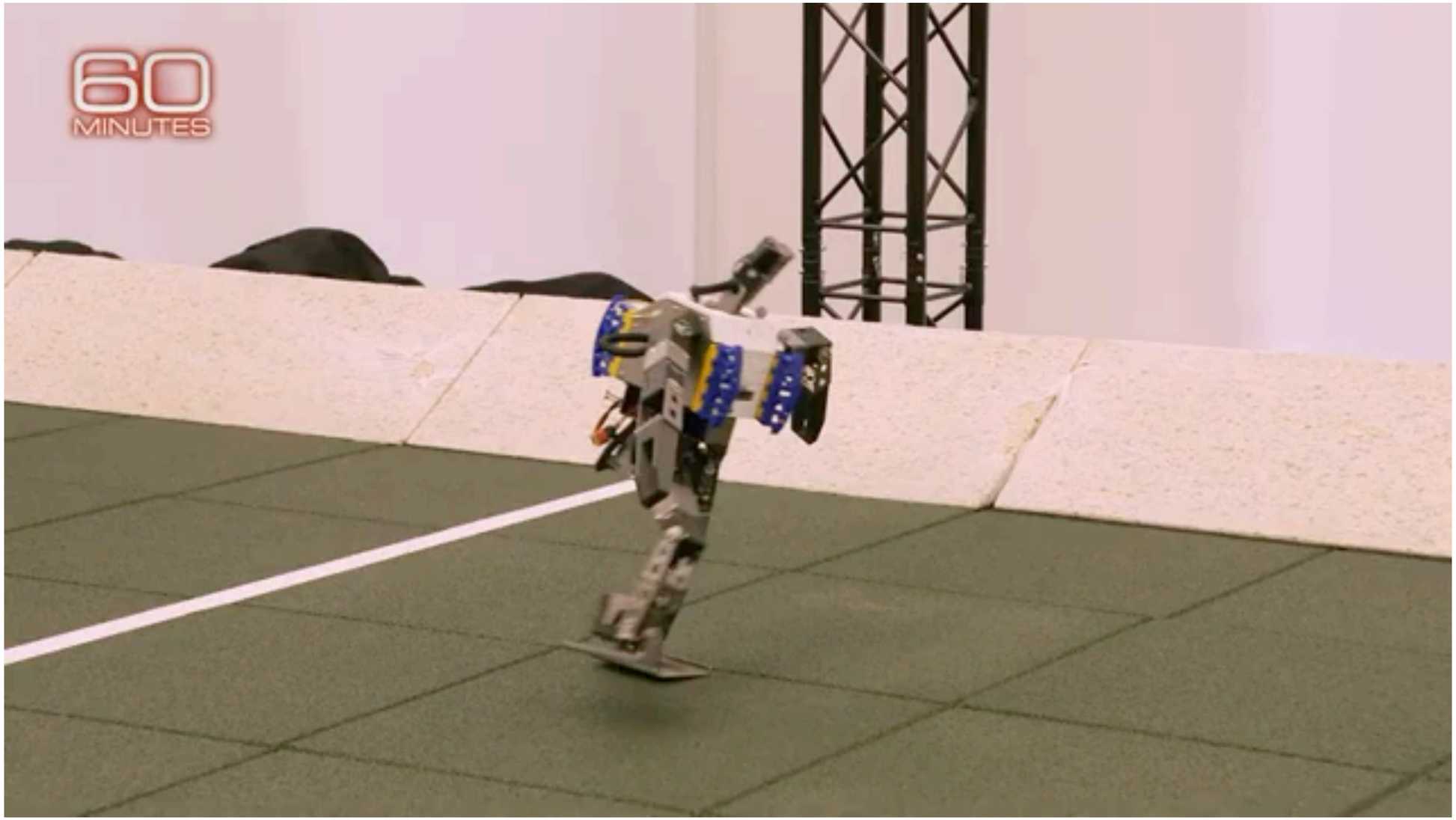
Robot Manipulation (2023)



Robot Soccer (2015)



Robot Soccer (2023)



Full body control of humanoids

Robotics

We will cover several topics relevant to robotics:

- Planning and search
- Reinforcement learning
- Time-series analysis
- State estimation and filtering

Game Playing

- Classic Moment: May, '97: Deep Blue vs. Kasparov
  - First match won against world champion
  - “Intelligent creative” play
  - 200 million board positions per second
  - Humans understood 99.9 of Deep Blue's moves
  - Can do about the same now with a PC cluster
- Open question:
  - How does human cognition deal with the search space explosion of chess?
  - Or: how can humans compete with computers at all??
- 1996: Kasparov Beats Deep Blue

  “I could feel --- I could smell --- a new kind of intelligence across the table.”
- 1997: Deep Blue Beats Kasparov

  “Deep Blue hasn't proven anything.”
- Huge game-playing advances recently, e.g. in Go

Text from Bart Selman, image from IBM’s Deep Blue pages

Game playing

We will cover AI for game playing in depth:

- Planning and search
- Adversarial / expectimax search
- Reinforcement learning

Logic

- Logical systems
  - Theorem provers
  - NASA fault diagnosis
- Methods:
  - Deduction systems
  - Constraint satisfaction
  - Satisfiability solvers (huge advances!)

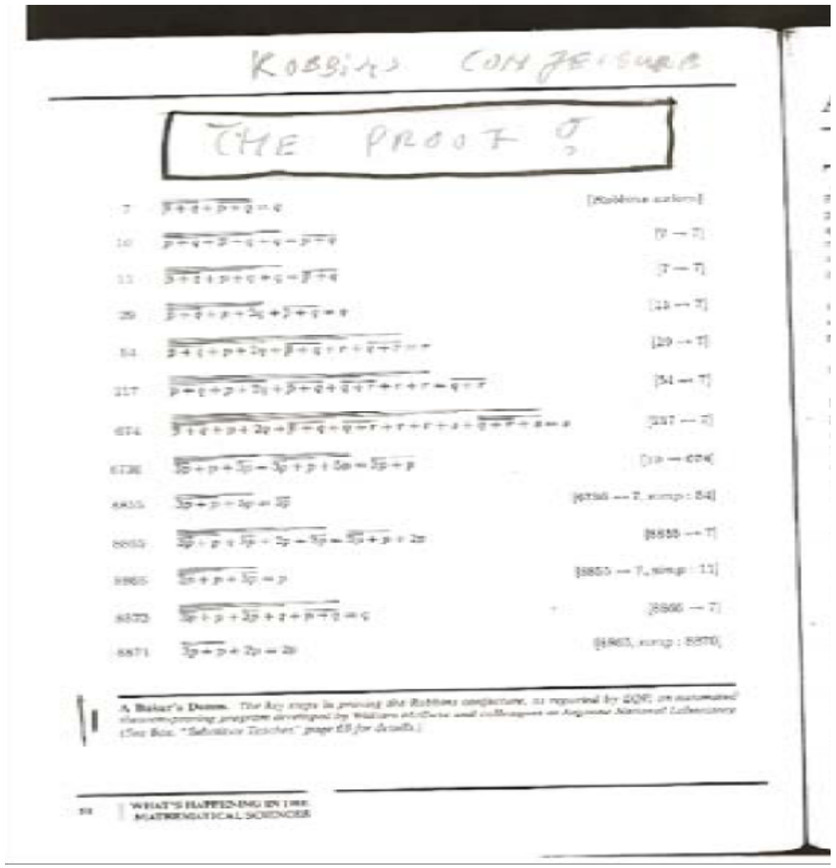


Image from Bart Selman

Course Topics

- Part I: Making Decisions
  - Fast search / planning
  - Constraint satisfaction
  - Adversarial and uncertain search
  - MDPs and Reinforcement learning

- Part II: Reasoning under Uncertainty
  - Bayes nets
  - Decision theory and value of information
  - Statistical Machine learning

- Throughout: Applications, Ethics, and Societal impacts

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 **performance measure, environment, actions, and sensing** dictate techniques for selecting rational actions
- **By then end of the course** you should understand:
  - General AI techniques for a variety of problem types
  - How to recognize when and how a new problem can be solved with an existing technique