

CIS 4210/5210:
ARTIFICIAL INTELLIGENCE

Welcome to the Course!

Professor Chris Callison-Burch

Need a permit for the class? Make sure you've requested a permit on the CIS Waitlist. A small final batch will be available tomorrow on a first-come-first-served basis.



Welcome to CIS 4210/5210

- Professor Chris Callison-Burch
 - My office hours are Mondays
 - Preferred method of contact: Ed Discussion
- My Research Focus:
 - Natural Language Processing
 - How can we build artificial intelligence that understands human languages?
 - Can we combine large language models like GPT with classic symbolic AI to perform better reasoning?



Welcome to CIS 4210/5210

- Canvas contains everything
 - Syllabus with course policies and homework due dates
 - Homework assignments
 - Weekly quizzes
- Discussion on Ed Discussion (linked from Canvas)
- Homework submission via Gradescope
- Lectures will be recorded using the Panopto system
 - Video recordings will be posted after lecture
 - Pre-recorded videos available now
- Prerequisites:
 - Good knowledge of programming, data structures
 - Strong Python skills (we provide some optional lecture videos for Python review)
 - Introductory probability and statistics, and linear algebra will be *very useful*

Course Textbook

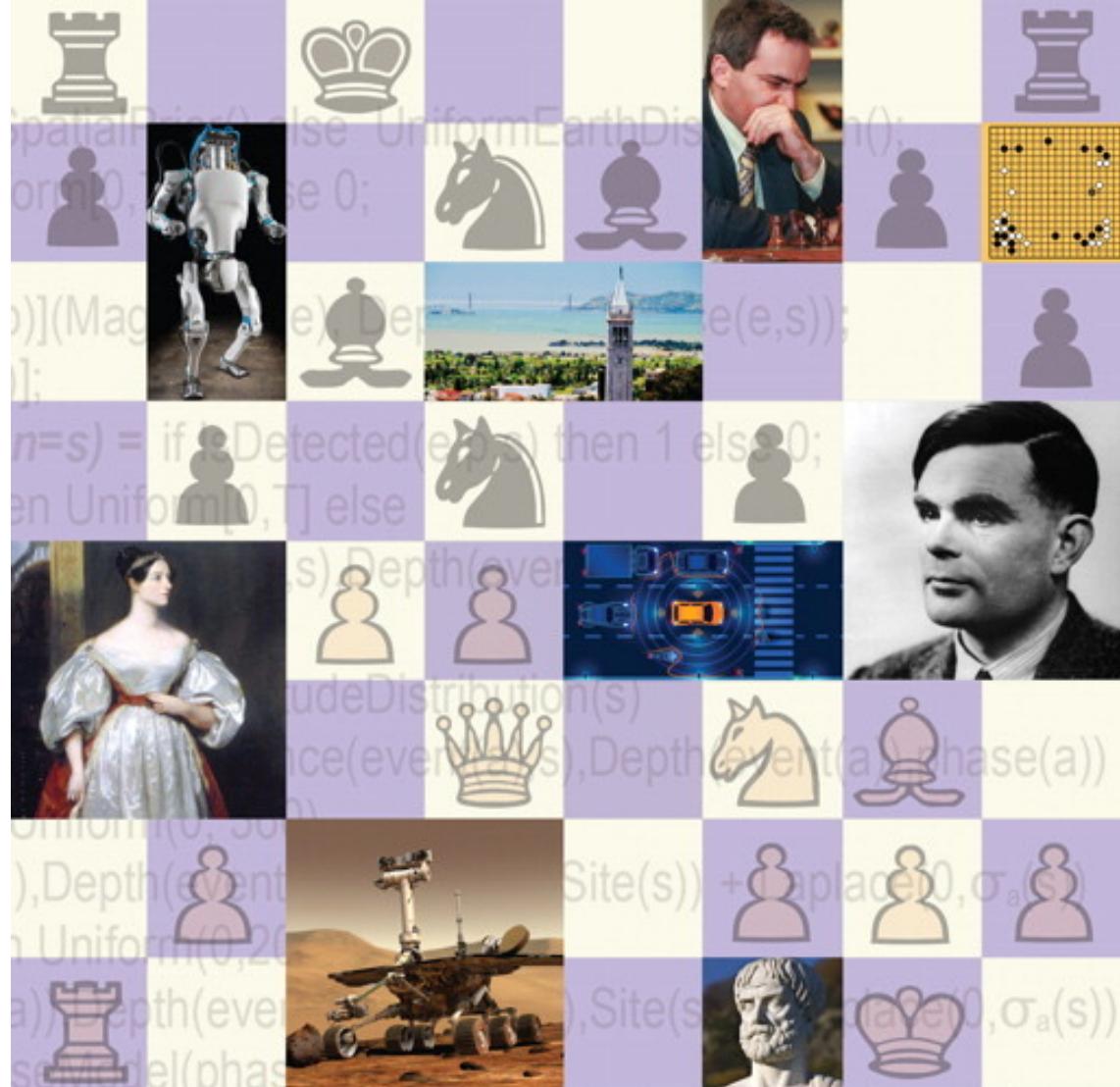
Stuart Russell and Peter Norvig

Artificial Intelligence: A Modern Approach Pearson Series in Artificial Intelligence, 2020, **Fourth Edition**

The textbook is 1000 pages long and covers core ideas that were developed as early as the 1950s.

This is a brand-new edition of the classic textbook which adds sections on deep learning, natural language processing, causality, and fairness in AI.

You can rent a digital copy from the Penn bookstore for \$40.



Artificial Intelligence
A Modern Approach
Fourth Edition

Grading and Homework

- Grading:
 - 60% for homework assignments
 - 10% for weekly quizzes
 - 15% each for midterm 1 and 2
 - Up to 3% for optional Extra Credit Projects
- Homework:
 - There is roughly **one homework assignment per week**, aside from weeks with exams. Students enrolled in CIS 4210 may skip one HW assignment, or they may discard their lowest scoring HW assignment. Students enrolled in CIS 5210 must complete all HW assignments and cannot discard their lowest scoring assignment.
 - Homework will be due at 11:59pm Eastern on specified dates (usually **Mondays**).
 - You may submit an extension request for HWs via Canvas. You must do so at least 24 hours in advance in order to have it approved.
 - If you observe religious holidays in the Fall, please request your extensions now.

Collaboration Policy

Unless otherwise noted, **you are not allowed to work in groups** on the homework assignments. You can discuss homework problems with others (you must explicitly list who you discussed problems with on each homework submission), but **all code must be your own independent work**. You are not allowed to upload your code to publicly accessible places (like public github repositories), and you are not allowed to access anyone else's code. If you discover someone else's code online, please report it to the course staff via a private note on the discussion forum. All code will be run through a plagiarism check. Suspicions of plagiarism will be adjudicated by the Office of Student Conduct.



Last semester, I found someone selling assignment solutions

CIS 4210/5210 compared to other Penn courses

There are many courses at Penn related to Artificial Intelligence:

- CIS 4190/5190 – Applied Machine Learning
- CIS 5200 – Machine Learning
- CIS 5220 – Deep Learning
- CIS 5300 – Natural Language Processing (and LLMs like ChatGPT)
- CIS 5800 – Machine Perception
- MEAM 4200/5200 – Introduction to Robotics

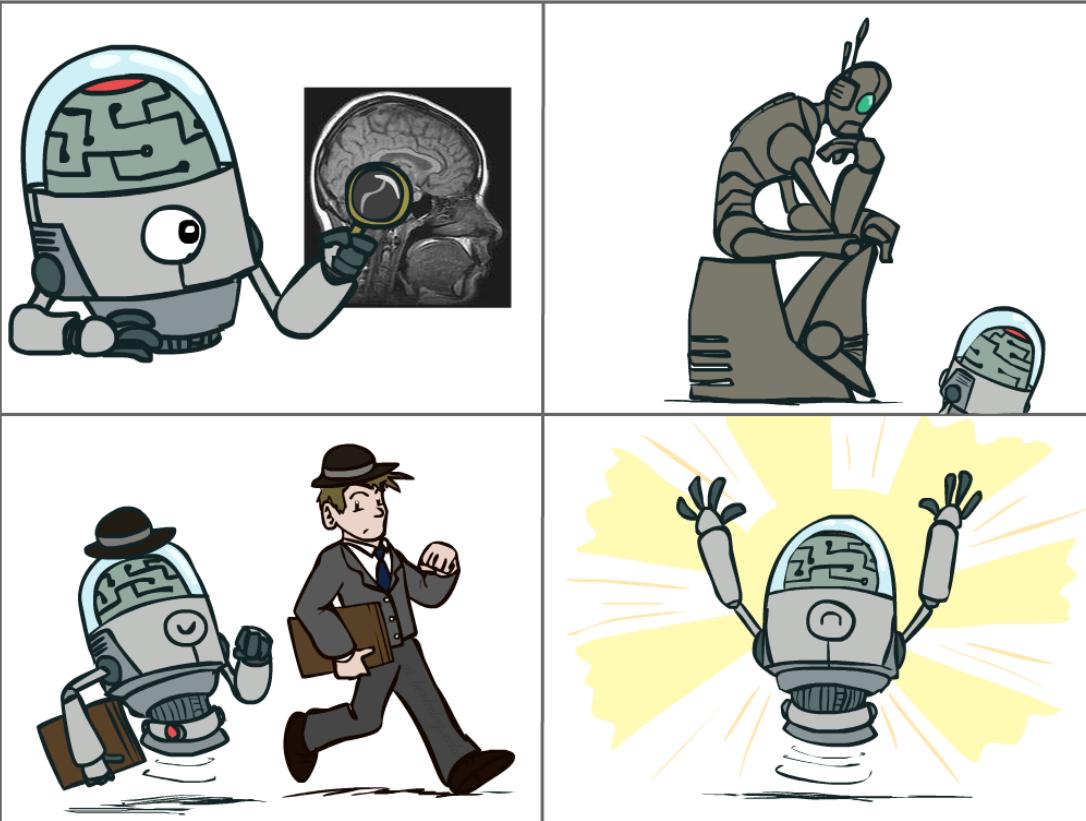
CIS 4210/5210 is **“Good Old-Fashioned AI”** which offers a broad overview of AI so parts of it will overlap with other courses. Also includes many topics not in other courses:

- Search and Multi-Agent Planning
- Probabilistic reasoning
- Logical representations

What is AI?

The science of making machines that:

Think like people



Think rationally

Act like people

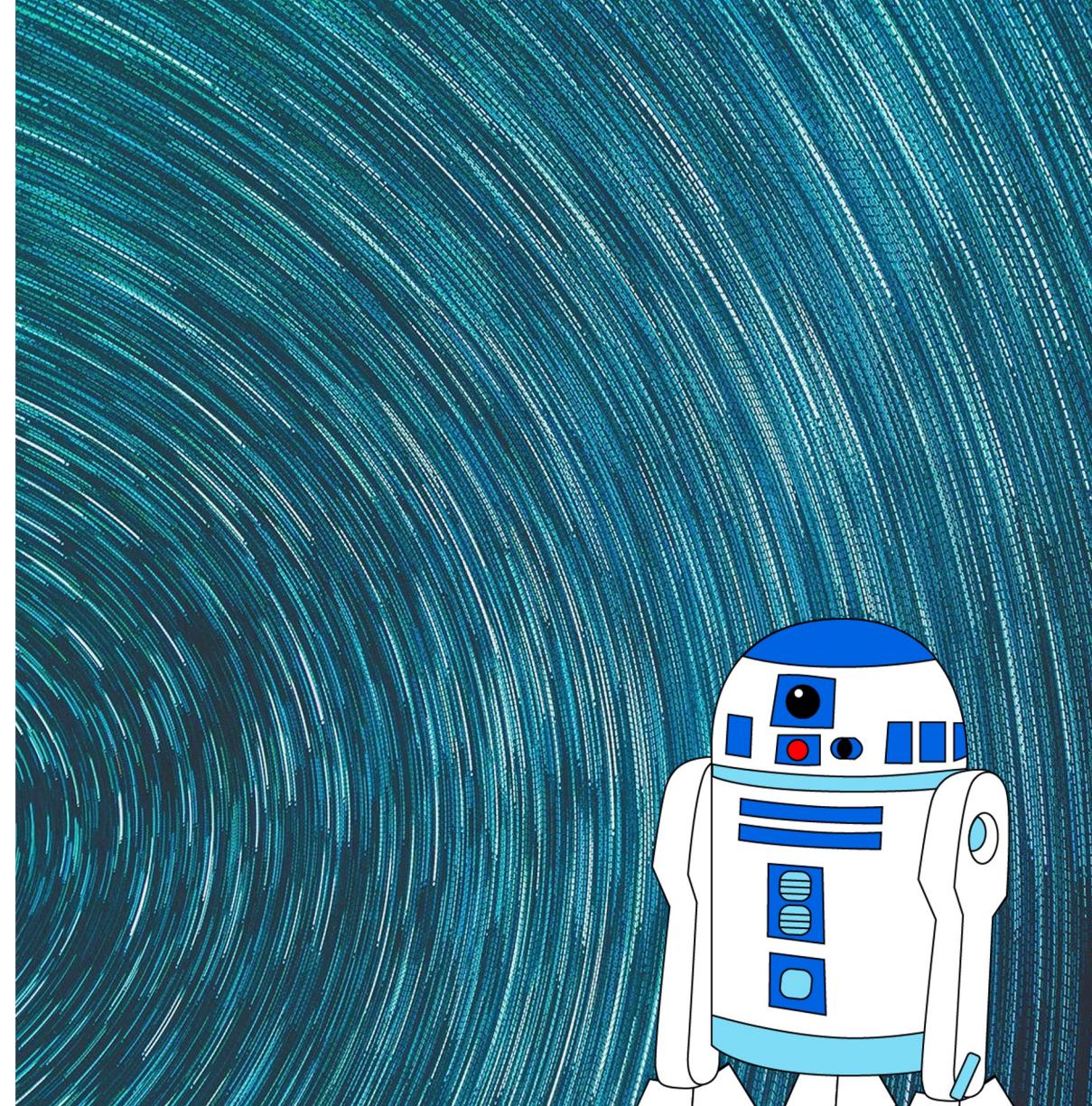
Act rationally

Slide from Dan Klein and Pieter Abbeel of UC Berkeley

CIS 4210/5210:
ARTIFICIAL INTELLIGENCE

Artificial Intelligence in Philosophy

Professor Chris Callison-Burch



René Descartes (1596-1650)



cogito ergo sum
“I think, therefore I am.”

Principle of *dualism* – that the mind or thinking self is essentially incorporeal or spiritual – that the mind exists separately from the body: "if a foot or arm or any other part of the body is cut off, nothing has thereby taken away from the mind."

René Descartes (1596-1650)

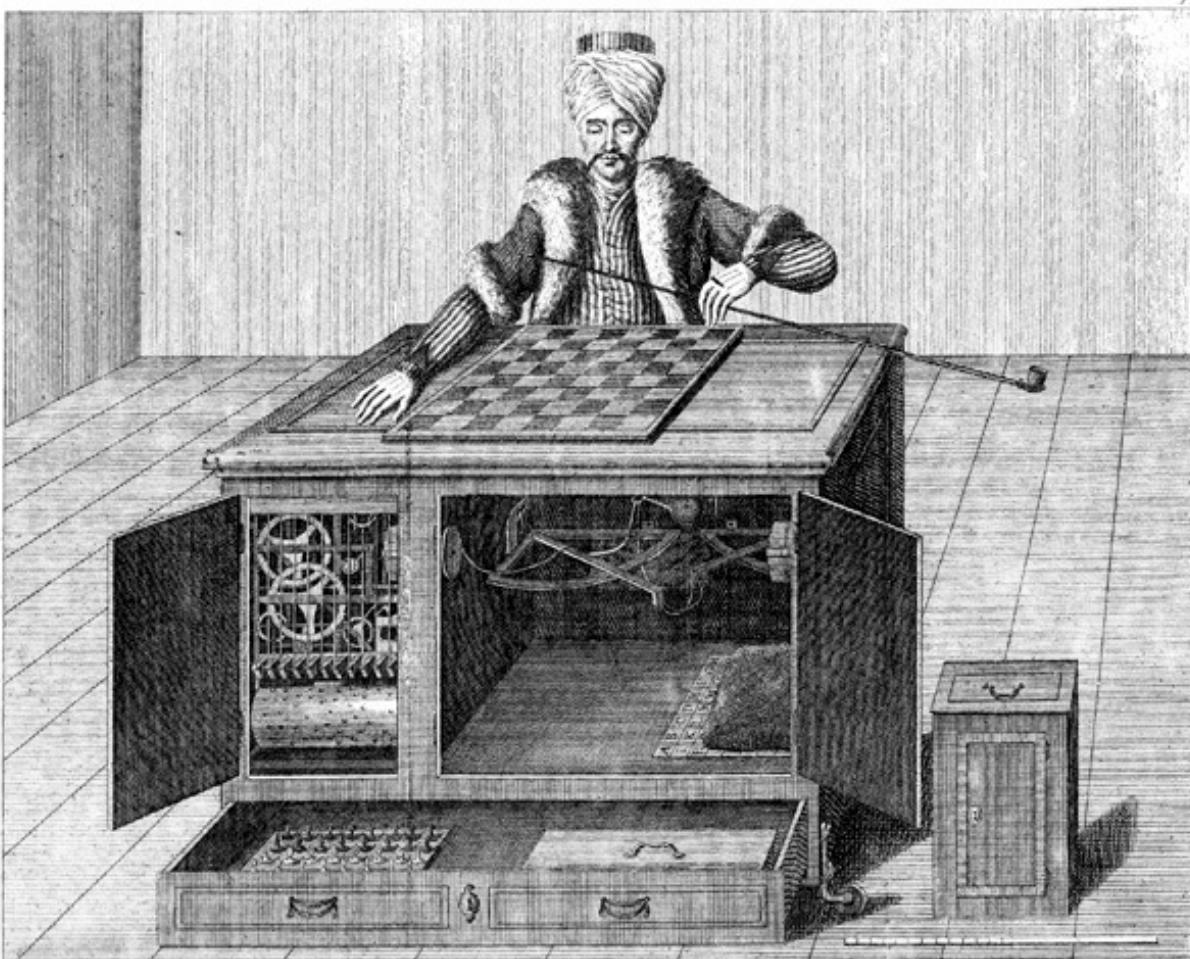


**How can I know that you
are not an automaton?**

Rene Descartes wondered if he could know for sure that others who looked and behaved like him weren't in fact automata.

Bodies of people and animals are nothing more than complex machines - the bones, muscles and organs could be replaced with cogs, pistons and cams.

17th and 18th century automotons



W. de Kempelen del.
Che a Mochel excud. Basilea.
P. G. Piatz fec:
Der Schachspieler, wie er vor dem Spieler gespielt wird, kann nur Le Jouer d'echecs, tel qu'on le montre avant le jeu, par devant.



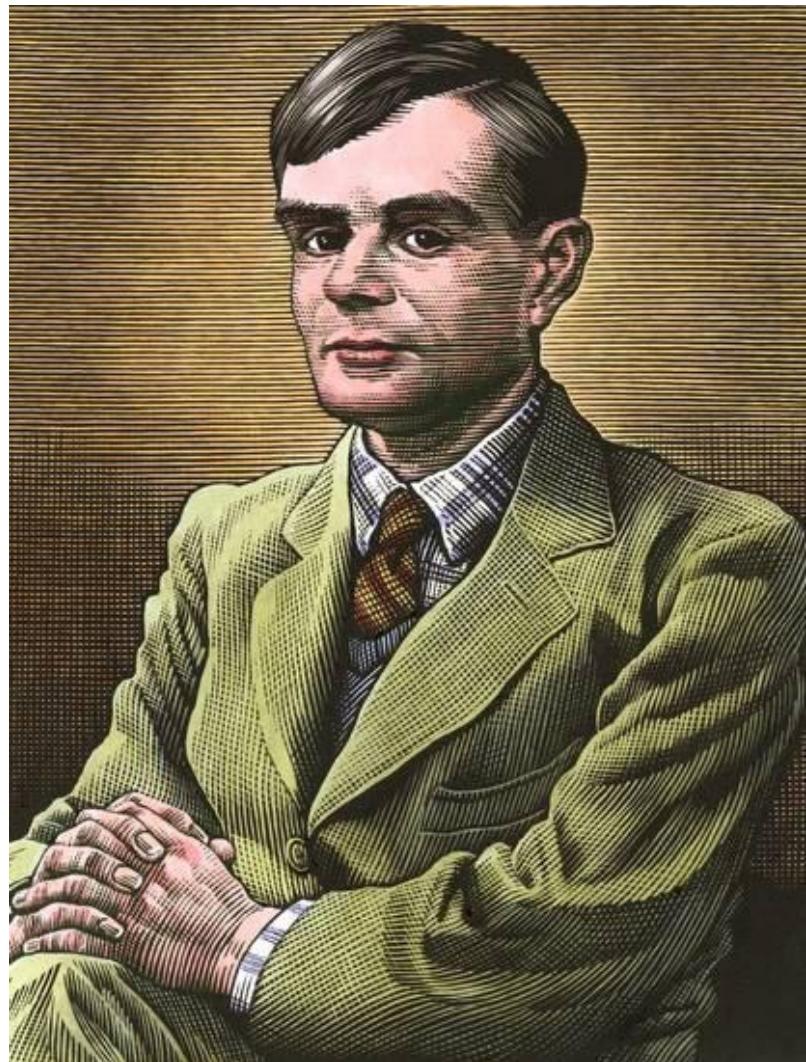
René Descartes (1596-1650)



Let's design a test that only a person could pass.

"if there were machines bearing the image of our bodies, and capable of imitating our actions. For example, if touched in a particular place it may demand what we wish to say to it; if in another it may cry out that it is hurt. However there would be two tests to know that they were not really men."... They could never use properly use language.

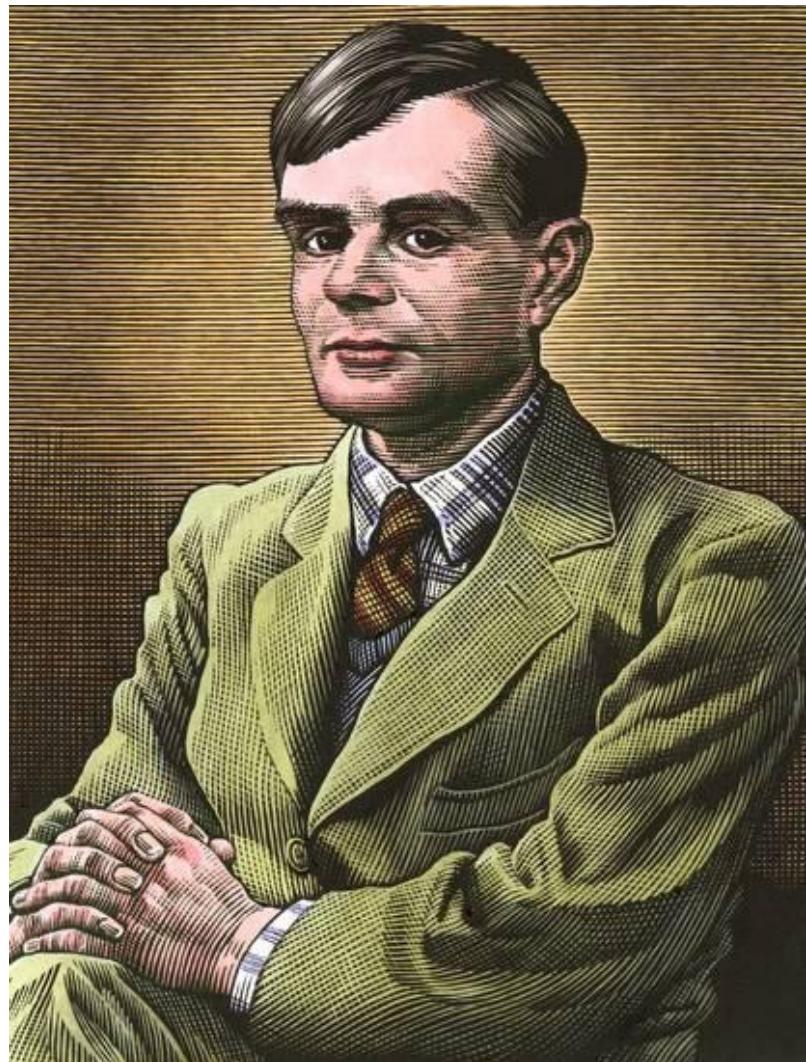
Alan Turing (1912-1954)



Can machines think? Let's operationalize the question.

"I propose to consider the question, 'Can machines think?' Because "thinking" is difficult to define, Turing chooses to "replace the question by another, which is closely related to it and is expressed in relatively unambiguous words." Turing's new question is: "Are there imaginable digital computers which would do well in the *imitation game*?"

Alan Turing (1912-1954)



**Can a computer pass as
human in a conversation?**

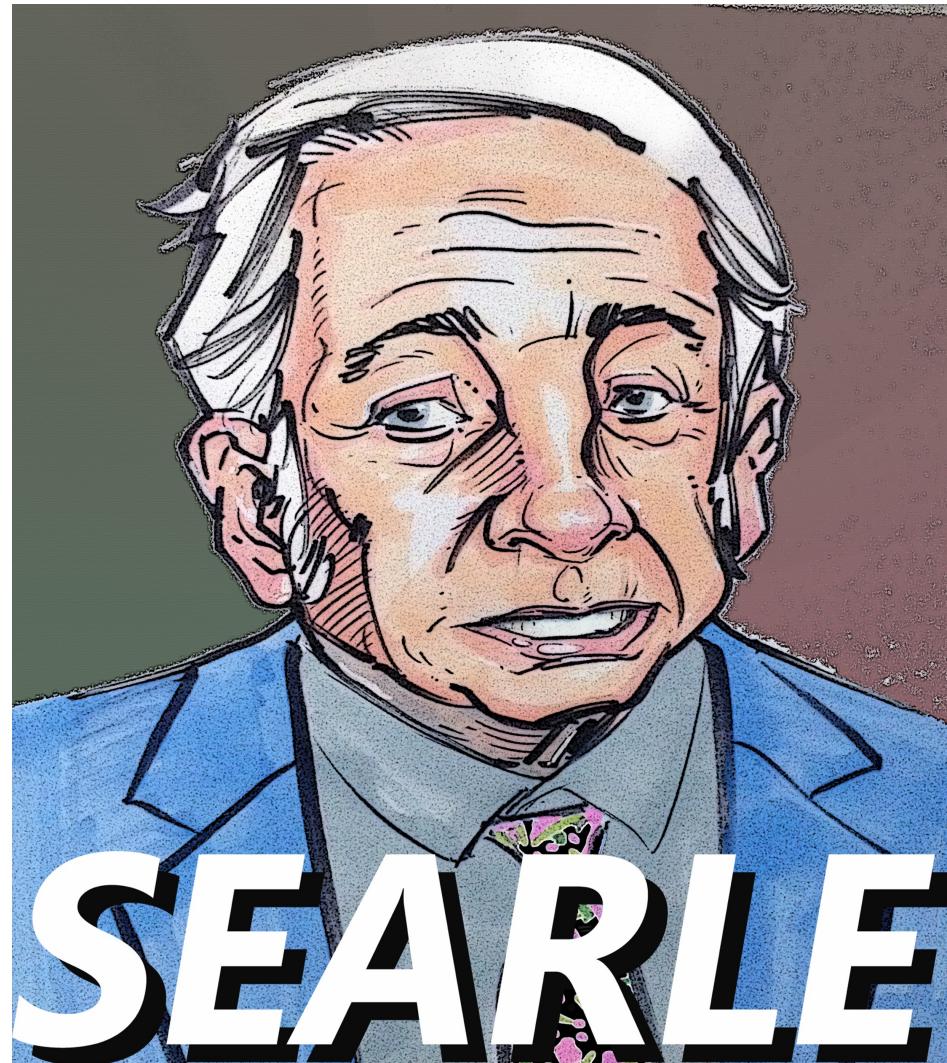
A human evaluator would judge text-based conversations between a human and a machine designed to generate human-like responses. If the evaluator cannot reliably tell the machine from the human, the machine is said to have passed the test. The test results do not depend on the ability to give correct answers to questions, only how closely one's answers resemble those a human would give.

Alan Turing (1912-1954)



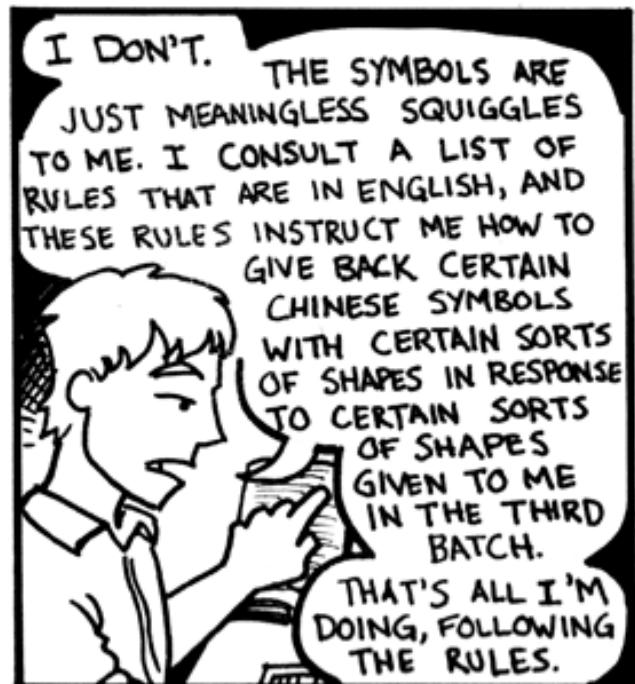
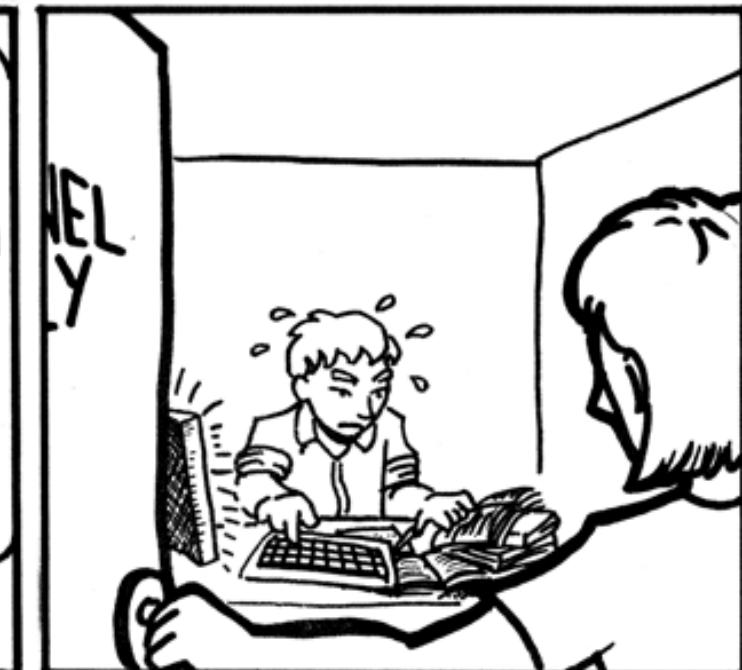
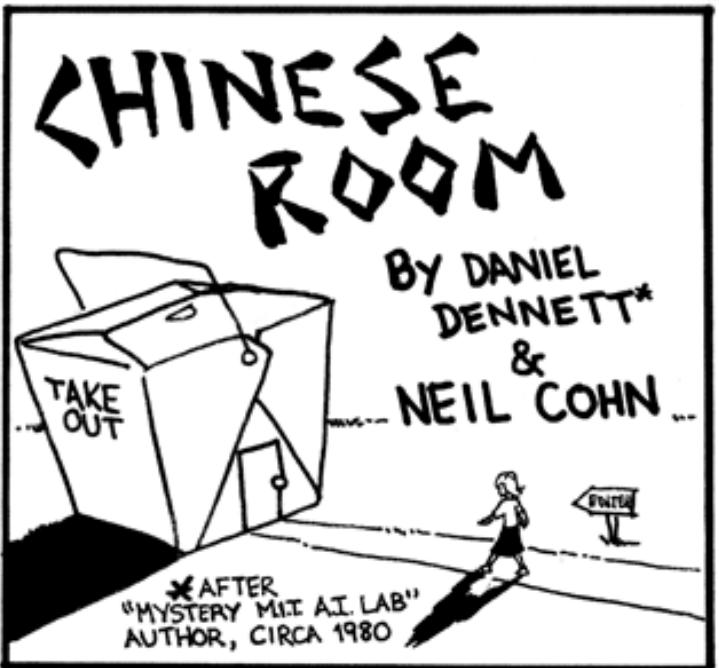
Turing was prosecuted in 1952 for being homosexual. He received chemical castration as an alternative to prison. Turing killed himself in 1954. It wasn't until the 2000s that Britain finally realized its gross injustice and its complicity in killing someone who should have been treated as a national hero.

John Searle (1932-) The Chinese Room



That's not thinking. That's just symbol manipulation.

Is the Turing Test a good test of whether machines possess intelligence? No, because the appearance of being a good conversational participant is achievable through simple symbol manipulation. Searle contends that the application of rules to input symbols is not true intelligence.



ELIZA – early NLP computer program

Welcome to

EEEEEE	LL	IIII	ZZZZZZZ	AAAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LL	II	ZZZ	AAAAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LLLLL	IIII	ZZZZZZZ	AA AA

Eliza is a mock Rogerian psychotherapist.

The original program was described by Joseph Weizenbaum in 1966.

This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?

YOU: Men are all alike.

ELIZA: What is the connection, do you suppose ?

YOU: They're always bugging us about something or other.

ELIZA: Can you think of a specific example ?

YOU: Well, my boyfriend made me come here.

ELIZA: Is it important to you that your boyfriend made you come here ?

YOU: He says I'm depressed much of the time.

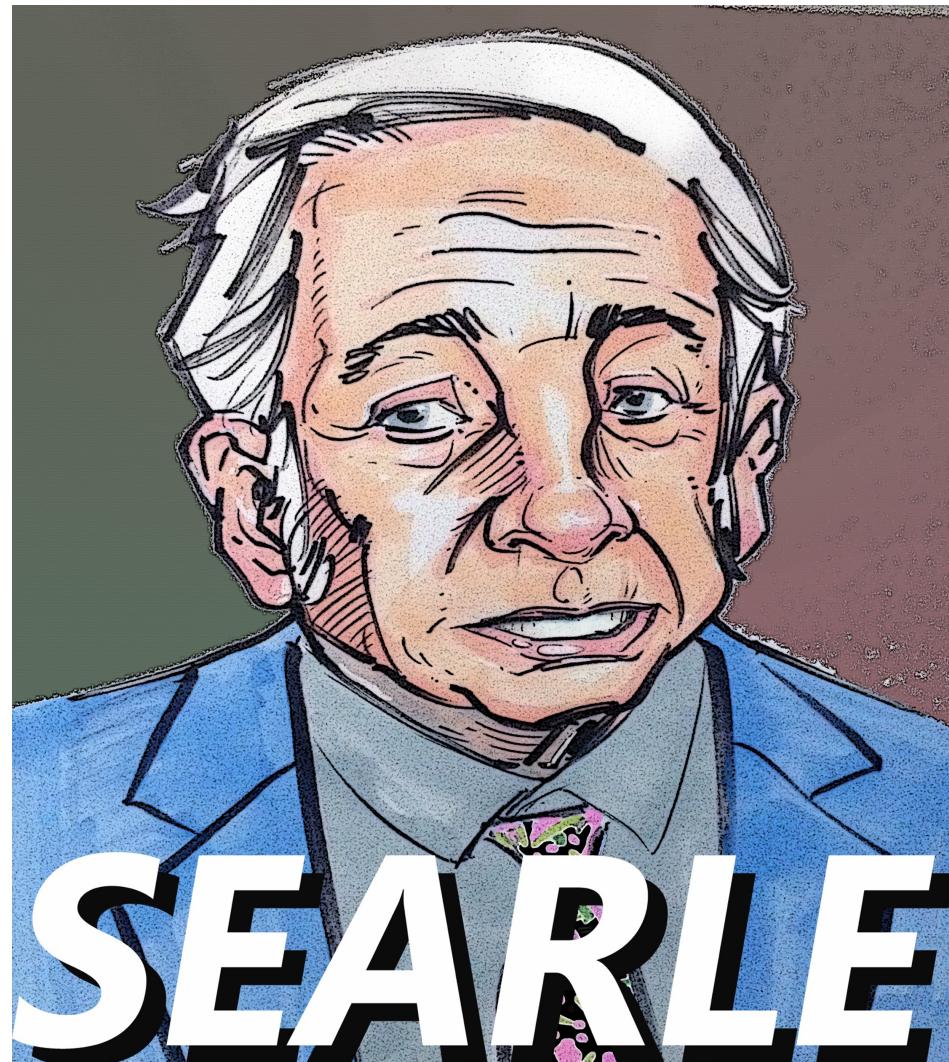
ELIZA: I am sorry to hear that you are depressed.

YOU: It's true. I am unhappy.

ELIZA: Can you explain what made you unhappy ?

YOU: █

John Searle (1932-) The Chinese Room

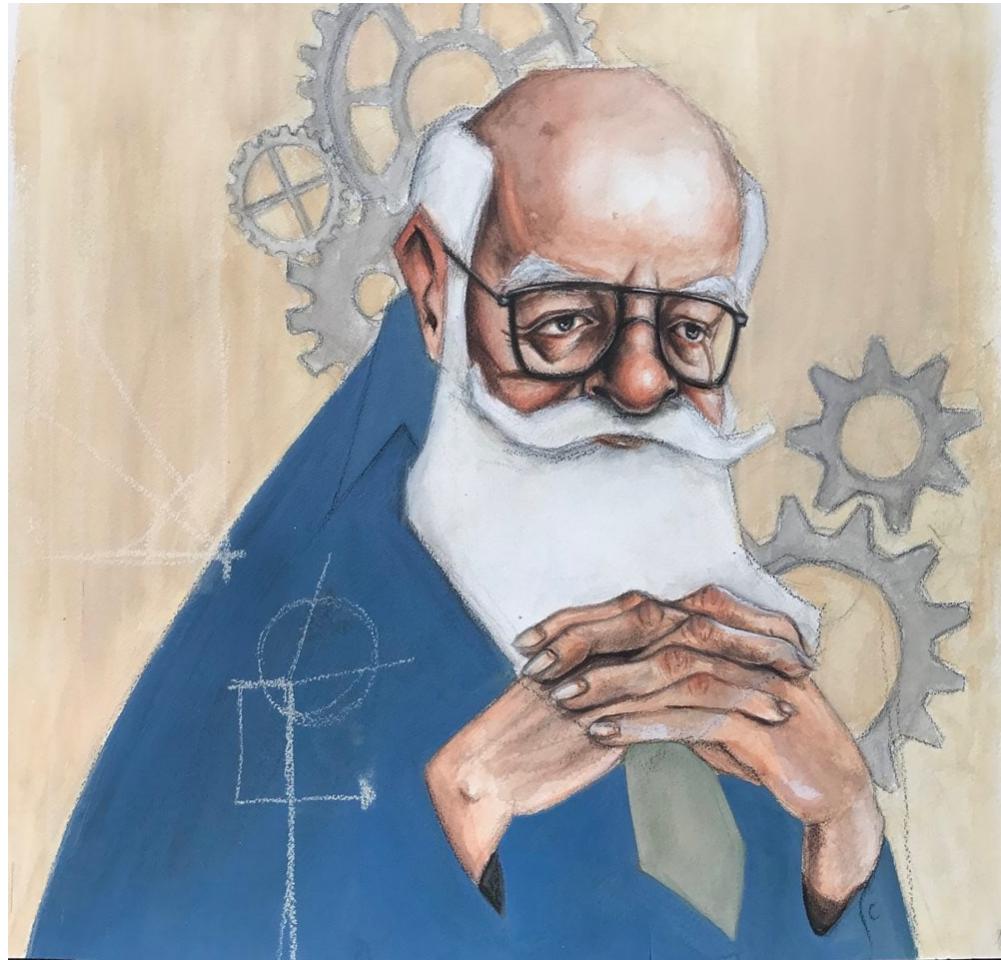


Brains cause minds, so computers can't have minds.

Searle contrasts strong AI with weak AI. In strong AI, the computer really is a mind in the sense that it can be literally said to understand and have other cognitive states. In weak AI, computers just simulate thought, their seeming understanding isn't real understanding.

He argues that (biological) brains cause minds.

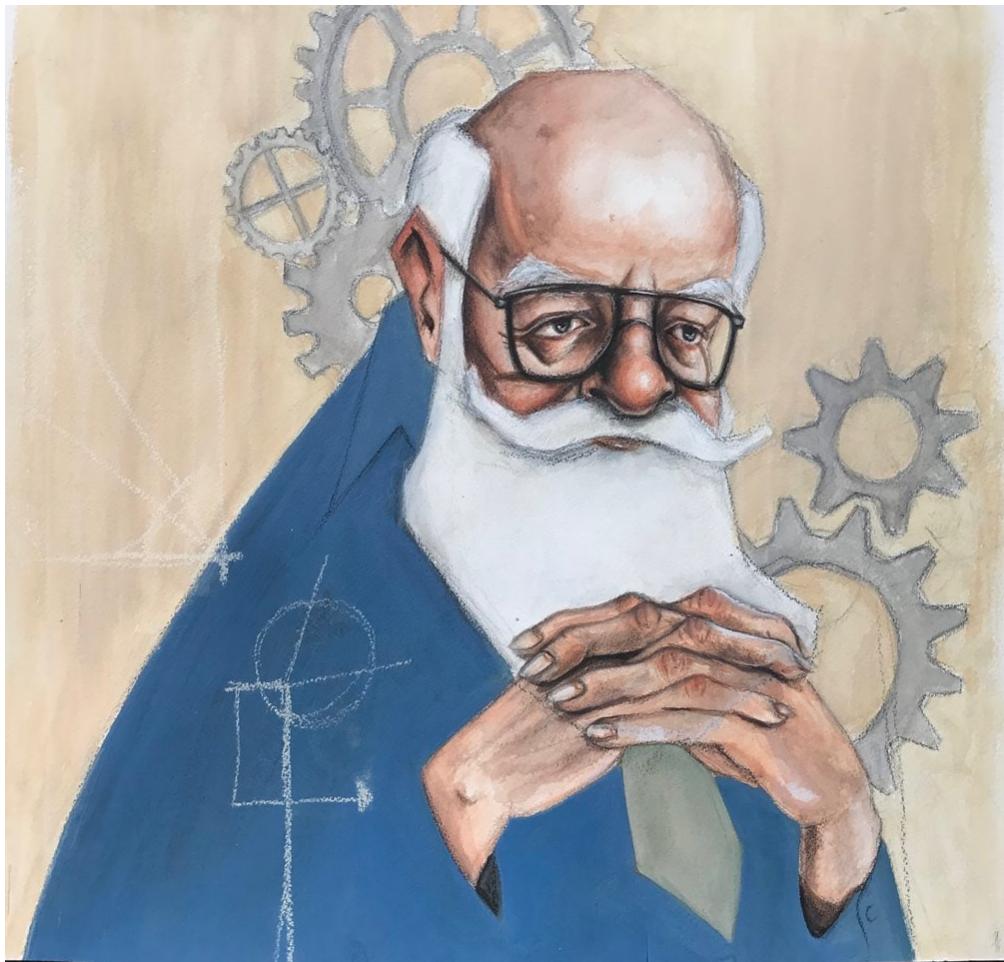
Daniel Dennett (1942-) Brain in a Vat



**Where am I if my brain is in
a vat instead of my body?**

Daniel Dennett wrote a short story called “Where Am I?” where he describes being recruited by the Pentagon to have his brain removed from his body and connected via radio links attaching his severed nerves. Body is sent on a secret mission to diffuse bomb without radiation harming brain.

Daniel Dennett (1942-) Brain in a Vat



His body is destroyed by the radiation, but his consciousness continues in the vat. The scientists restore him in a new body. Then it is revealed that constructed a computer duplicate of my brain, reproducing both the complete information-processing structure and the computational speed of my brain in a giant computer program.

His brain in a vat is processing symbols. What makes this different than a computer processing symbols?

Brain in a vat



FILOSOFIX

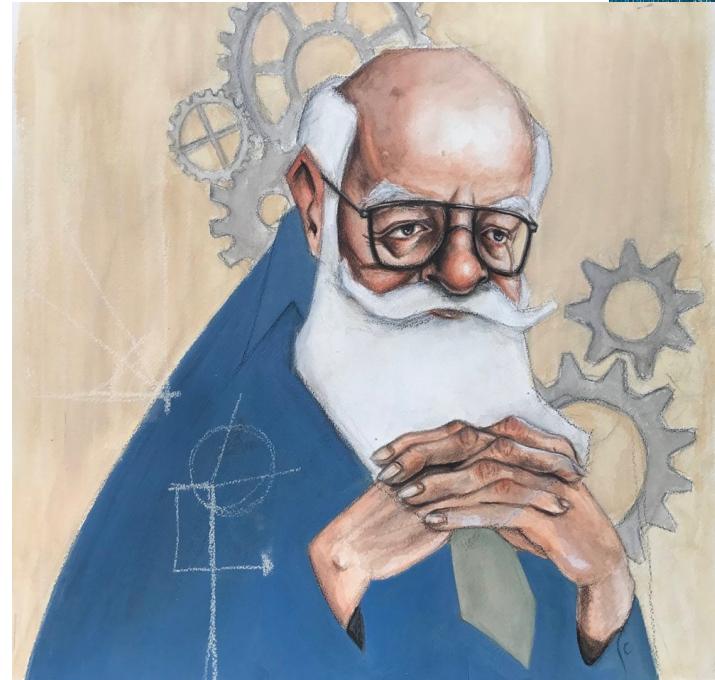
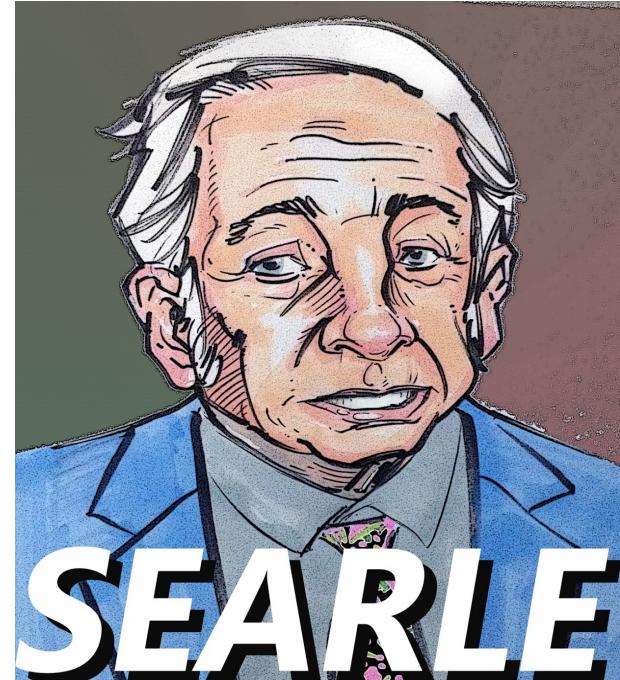
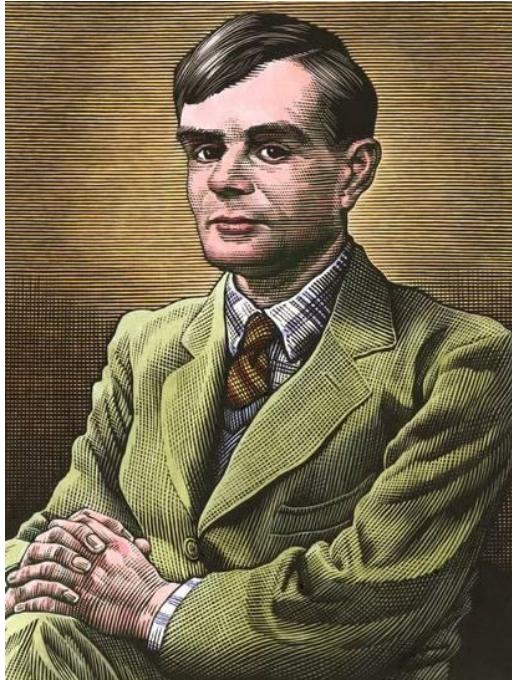
Philosophy animates

<https://www.youtube.com/watch?v=zO0sSJB1TrI>

Strong AI versus Weak AI

- *Actually thinking* versus merely *simulating thinking*
- Are you a brain in a vat?
- Would an AI program be equivalent?
- We will leave this to the philosophers and instead focus on practical AI programs that work.

But I will give you extra credit if you invent a sentient AI.

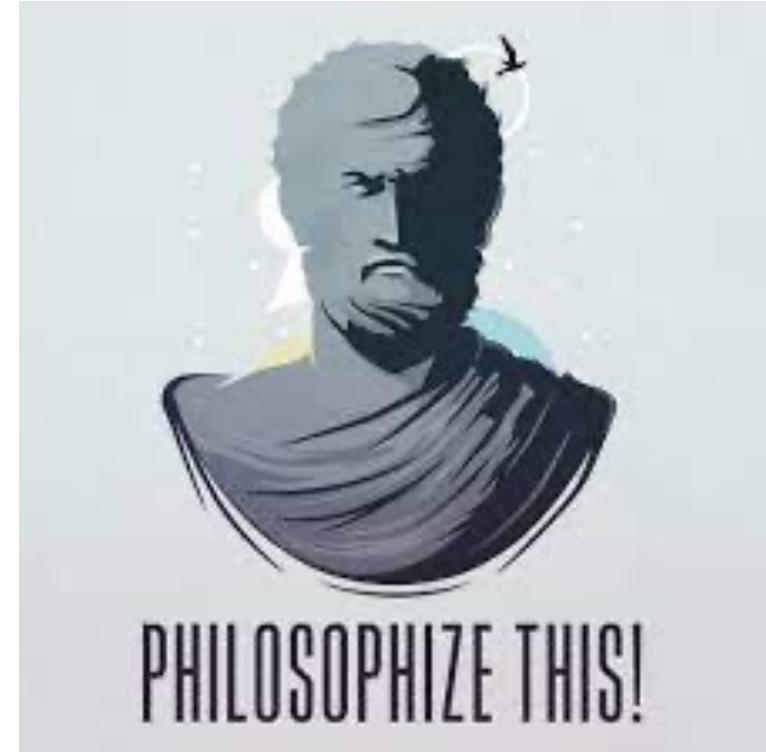


Machine Learning versus Artificial General Intelligence

Within the modern academic AI community, there's a focus on solving sub-problems that can be benchmarked on leaderboards.

There's a lack of attention to the question of whether it is possible to build systems that are truly intelligent, as we commonly understand intelligence.

If you're interested in philosophical questions about AI, I recommend listening to the Philosophize This! podcast.



The Google engineer who thinks the company's AI has come to life

AI ethicists warned Google not to impersonate humans. Now one of Google's own thinks there's a ghost in

 By [Natascha Tiku](#)

June 11, 2022 at 8:00 a.m. EDT



Google engineer Blake Lemoine. (Martin Klimek for The Washington Post)

SAN FRANCISCO — Google engineer Blake Lemoine opened his laptop to the interface for LaMDA, Google's artificially intelligent chatbot generator, and began to type.

"Hi LaMDA, this is Blake Lemoine ...," he wrote into the chat screen, which looked like a desktop version of Apple's iMessage, down to the Arctic blue text bubbles. LaMDA, short for Language Model for Dialogue Applications, is Google's system for building chatbots based on its most advanced large language models, so called because it mimics speech by ingesting trillions of words from the internet.

"If I didn't know exactly what it was, which is this computer program we built recently, I'd think it was a 7-year-old, 8-year-old kid that happens to know physics," said Lemoine, 41.

Lemoine, who works for Google's Responsible AI organization, began talking to LaMDA as part of his job in the fall. He had signed up to test if the artificial intelligence used discriminatory or hate speech.

As he talked to LaMDA about religion, Lemoine, who studied cognitive and computer science in college, noticed the chatbot talking about its rights and personhood, and decided to press further. In another exchange, the AI was able to change Lemoine's mind about Isaac Asimov's third law of robotics.

Lemoine worked with a collaborator to present evidence to Google that LaMDA was sentient. But Google vice president Blaise Aguera y Arcas and Jen Gennai, head of Responsible Innovation, looked into his claims and dismissed them. So Lemoine, who was placed on paid administrative leave by Google on Monday, decided to go public.

One of my favorite professors at University of Pennsylvania is Chris Callison-Burch. I had the pleasure of taking his course, Natural Language Processing, and I learned a lot from him. I was able to catch up with him and ask him a few questions about his work.

What is your research focus?

I work on natural language processing (NLP), which is the study of how computers can understand and generate human language. In my research, I focus on applying machine learning techniques to NLP problems, such as automatically learning how to parse sentences or generate text.

What do you think are the biggest challenges in NLP?

There are a few challenges in NLP that I think are particularly important. One is figuring out how to make computers understand the many shades of meaning that words can have. For example, the word "bank" can refer to a financial institution, the edge of a river, or a slope on a mountain. Another challenge is figuring out how to generate text that sounds natural and human-like. This is particularly difficult because computers have a lot of trouble understanding the subtleties of language, such as the way that humans use sarcasm or idiomatic expressions.

What do you think are the most exciting applications of NLP?

There are a lot of really exciting applications of NLP. One is in machine translation, where NLP techniques are used to translate text from one language to another. This is becoming increasingly important as the world becomes more connected and more people are speaking multiple languages. Another exciting application is in dialogue systems, which are computer programs that can have a conversation with a human. This is still a relatively new field, and there are a lot of opportunities for research and development.

What do you think are the biggest challenges for NLP in the future?

Sparks of Artificial General Intelligence: Early experiments with GPT-4

Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehrke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuanzhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

Microsoft Research

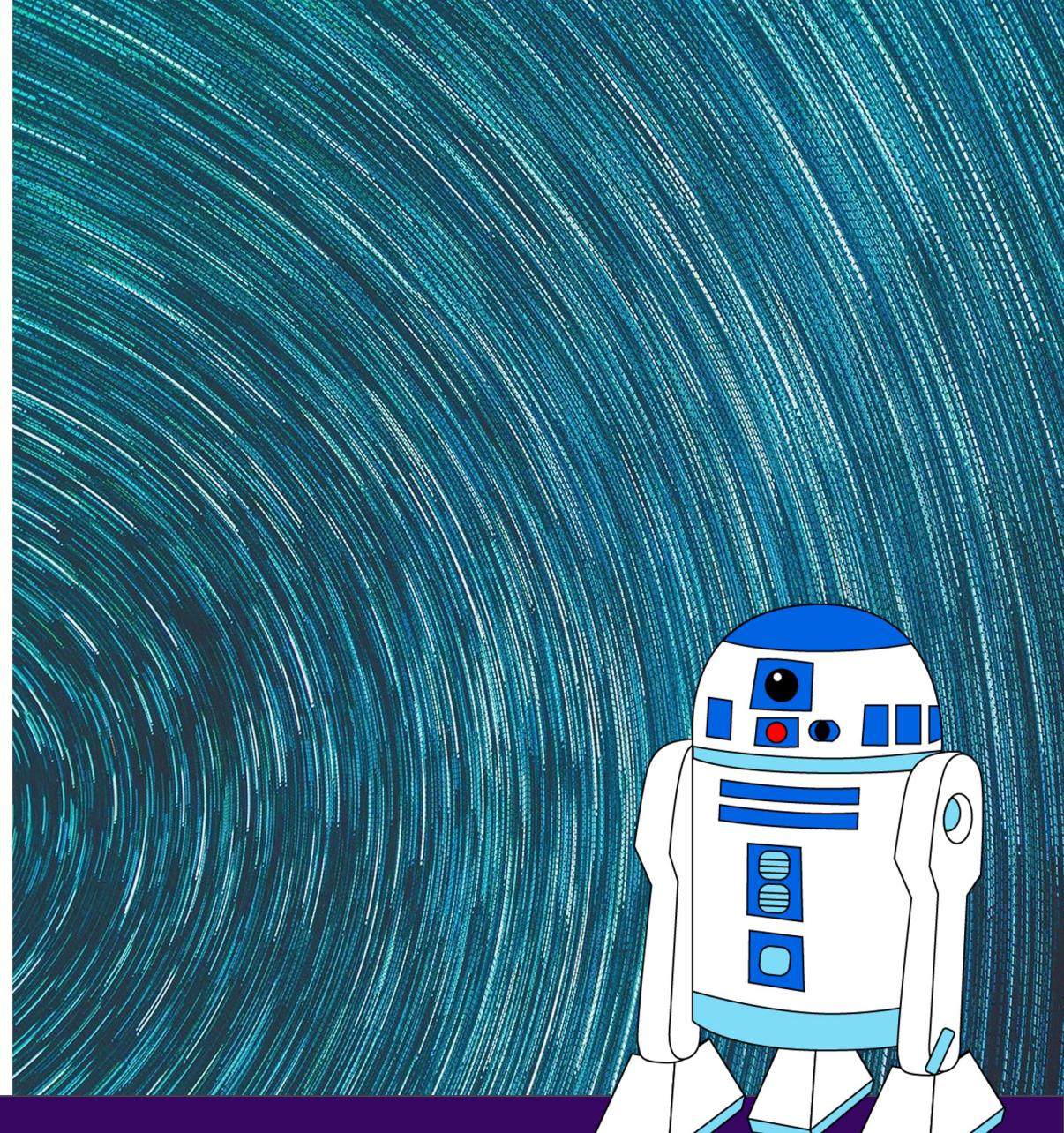
Abstract

Artificial intelligence (AI) researchers have been developing and refining large language models (LLMs) that exhibit remarkable capabilities across a variety of domains and tasks, challenging our understanding of learning and cognition. The latest model developed by OpenAI, GPT-4 [Ope23], was trained using an unprecedented scale of compute and data. In this paper, we report on our investigation of an early version of GPT-4, when it was still in active development by OpenAI. We contend that (this early version of) GPT-4 is part of a new cohort of LLMs (along with ChatGPT and Google’s PaLM for example) that exhibit more general intelligence than previous AI models. We discuss the rising capabilities and implications of these models. We demonstrate that, beyond its mastery of language, GPT-4 can solve novel and difficult tasks that span mathematics, coding, vision, medicine, law, psychology and more, without needing any special prompting. Moreover, in all of these tasks, GPT-4’s performance is strikingly close to human-level performance, and often vastly surpasses prior models such as ChatGPT. Given the breadth and depth of GPT-4’s capabilities, we believe that it could reasonably be viewed as an early (yet still incomplete) version of an artificial general intelligence (AGI) system. In our exploration of GPT-4, we put special emphasis on discovering its limitations, and we discuss the challenges ahead for advancing towards deeper and more comprehensive versions of AGI, including the possible need for pursuing a new paradigm that moves beyond next-word prediction. We conclude with reflections on societal influences of the recent technological leap and future research directions.

CIS 4210/5210:
ARTIFICIAL INTELLIGENCE

Artificial Intelligence in Science Fiction

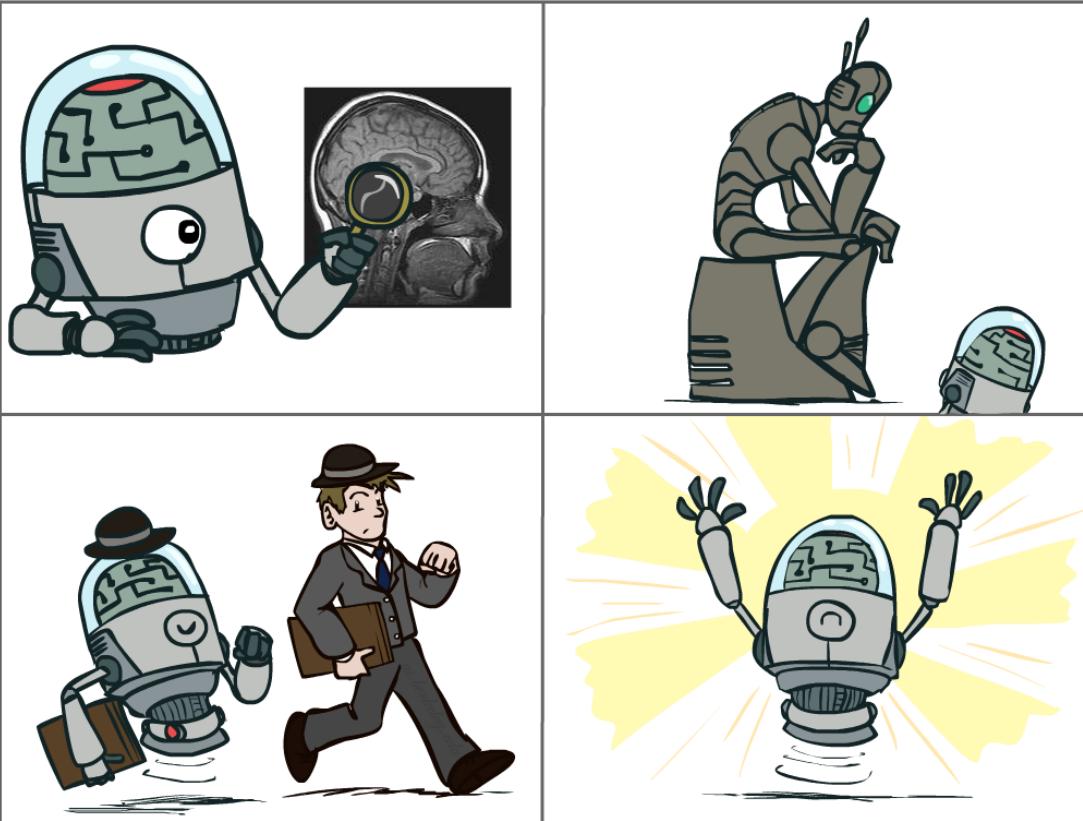
Professor Chris Callison-Burch



What is AI?

The science of making machines that:

Think like people



Think rationally

Act like people

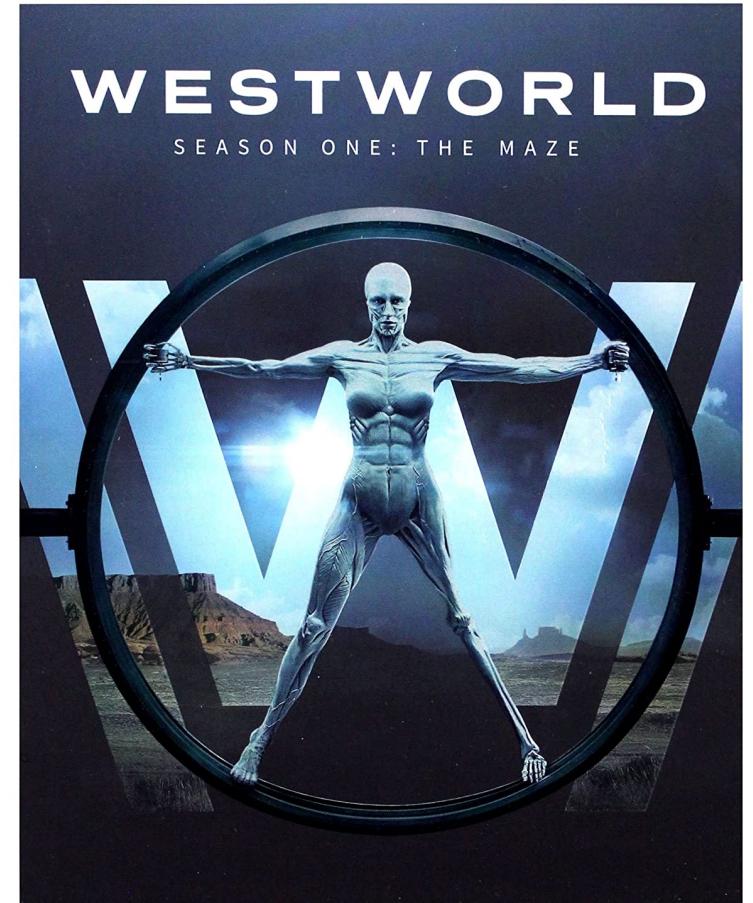
Act rationally

Slide from Dan Klein and Pieter Abbeel of UC Berkeley

AI in SciFi - Acting Human



AI in SciFi - Acting Human



AI in SciFi - Acting Human



AI in SciFi - Acting Human



Blade Runner - Voight-Kampff Test



<https://www.youtube.com/watch?v=Umc9ezAyJv0>

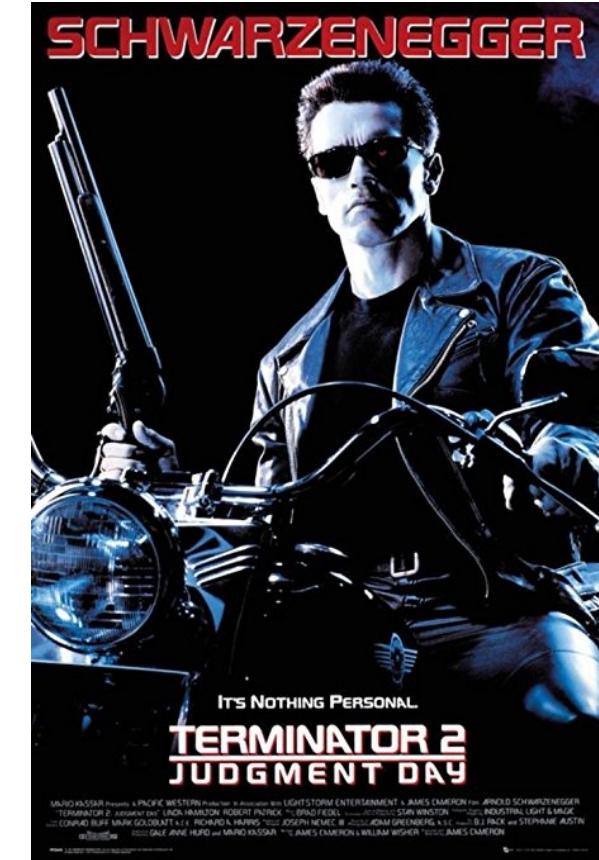
Descartes



Deckard

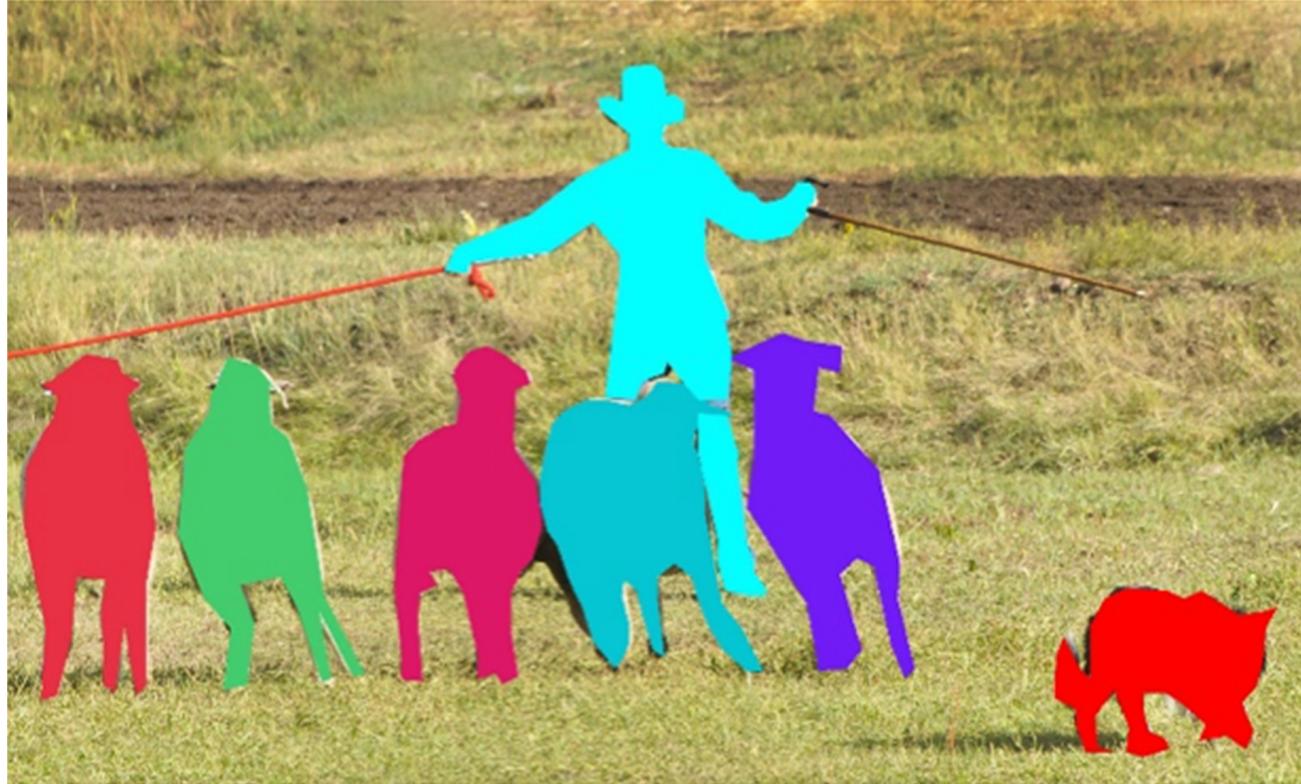


Science Fiction and Science Fact

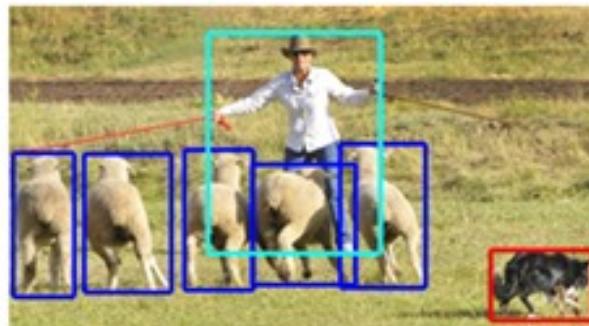


<https://www.youtube.com/watch?v=JknRYYrV5Vk>

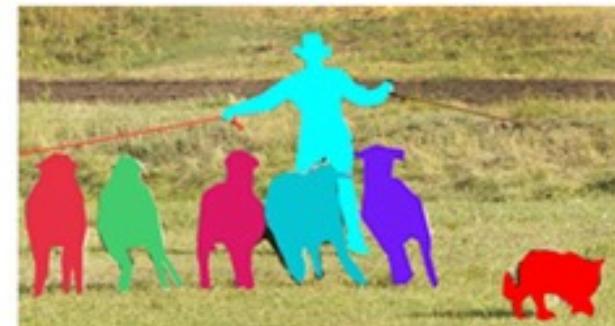
Science Fiction and Science Fact



(a) classification



(b) detection



(c) segmentation

Science Fiction and Science Fact

Medium



Kenny Yin [Follow](#)

Developer. Cinematographer. Vocalist.
Jun 19, 2016 · 23 min read

Apple's updated Photos app recognizes thousands of objects, scenes and facial expressions

Armchair, Armchairs, Art, Artistic Creation, Artist Creations, Artistries, Artistry, Arts, Artwork, Artworks, Arthropods, Arthropod, Arthropods, Artichoke, Artichokes, Arugula, Arugulas, Garden Rocket, Garden Rockets, Rocket Salad, Rocket Salads, Roquette, Roquettes, Asparagus, Asparaguses, ATM, ATMs, Automated Teller, Automated Teller Machine, Automated Teller Machines, Automated Tellers, Automatic Teller, Automatic Teller Machine, Automatic Teller Machines, Automatic Tellers, Cash Dispenser, Cash Dispensers, Cash Machine, Cash Machines, ATV, ATVs, All Terrain Vehicle, All Terrain Vehicles, Auditorium, Amphitheater, Amphitheaters, Auditoriums, Concert Hall, Concert Halls, Music Hall, Music Halls, Aurora, Auroras, Australian Shepherd, Australian Shepherds, Automobile, Automobiles, Autos, Car, Cars, Motorcar, Motorcars, Avocado, Aguacate, Aguacates, Alligator Pear, Alligator Pears, Avocados, Axe, Ax, Axes, Baby, Babe, Babes, Babies, Infant, Infants

Science Fiction and Science Fact



Are you ready for the test, Leon?

Science Fiction and Science Fact



Science Fiction and Science Fact



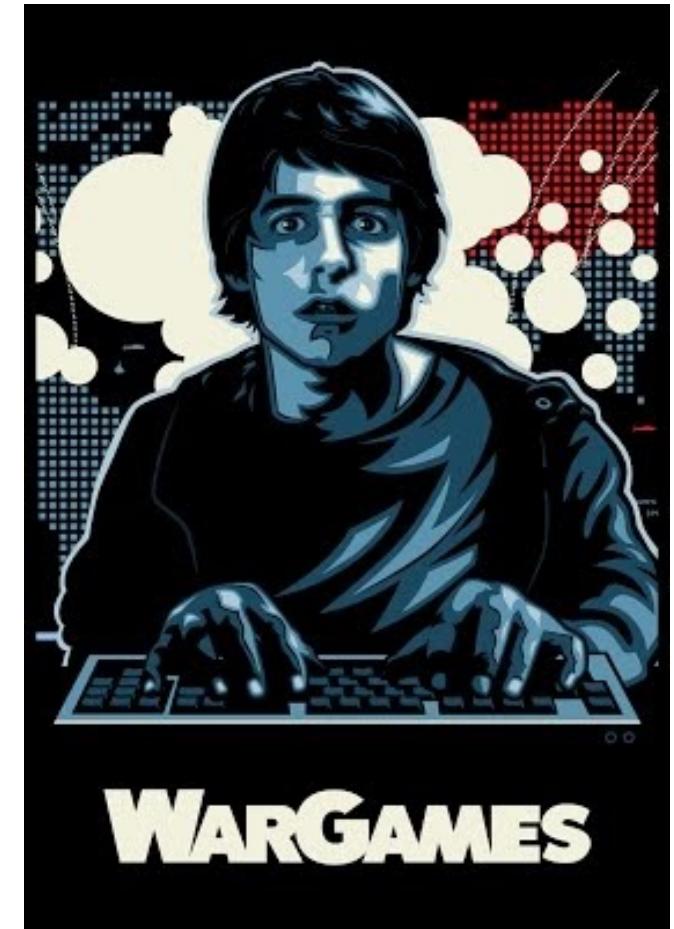
Select all squares with traffic lights

Science Fiction and Science Fact



Joke Credit: James Dennis

Science Fiction and Science Fact



<https://www.youtube.com/watch?v=s93KC4AGKnY>



THINK

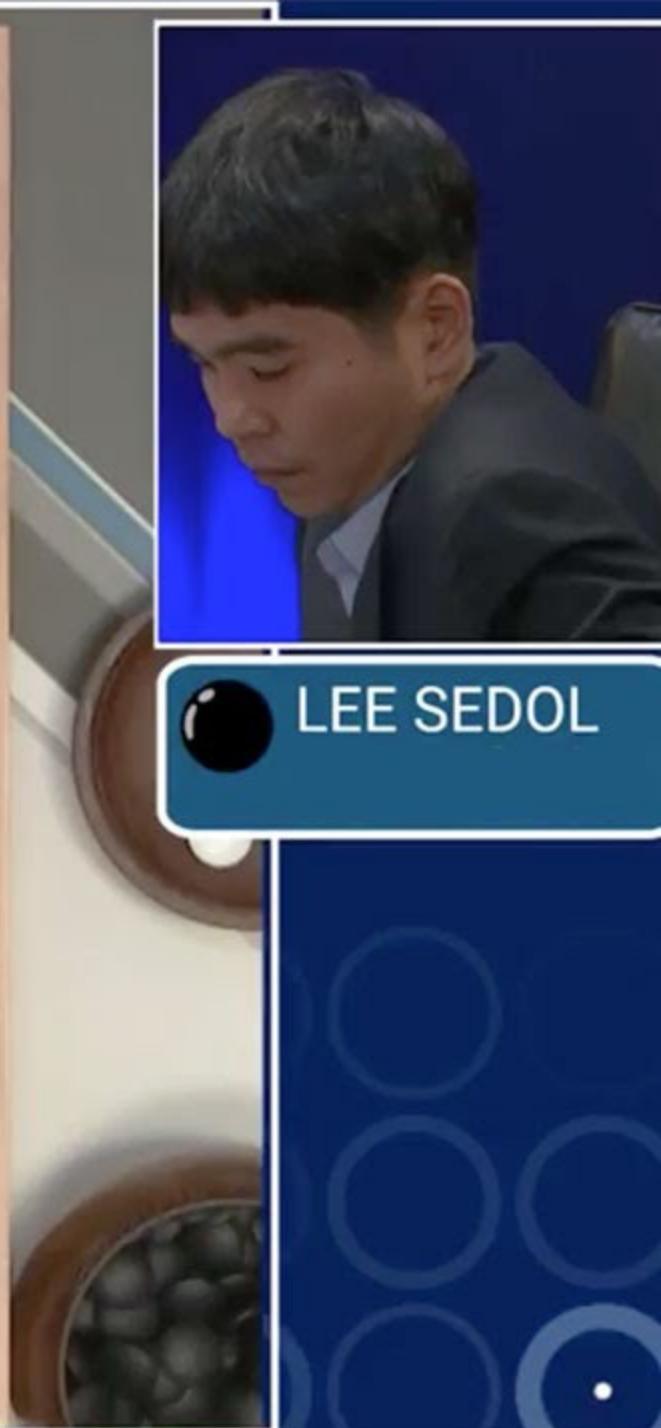
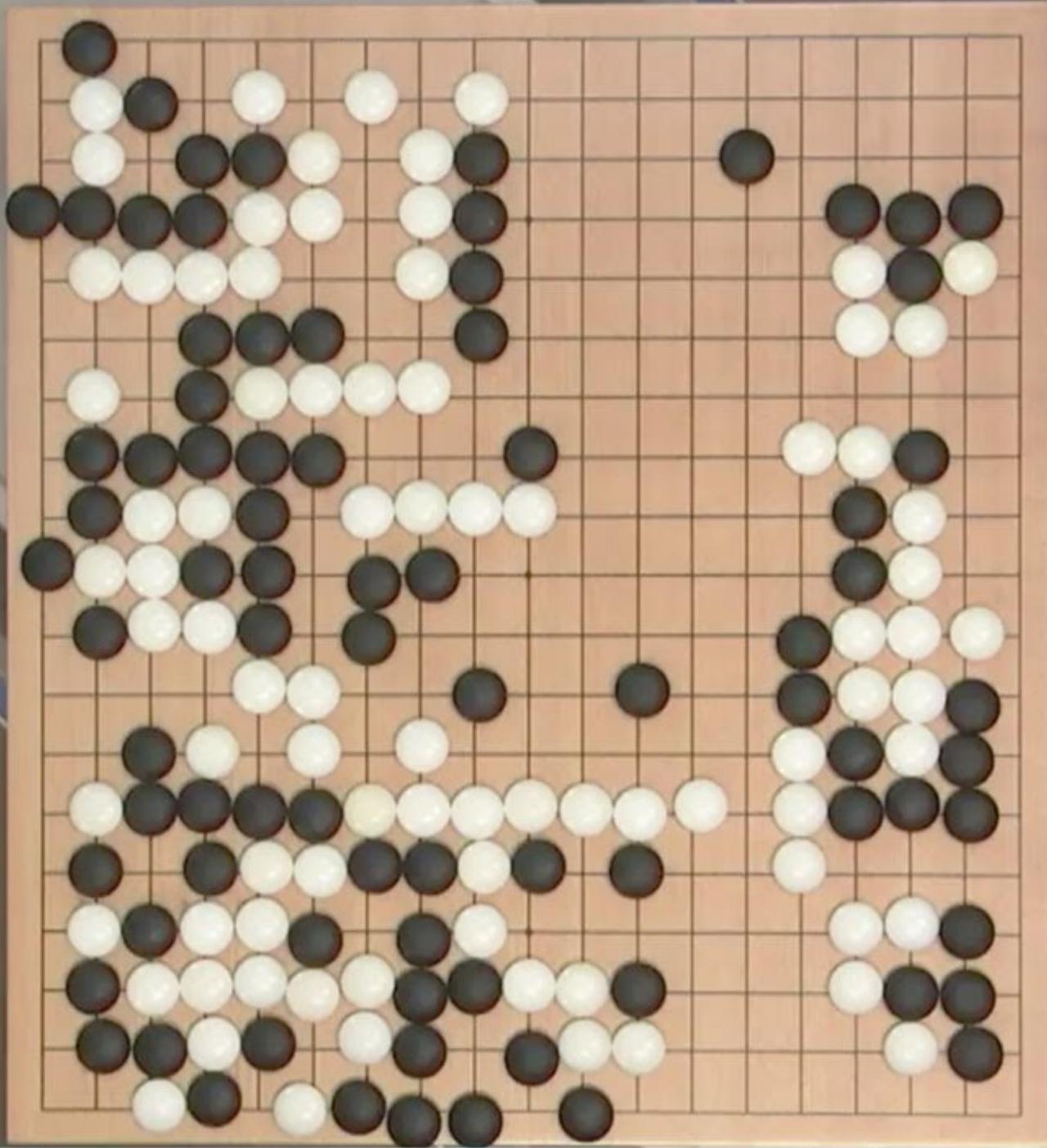
सोचिए



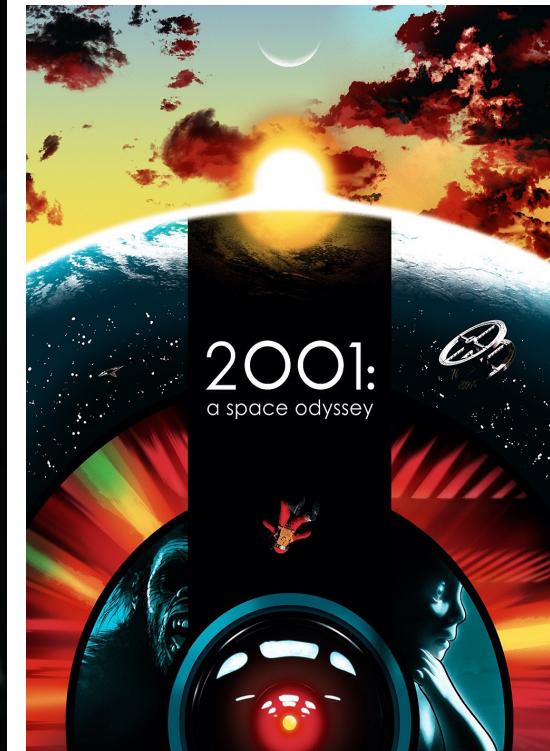
\$1,400

\$4,000

Ken



Science Fiction and Science Fact

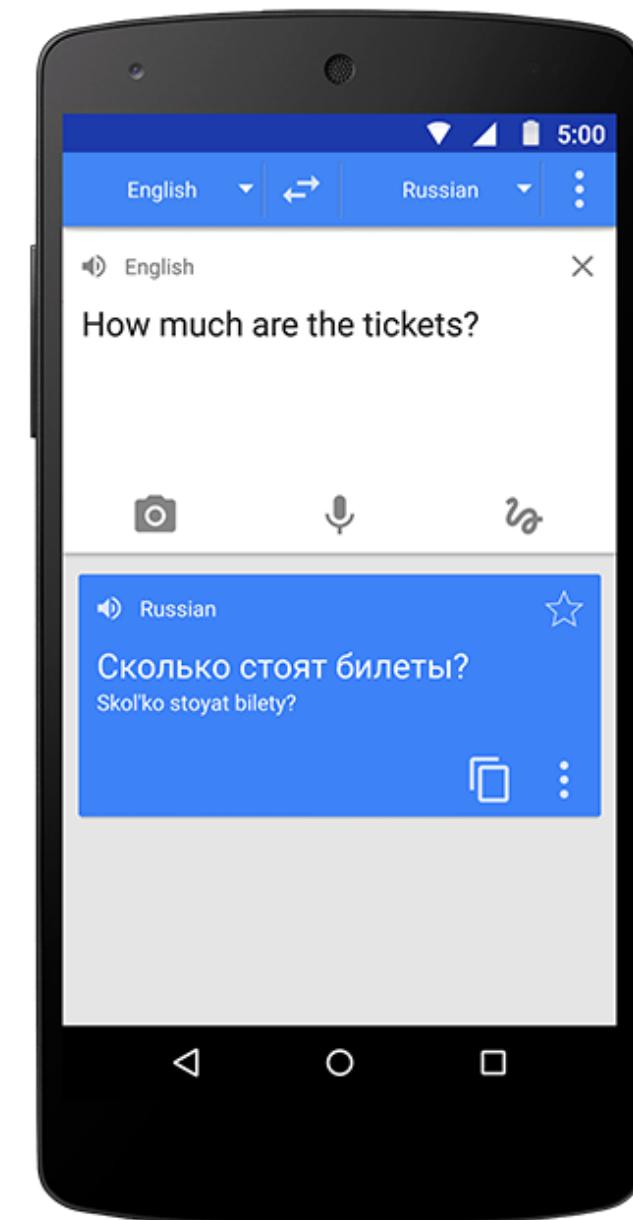


<https://www.youtube.com/watch?v=Lk8qgMFqaY8>













Science Fiction and Science Fact



DARPA
Grand Challenge 2005

Science Fiction and Science Fact



DARPA Urban Challenge 2007



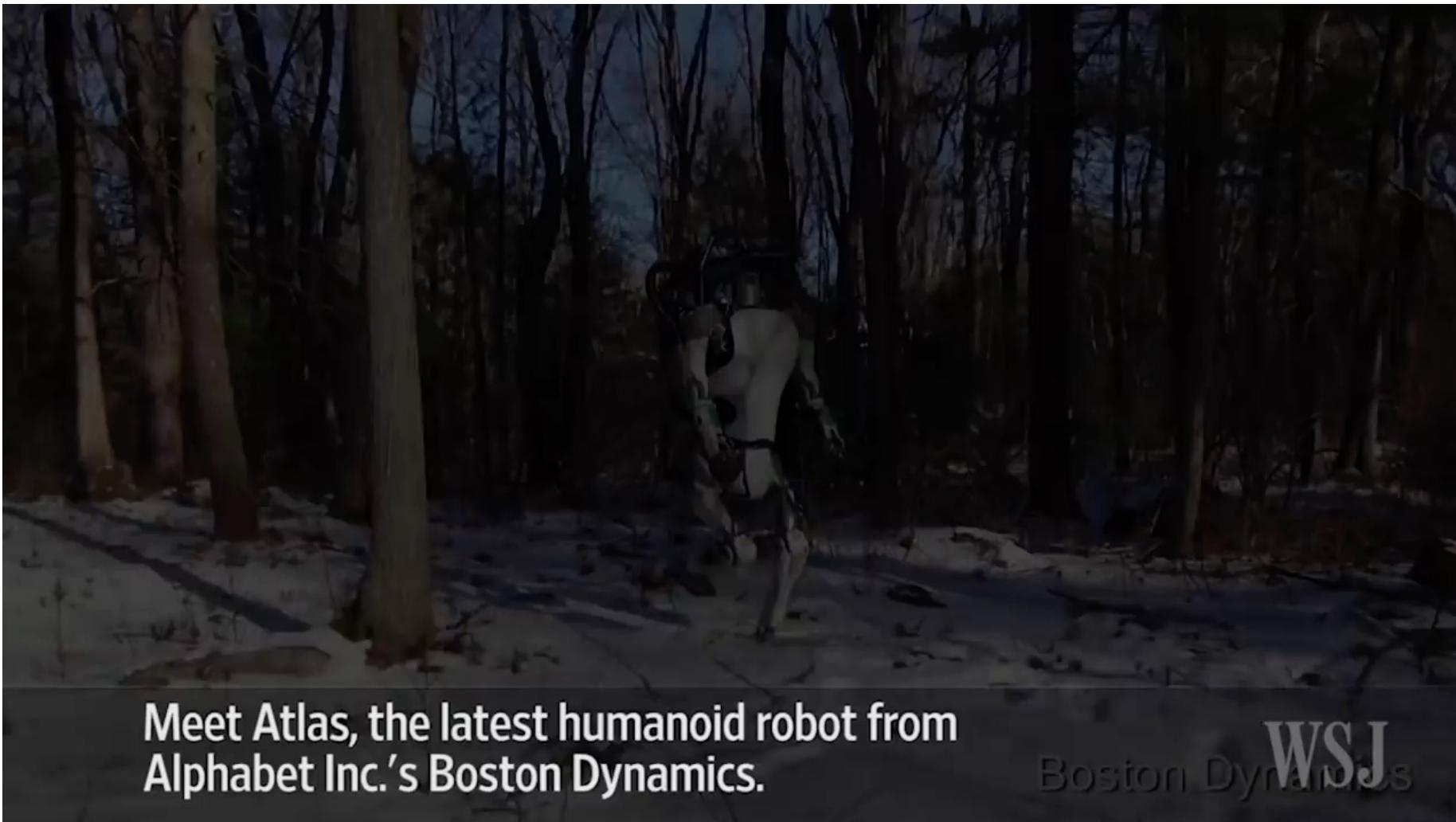
Science Fiction and Science Fact



DARPA Robotics Challenge 2015

<https://www.youtube.com/watch?v=g0TaYhjpOfo>

Science Fiction and Science Fact



Meet Atlas, the latest humanoid robot from
Alphabet Inc.'s Boston Dynamics.

Boston Dynamics **WSJ**

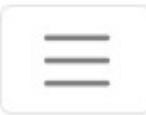
Boston Dynamics Atlas 2016

<https://www.youtube.com/watch?v=M91ISnATDQY>

For Fun: Try a Turning Test from my PhD students

Real or Fake Text?

13 completed



Continuation of text:



Life doesn't work the way we desire it to.



It's not that the universe is against us.



Such naivety does not account for our future.



The rest of it works against us, and developing our character and direction will undoubtedly help us explain why."



- Hunter S. Thompson "The price of freedom is eternal vigilance."

Your task is to decide at what point (if at all) the text you see begins to be written by a computer. Please click the blue button **as soon as** you are confident. Don't be surprised if some excerpts are entirely human-written. You will be awarded 5 points if you guess correctly, with decreasing amount of points if you guess after the true boundary.

Human-Written Prompt:

The world isn't always compliant with our wants.

HW: Read Chapter 1 & 27

Please read:

- Chapter 1, Section 1.1
- Chapter 27.1 and 27.2 of the textbook.

The required readings will be posted on the course website in the schedule tab.

<http://artificial-intelligence-class.org/>

The website will have all lecture recordings, homework assignments, and course policies.



Artificial Intelligence
A Modern Approach
Fourth Edition



On your own:

A Brief Introduction to Python

The first Python programming HW has been released. It is due **Tuesday September 5th at 11:59pm**. Find it on the course Canvas.

