

CIS 421/521 Waitlist Update

We reached the max course enrollment of 150 students. There are an additional 300 students still on the waitlist.

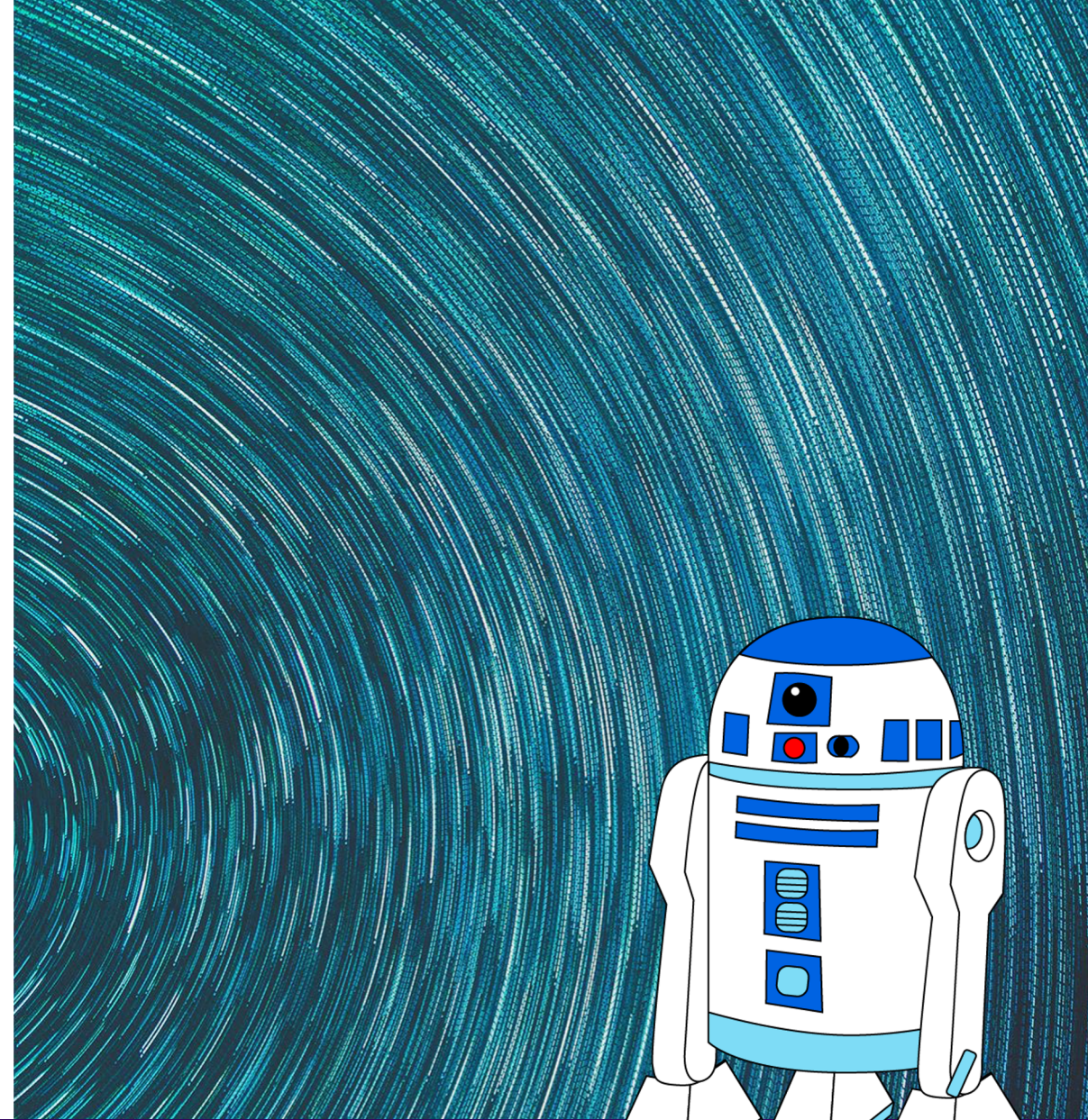
I'm going to offer another section of the class on MW from 8:30am-10am, which will **open an additional 200 slots.**

If you're in the room now and don't have a permit – please leave now and come to Heilmeier auditorium at 8:30am on Wednesday instead. Or watch the videos online.

If you are a graduate student who is unable to come to Penn this semester, you can also take the course online. I've created a special section for that. Email Lee Dukes to register for it:
Lee Dukes <ldukes@seas.upenn.edu>

CIS 421/521:
ARTIFICIAL INTELLIGENCE

Welcome to the Course!



Welcome to CIS 421/521

Professor Chris Callison-Burch (he/him)

- Office hours are Fridays from 2pm-3pm via Zoom
- Preferred method of contact: Piazza

My Research Focus:

- Natural Language Processing
- How can we build artificial intelligence that understands human languages?
- Translation between different languages
- Natural language inference

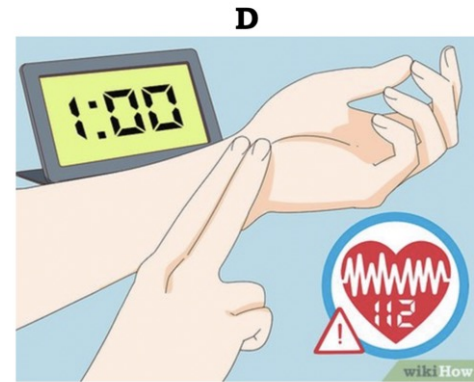
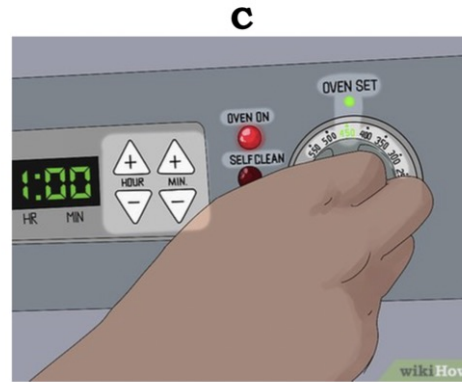
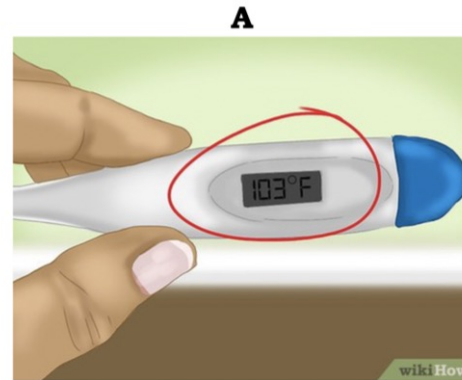
About me: I've got two kids ages 4.5 and 8. I like to listen to roleplaying game podcasts. I want to build an AI that can play Dungeons and Dragons.



Yue Yang - Head TA



How to Bake White Fish ?



Course staff



Amy Guo (she/her)



Anna Orosz (she/her)



Artemis Panagopoulou
(she/her)



Coco Zhao (she/her)



Etan Ginsberg (he/him)



Hanbang Wang (he/him)



Harry Wang (he/him)



Henrique Martins dos Santos
(he/him)



John Wallison (he/him)



Kit Wiggin (he/him)



Lisa Zhao (she/her)



Samar Haider (he/him)



Welcome to CIS 421/521

- Course web page: <http://artificial-intelligence-class.org>
 - Lecture slides on web page
 - Homeworks on web page
- Discussion on Piazza (link on course home page)
- Homework submission via Gradescope
- You do not have to come to class
 - Lectures will be recorded and posted after lecture
 - Pre-recorded videos are also available!
- Prerequisites:
 - Good knowledge of programming, data structures
 - Enough programming experience to *master* Python after introductory lecture
 - Introductory probability and statistics, and linear algebra will be *very useful*

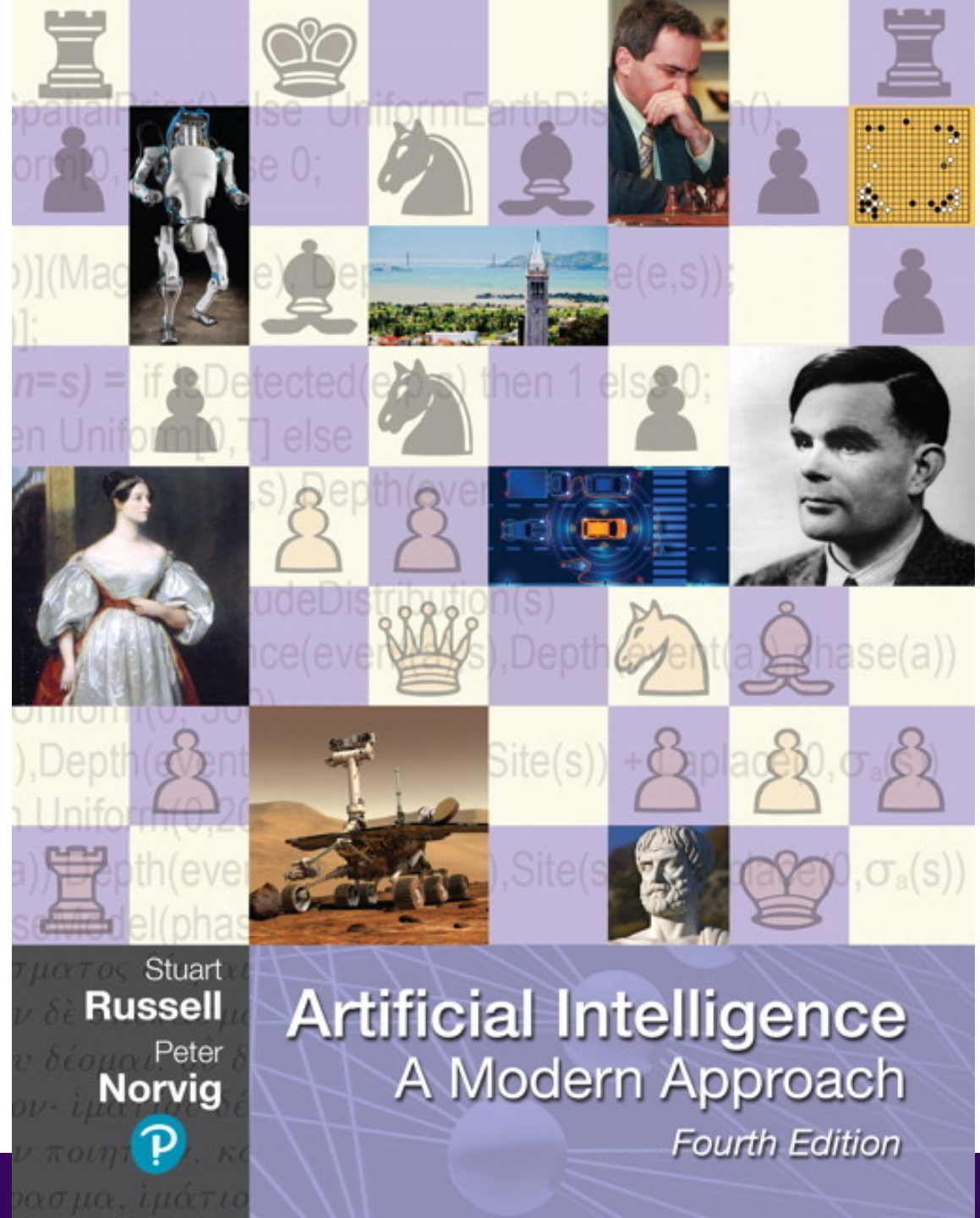
Course Textbook

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.



Grading and Homework

- Grading:
 - 70% for homework assignments
 - 30% for exams and quizzes
 - Up to about 3-5% for optional Extra Credit Projects
- Homework:

There is roughly one homework assignment per week, aside from weeks with exams. Students enrolled in CIS 421 may skip one HW assignment, or they may discard their lowest scoring HW assignment. You do not get late days back on the homework that you discard. Students enrolled in CIS 521 must complete all HW assignments and cannot discard their lowest scoring assignment.

Homework will be due at 11:59pm Eastern on specified dates (usually Tuesdays).

Each student has 5 free “late days”. Homeworks can be submitted at most two days late. If you are out of late days, then you will not be able to get credit for subsequent late assignments. One “day” is defined as anytime between 1 second and 24 hours after the homework deadline.

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 Piazza. All code will be run through a plagiarism check. Suspicions of plagiarism will be adjudicated by the Office of Student Conduct.

For extra credit assignments, you may work in pairs

Last semester, I found someone selling assignment solutions

CIS 421/521 compared to other Penn courses

There are many courses at Penn related to Artificial Intelligence:

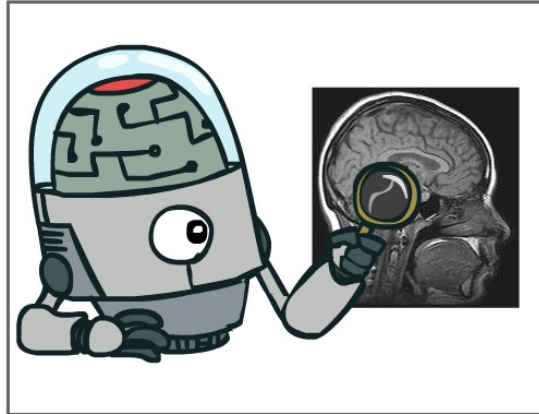
- CIS 419/519 – Applied Machine Learning
- CIS 520 – Machine Learning
- CIS 522 – Deep Learning
- CIS 530 – Computational Linguistics
- CIS 580 – Machine Perception
- MEAM 420/520 – Introduction to Robotics

CIS 421/521 offers a broad overview of AI so parts of it will overlap with other courses.

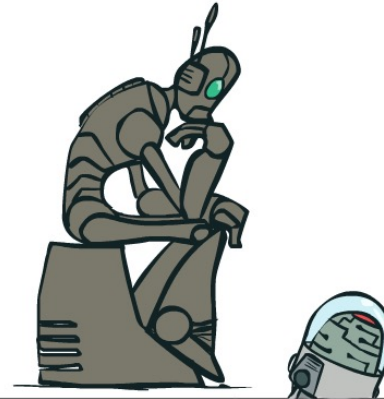
What is AI?

The science of making machines that:

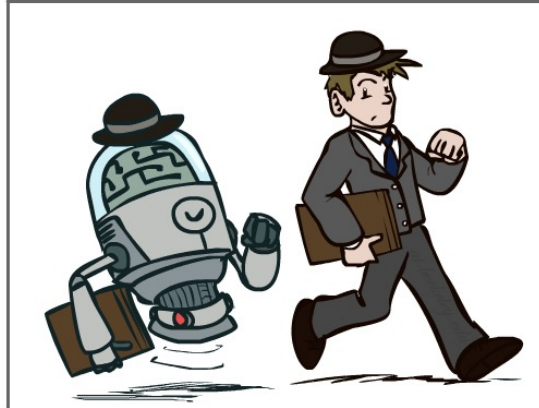
Think like people



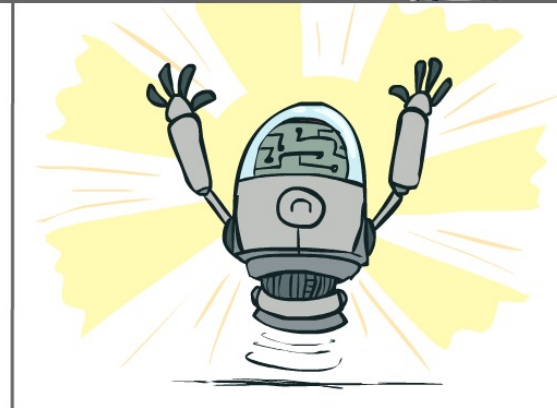
Think rationally



Act like people



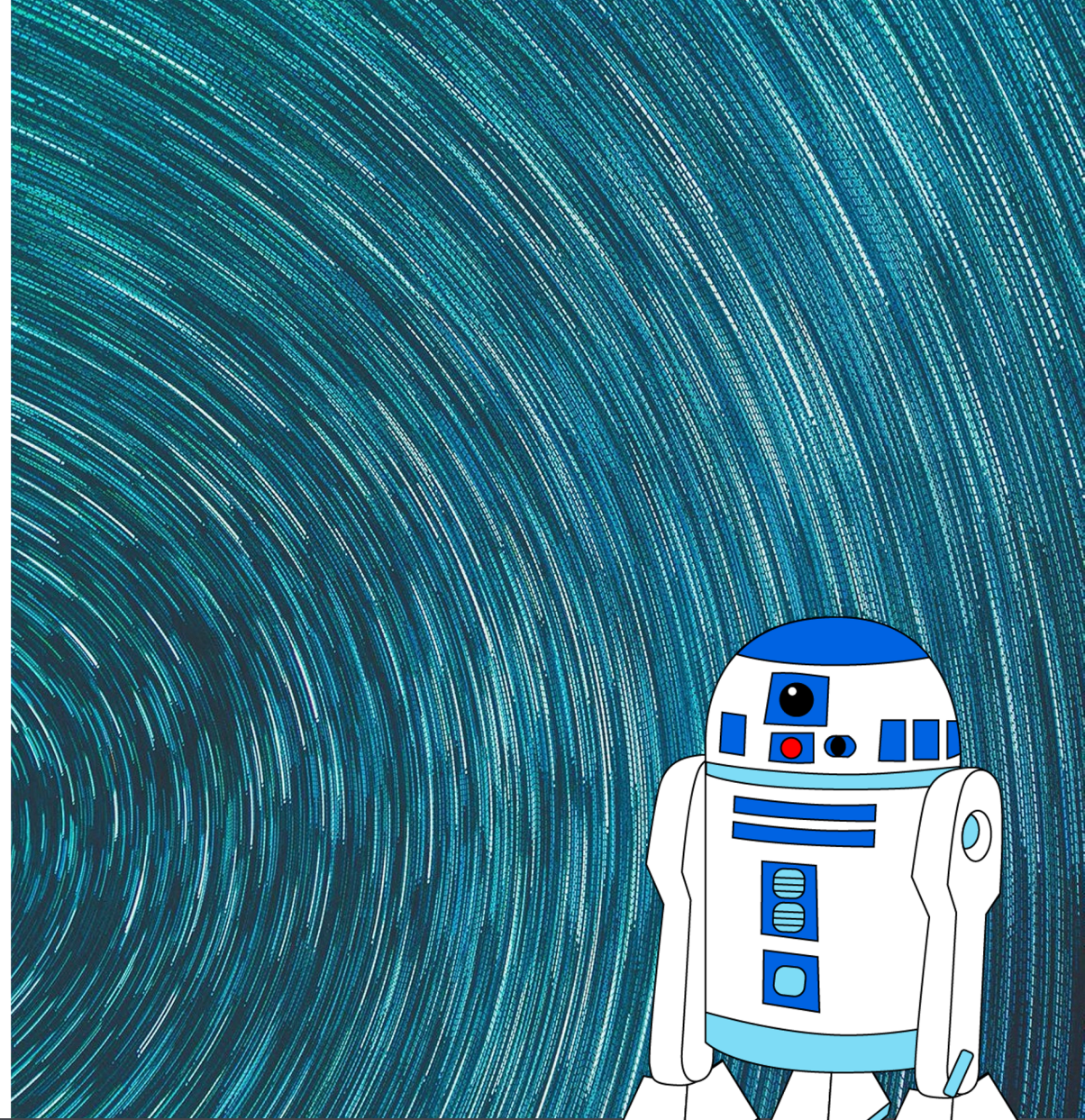
Act rationally



Slide from Dan Klein and Pieter Abbeel of UC Berkeley

CIS 421/521:
ARTIFICIAL INTELLIGENCE

Artificial Intelligence in Philosophy



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)

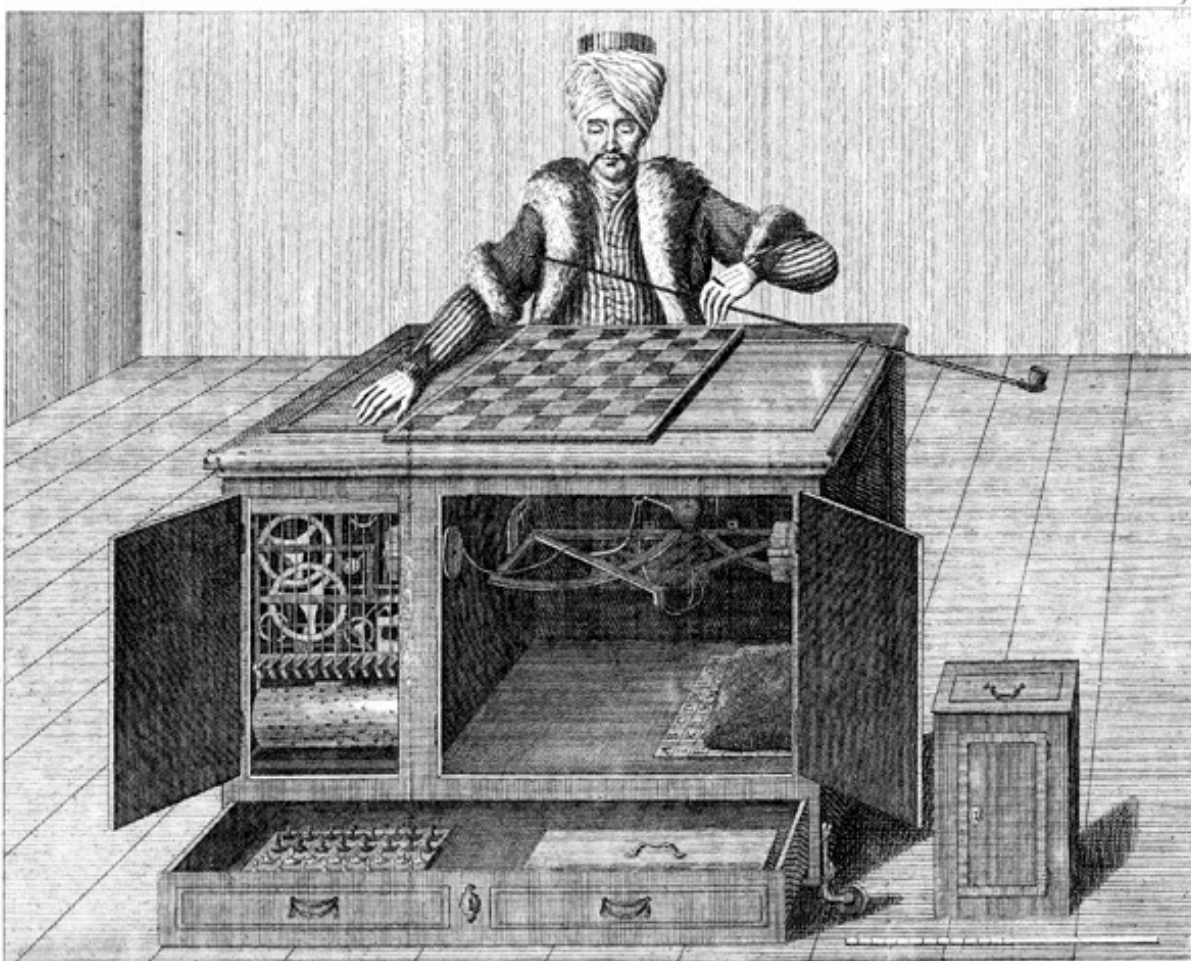


How can I know that you are not an automaton?

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 automatons



W. de Kempelen del. Che. à Mechel occid. Basilea. P.G. Pintz. sc.
Der Schach-Spieler, wie er vor dem Spiel gezeigt wird, von vorn. Le Joueur d'Échecs, tel qu'on le montre avant le jeu, par devant.



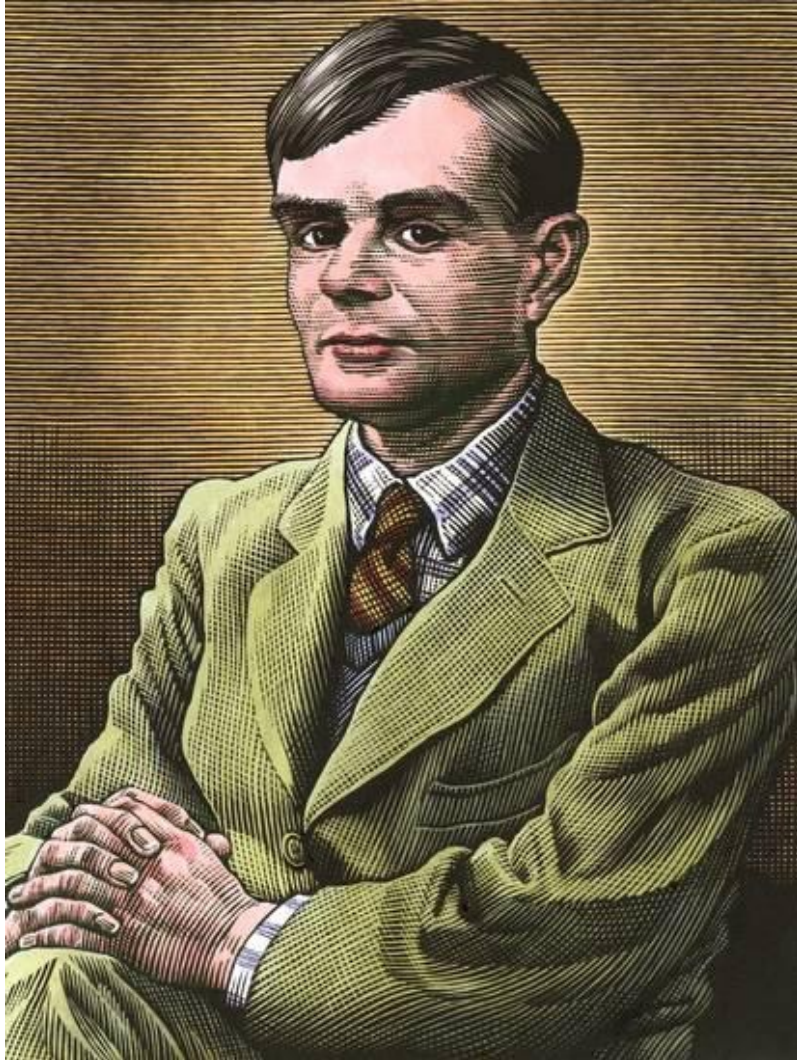
René Descartes (1596-1650)



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.

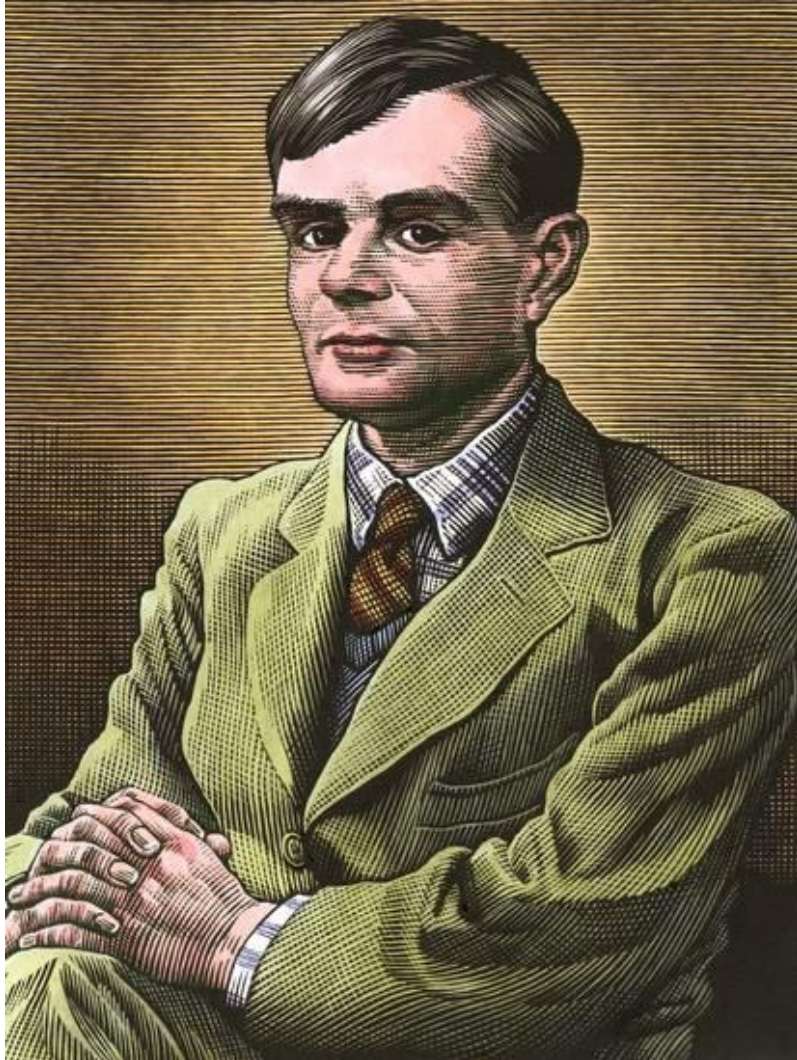
Alan Turing (1912-1954)



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?

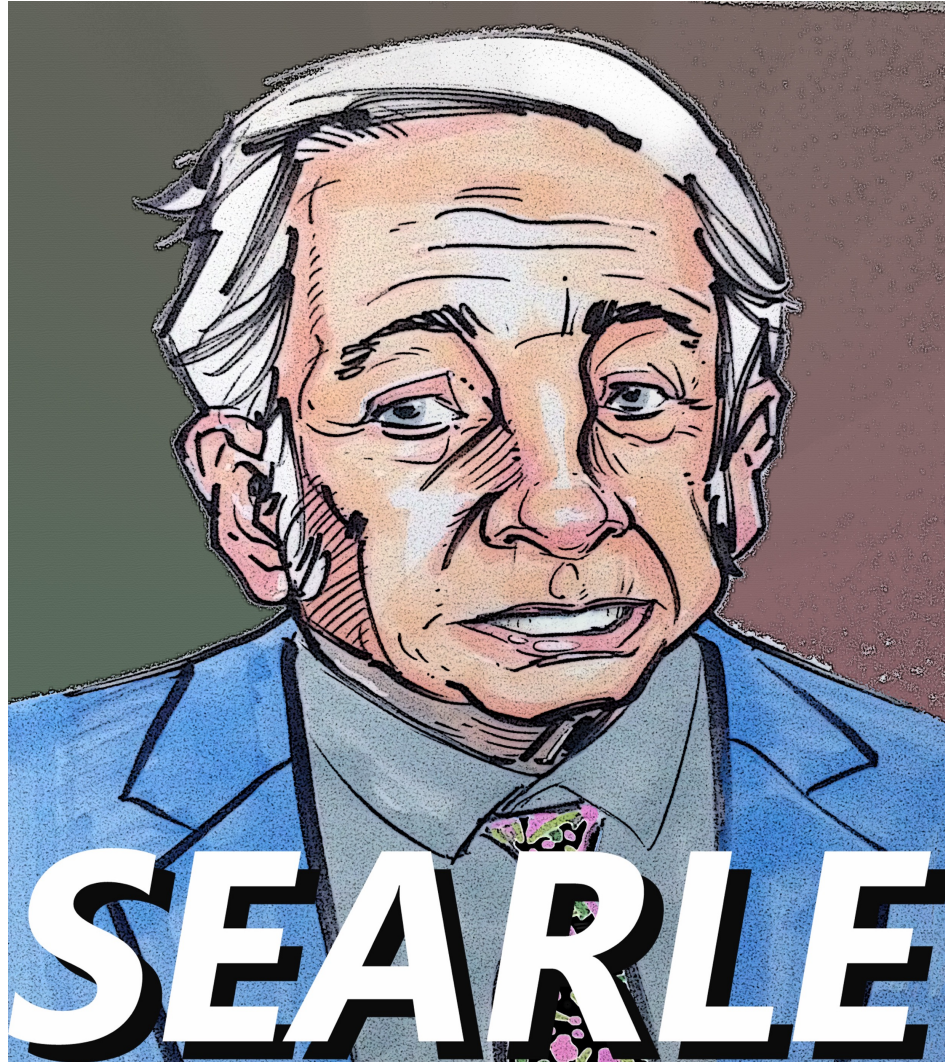
A human evaluator would judge text-based conversations between a human and a machine designed to generate human-like responses. If the evaluator cannot reliably tell the machine from the human, the machine passes the test. The test results do not depend on the ability to give correct answers to questions,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

Alan Turing (1912-1954)



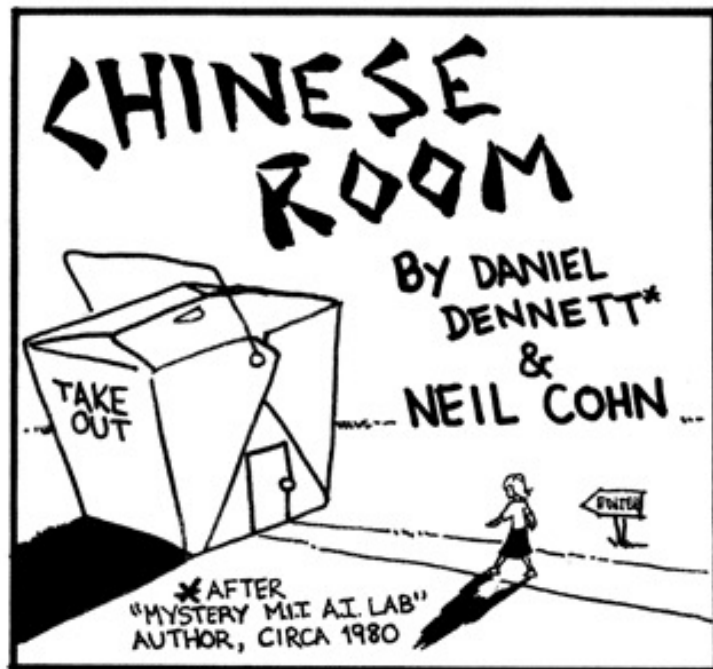
Turing was prosecuted in 1952 for being homosexual. He received chemical castration as an alternative to prison. Turing killed himself in 1954. It wasn't until the 2000s that Britain finally realized its gross injustice and its complicity in killing someone who should have been treated as a national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

```
Desktop — ssh nlpgrid.seas.upenn.edu
File Edit Options Buffers Tools Text Help
ELIZA: I am the psychotherapist. Please, describe your problems. Each time
you are finished talking, type RET twice.

YOU: I'm feeling very isolated.

ELIZA: Why do you say you are feeling very isolated?

YOU: I feel like I haven't left the house in a year.

ELIZA: What makes you believe that?

YOU: There's a global pandemic and I'm afraid to catch the virus.

ELIZA: Maybe you should consult a medical doctor; I am a psychotherapist. .

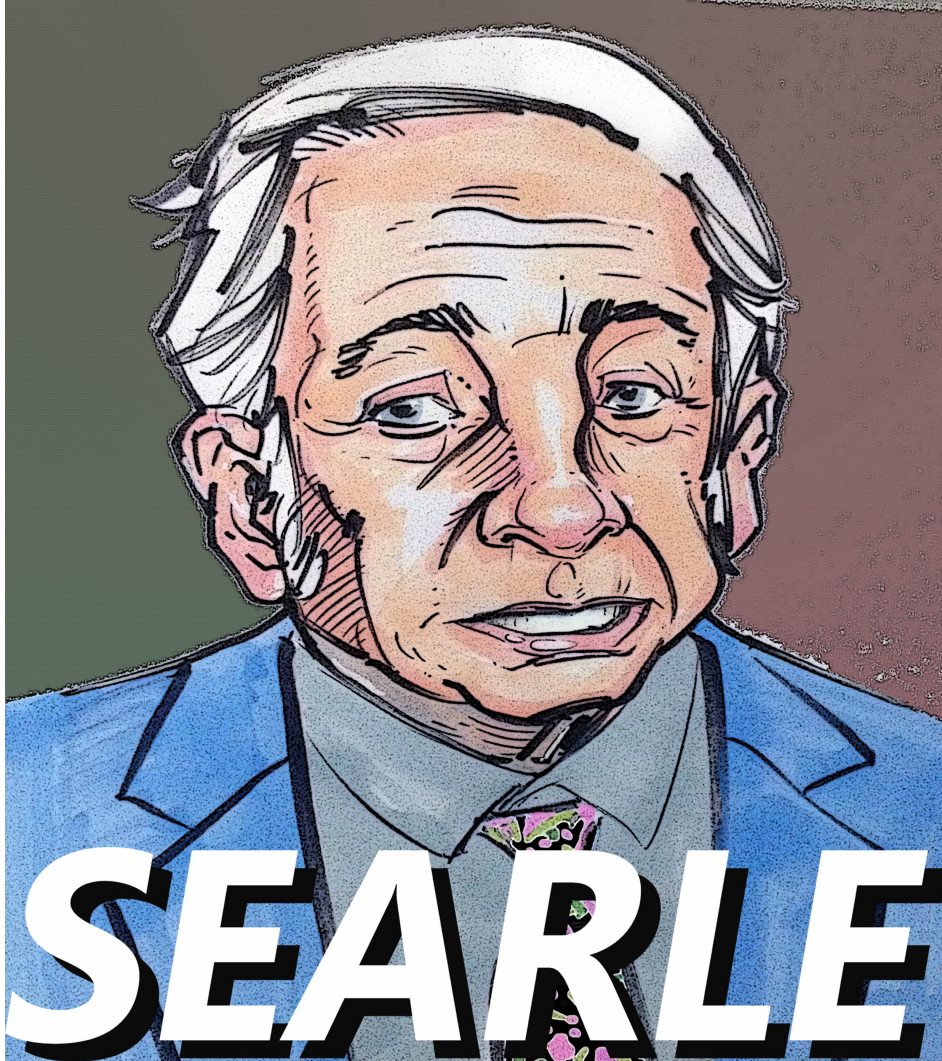
YOU: I feel sad that I cannot see my mother.

ELIZA: Tell me something about your family.

|

-UUU:**--F1 *doctor* All L20 (Doctor Fill) -----
```

John Searle (1932-) The Chinese Room

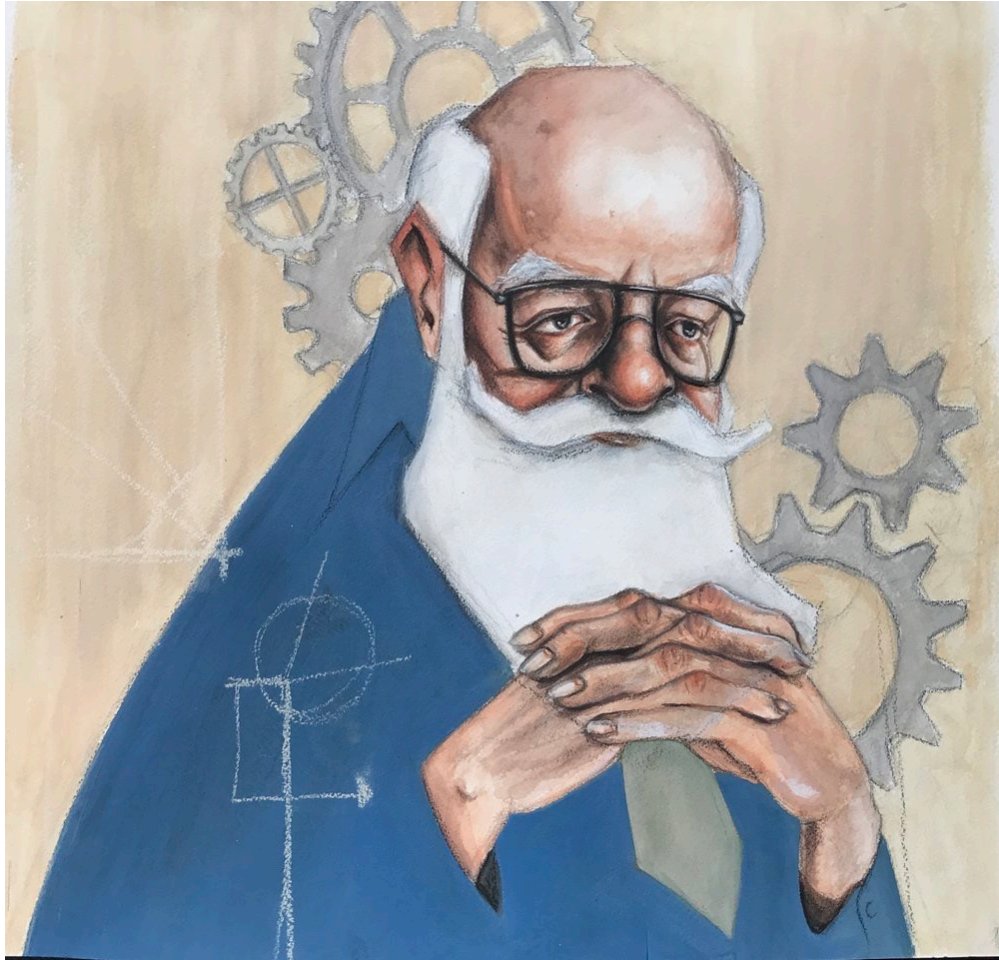


Brains cause minds, so computers can't have minds.

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.

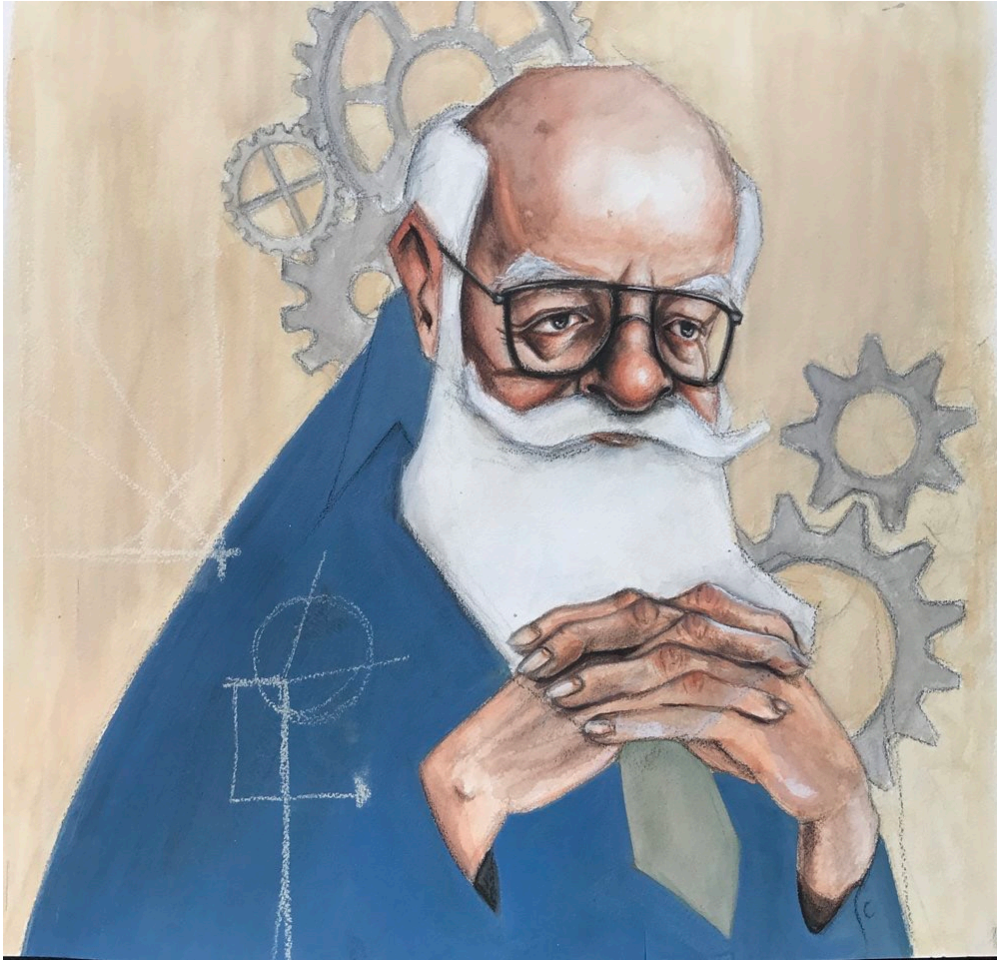
Daniel Dennett (1942-) Brain in a Vat



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

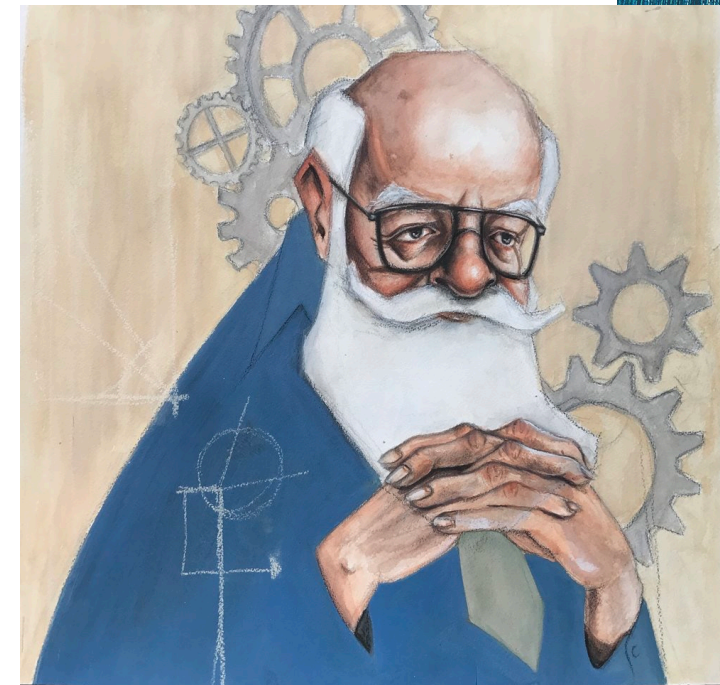
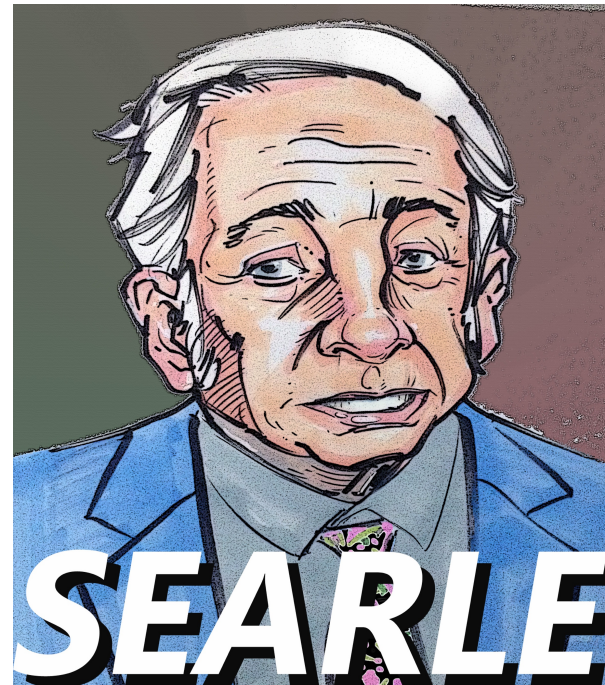
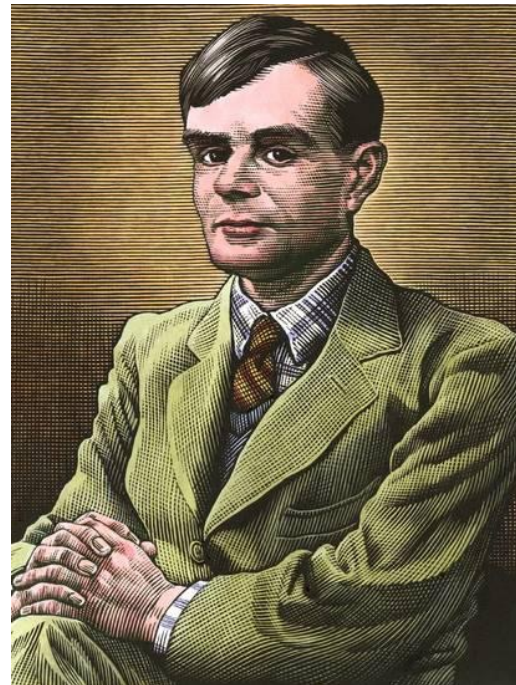
FILOSOFIX
Philosophy animates

<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.

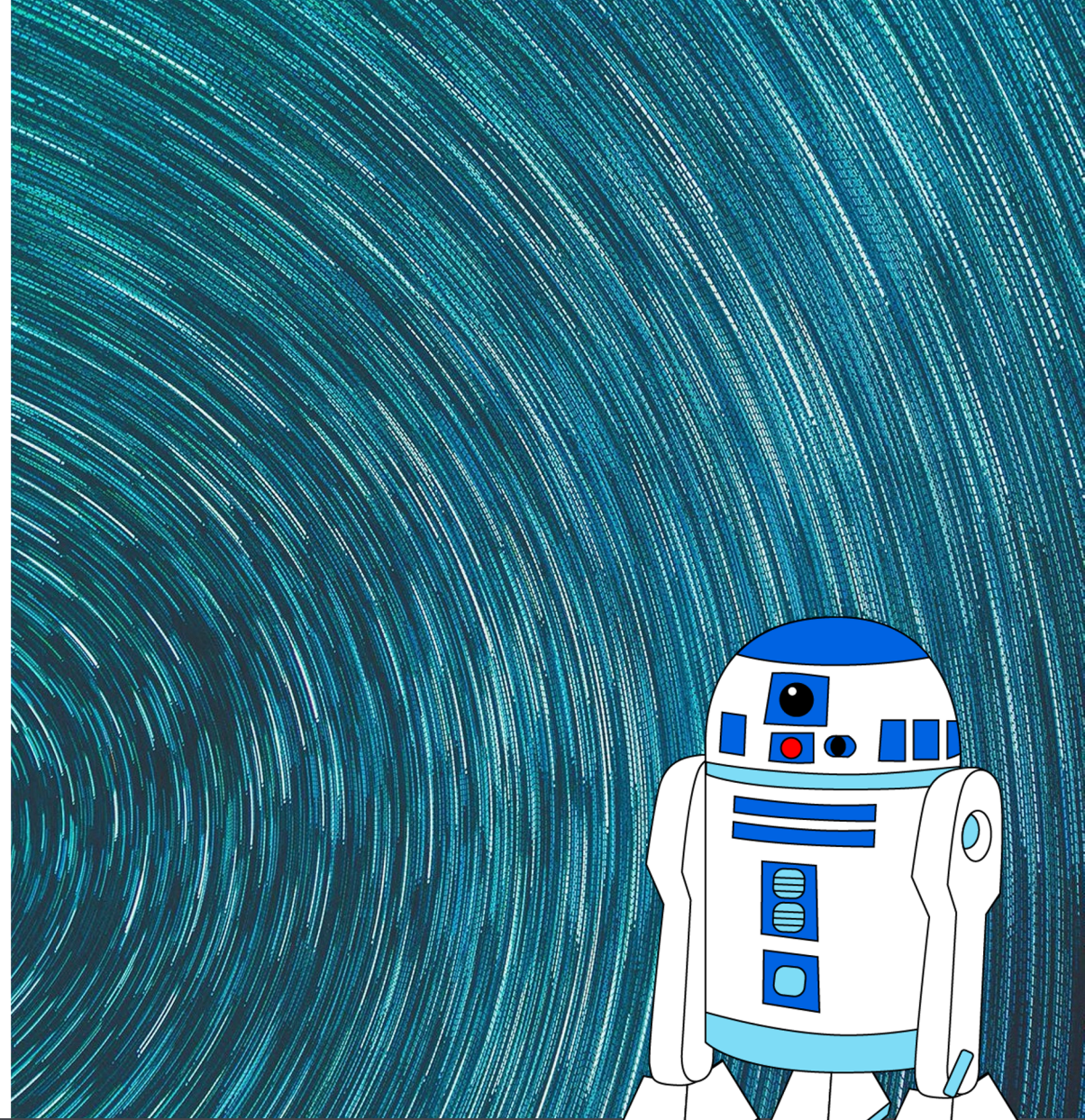
There's a lack of attention to the question of whether it is possible to build systems that are truly intelligent, as we commonly understand intelligence.

If you're interested in philosophical questions about AI, I recommend listening to Lex Fridman's podcast.



CIS 421/521:
ARTIFICIAL INTELLIGENCE

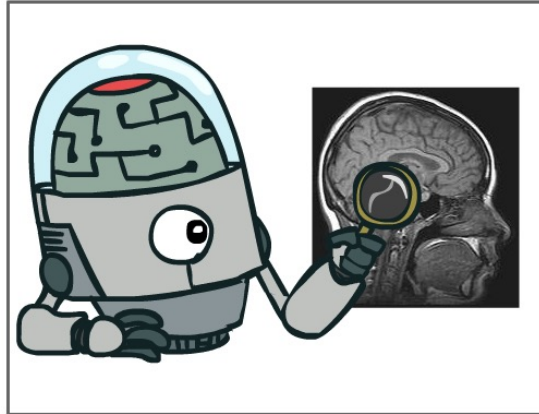
Artificial Intelligence in Science Fiction



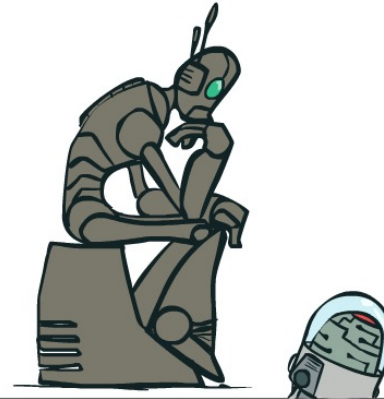
What is AI?

The science of making machines that:

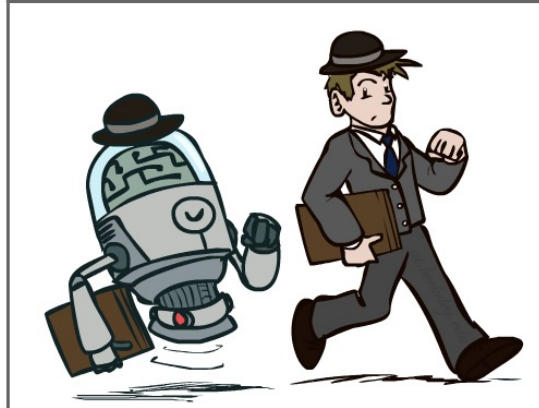
Think like people



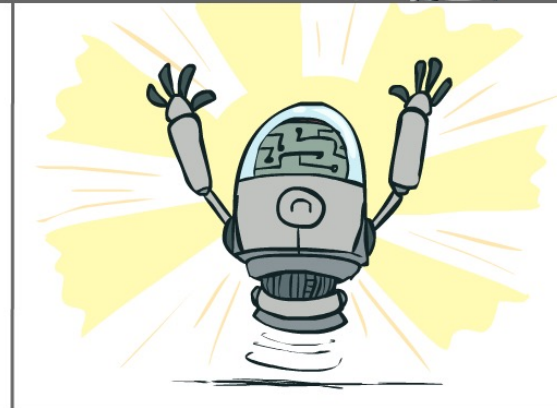
Think rationally



Act like people



Act rationally

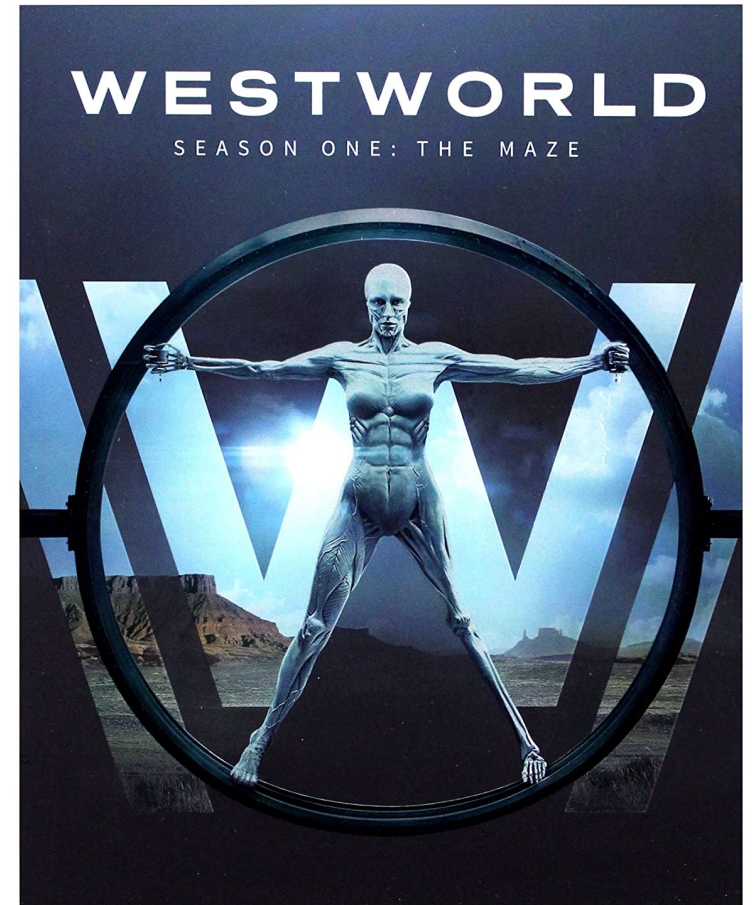


Slide from Dan Klein and Pieter Abbeel of UC Berkeley

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>

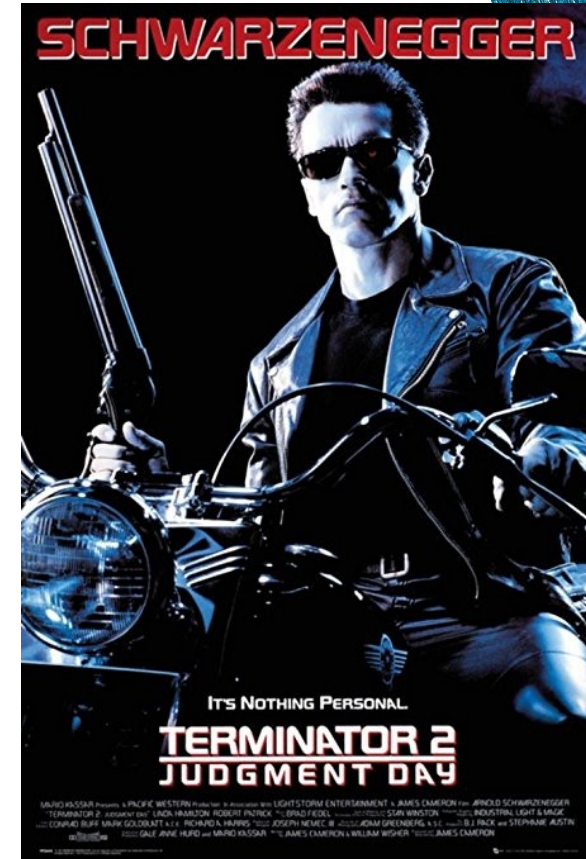
Descartes



Deckard

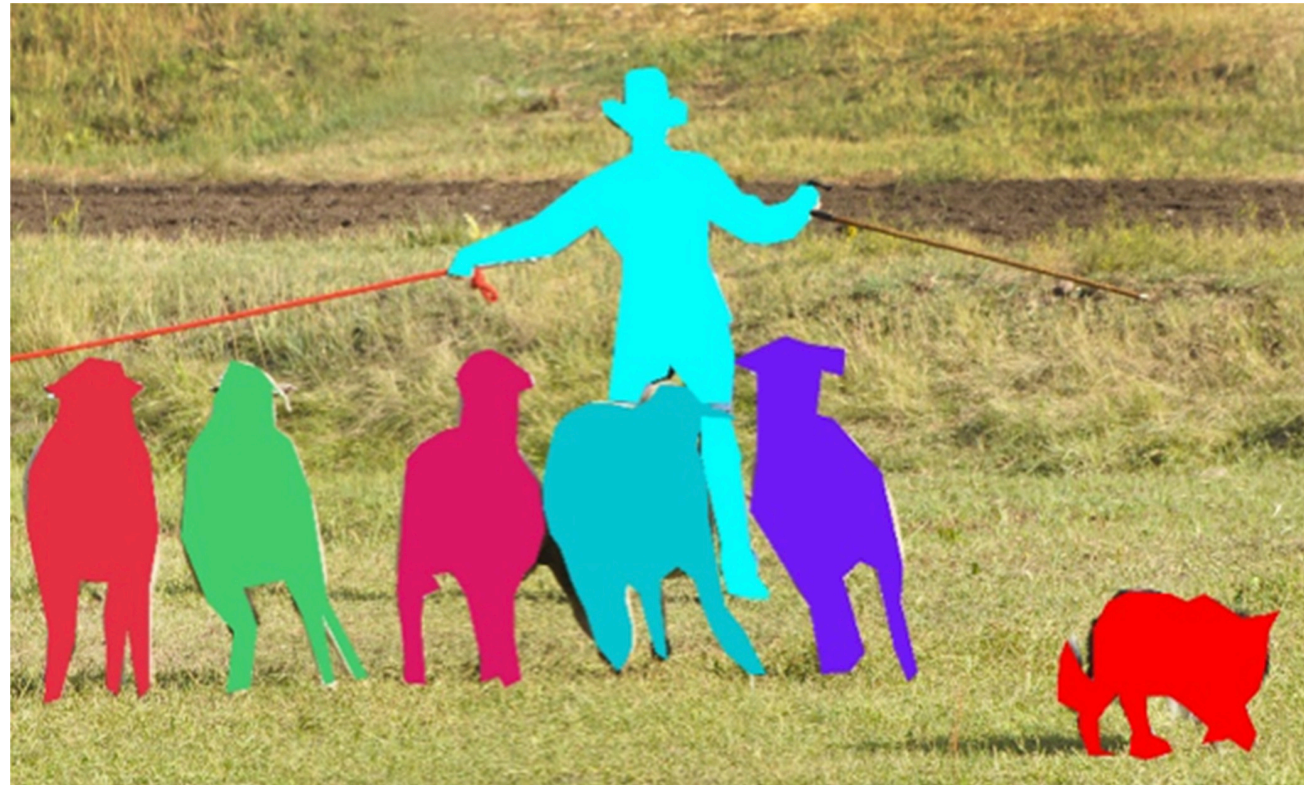


Science Fiction and Science Fact

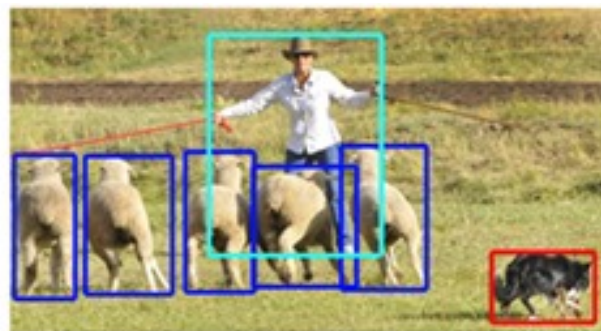


<https://www.youtube.com/watch?v=JknRYYrV5Vk>

Science Fiction and Science Fact



(a) classification



(b) detection



(c) segmentation

Science Fiction and Science Fact

Medium

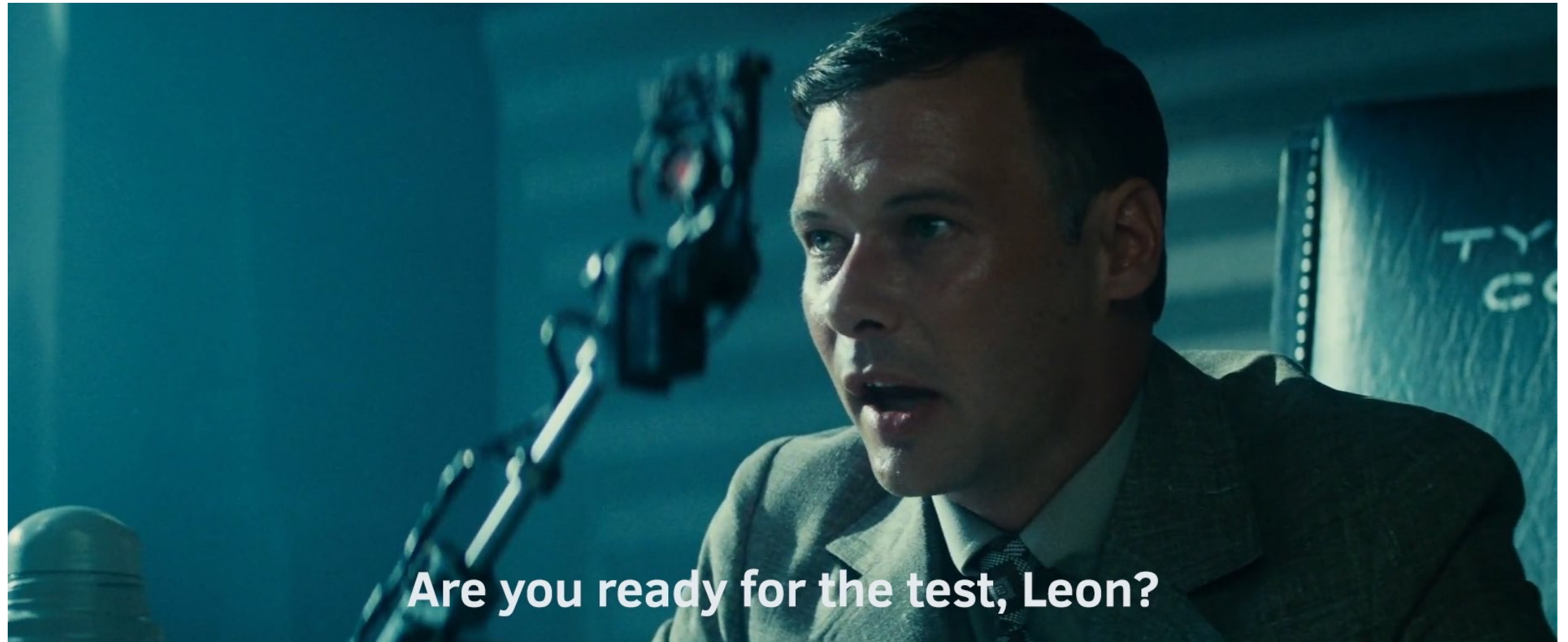


Kenny Yin [Follow](#)
Developer. Cinematographer. Vocalist.
Jun 19, 2016 · 23 min read

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

Armchair, Armchairs, Art, Artistic Creation, Artistic 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, Amphitheatres, Auditoriums, Concert Hall, Concert Halls, Music Hall, Music Halls, Aurora, Auroras, Australian Shepherd, Australian Shepherds, Automobile, Auto, Automobiles, Autos, Car, Cars, Motorcar, Motorcars, Avocado, Aguacate, Aguacates, Alligator Pear, Alligator Pears, Avocados, Axe, 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



Science Fiction and Science Fact



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





THINK

SMACMIS

INK

सोचिए



\$1,400

Iron

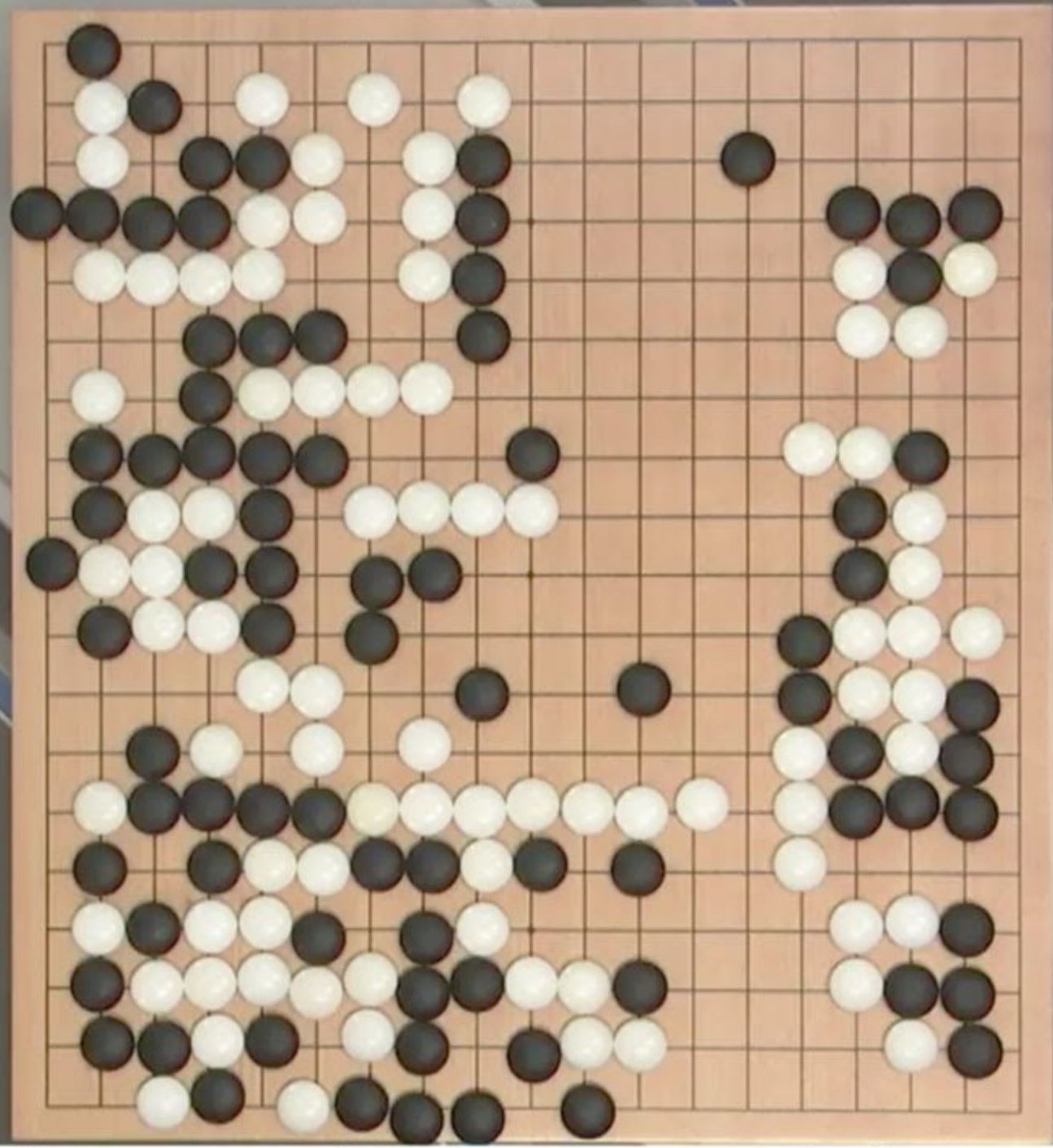
\$4,000

TABLE

● ALPHAGO



AlphaGo
Google DeepMind



● LEE SEDOL

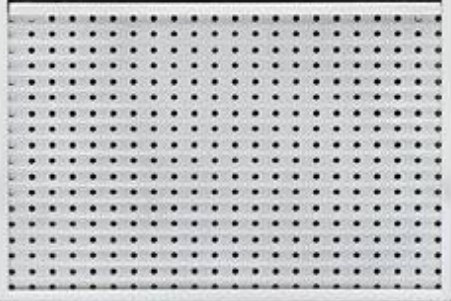


Science Fiction and Science Fact



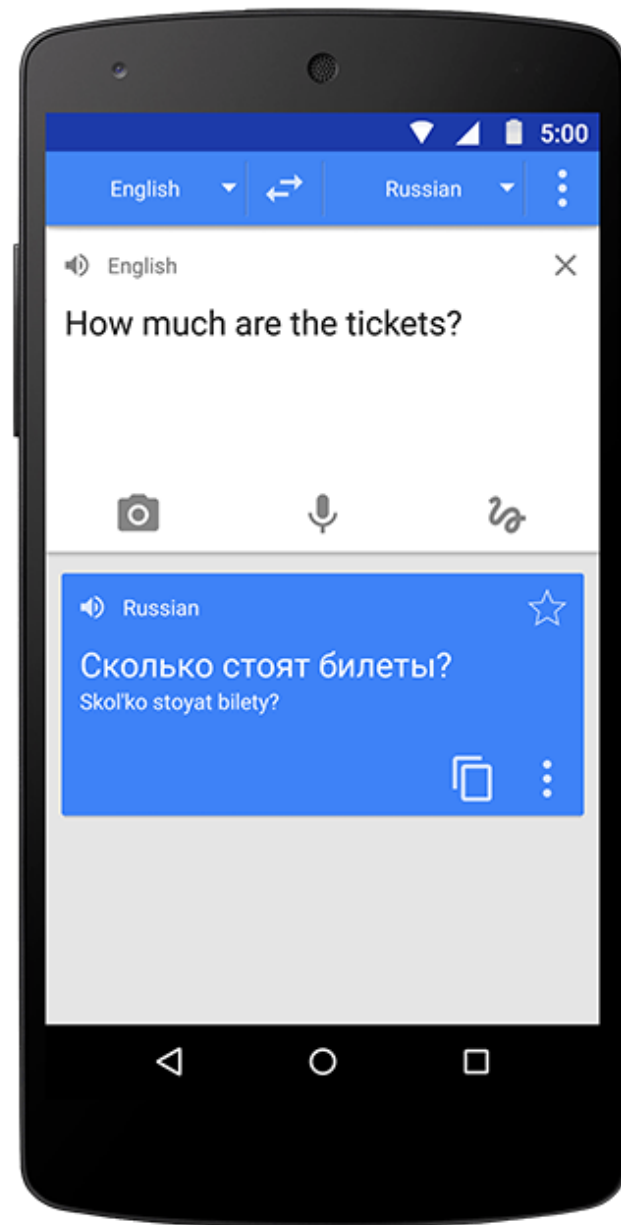
<https://www.youtube.com/watch?v=Lk8qgMFqaY8>

HAL 9000













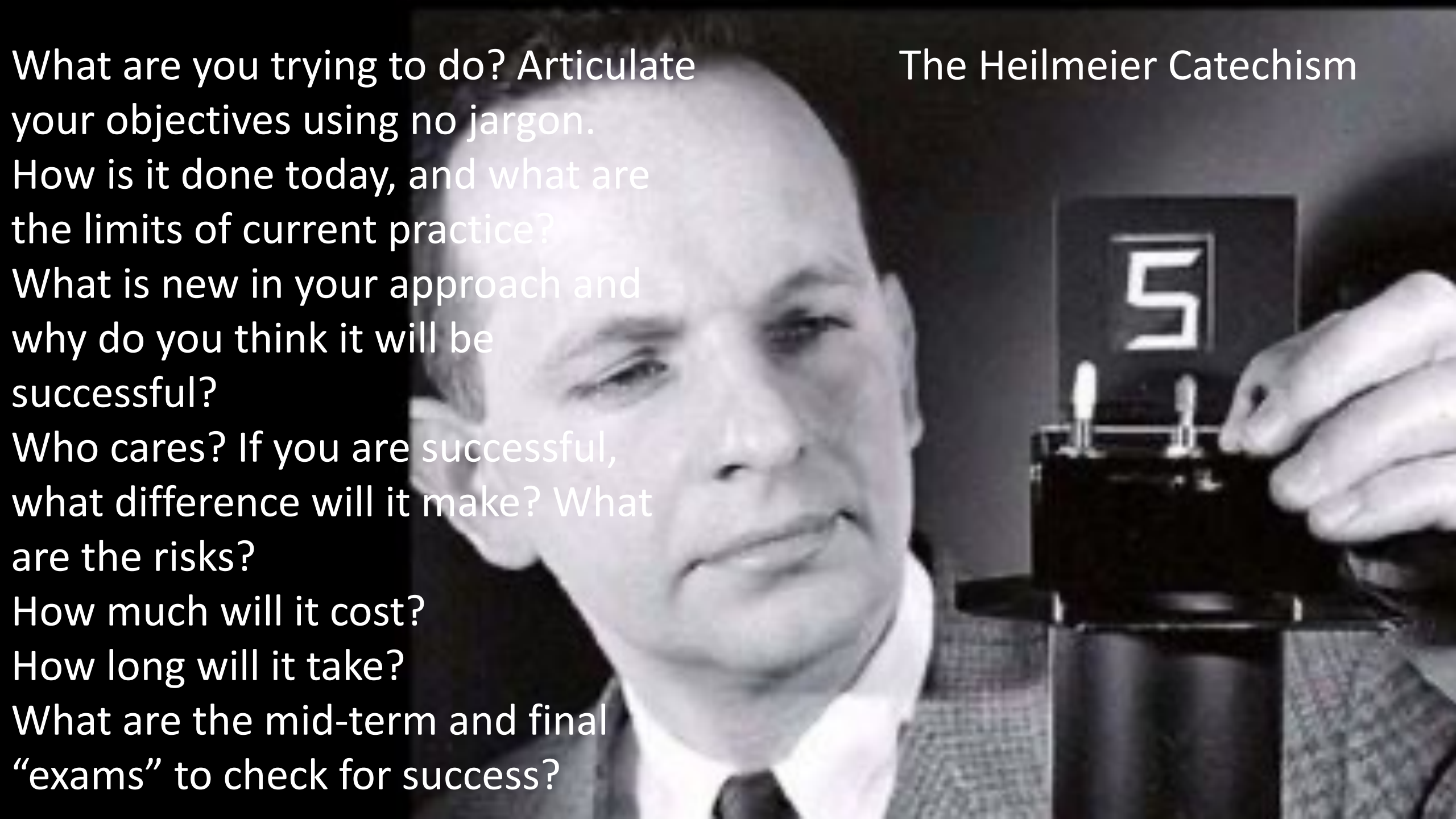


DARPA

DARPA

BREAKTHROUGH TECHNOLOGY

PAST | PRESENT | FUTURE



What are you trying to do? Articulate your objectives using no jargon.
How is it done today, and what are the limits of current practice?
What is new in your approach and why do you think it will be successful?
Who cares? If you are successful, what difference will it make? What are the risks?
How much will it cost?
How long will it take?
What are the mid-term and final “exams” to check for success?

The Heilmeier Catechism

Science Fiction and Science Fact



DARPA
Grand Challenge 2005

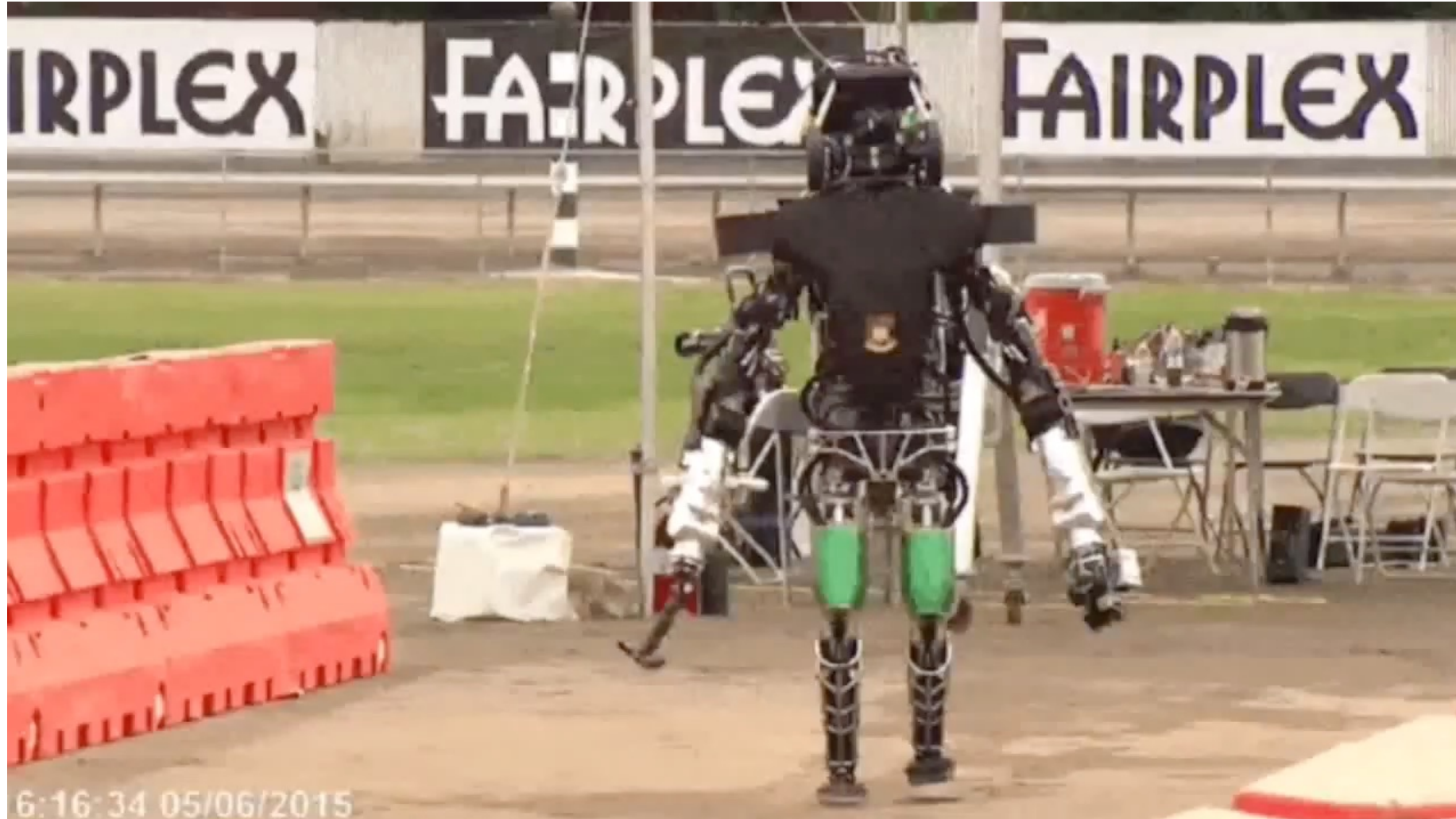
Science Fiction and Science Fact



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

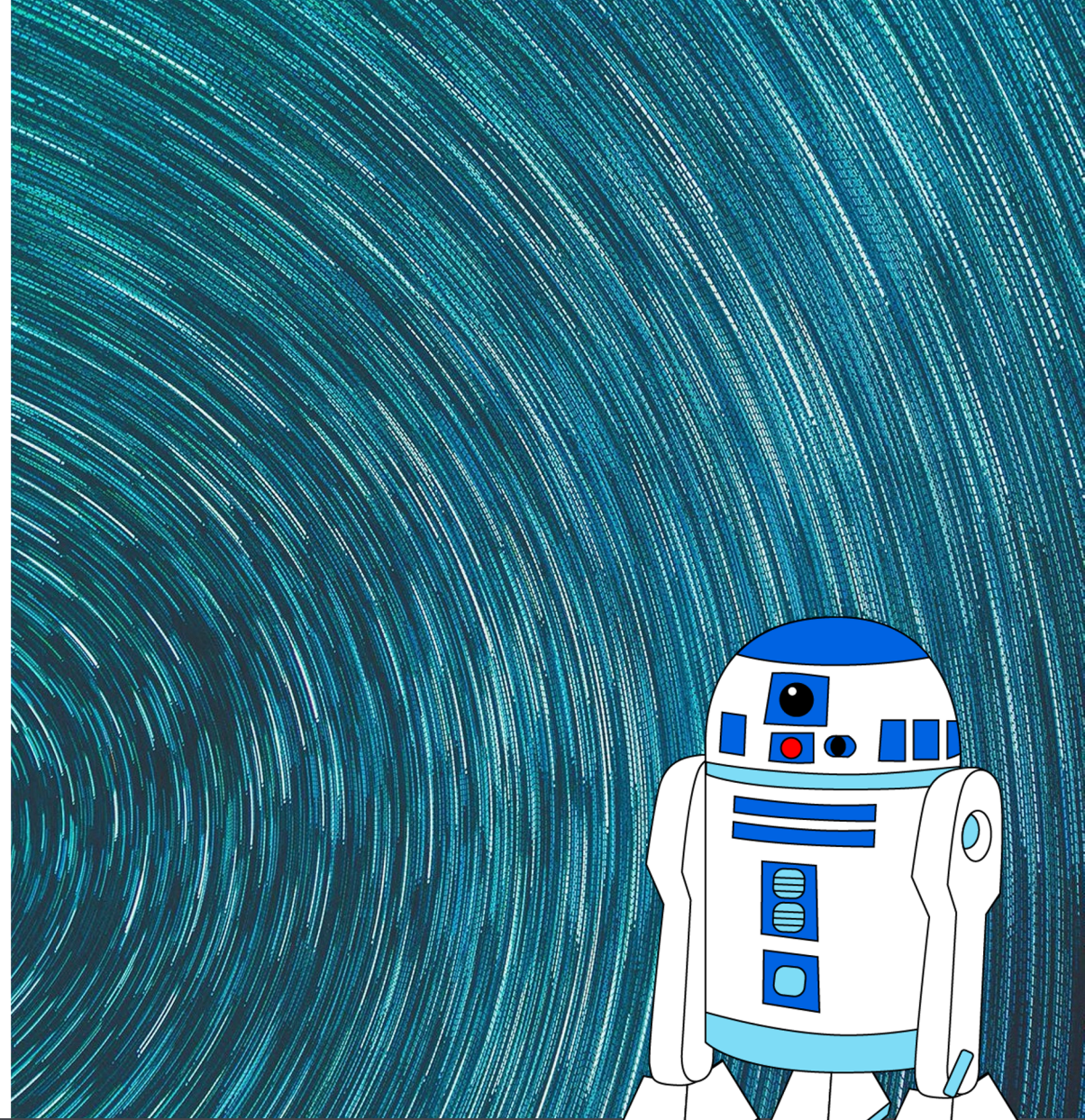


Boston Dynamics Atlas 2016

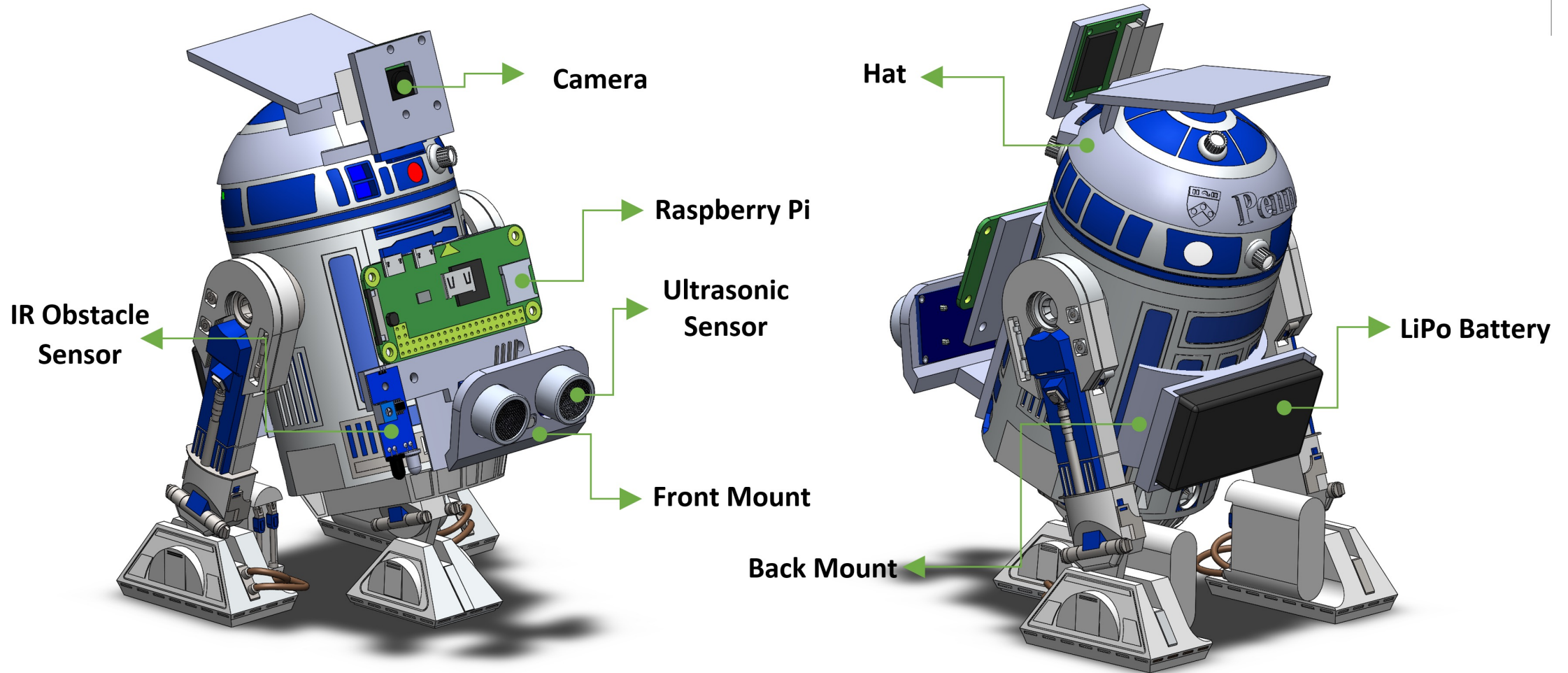
<https://www.youtube.com/watch?v=M91ISnATDQY>

CIS 421/521:
ARTIFICIAL INTELLIGENCE

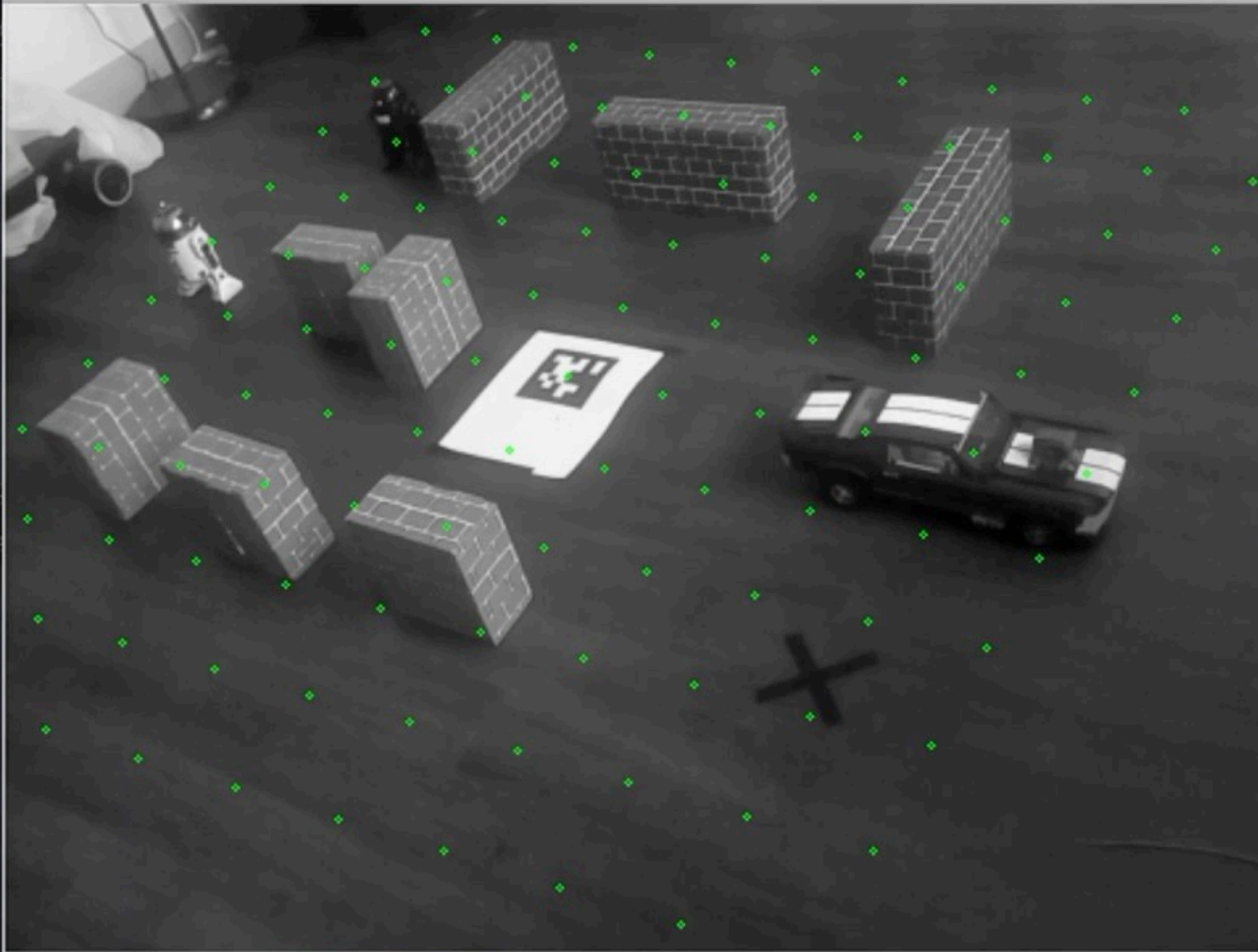
R2D2 Assignments



Programmable Toy Robots + Custom Sensor Packs



maze



(x=325, y=18) ~ R:63 G:63 B:63

Try a Turning Test from my PhD students

<https://roft.io/>

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."

Is the following written by a person or by a machine?

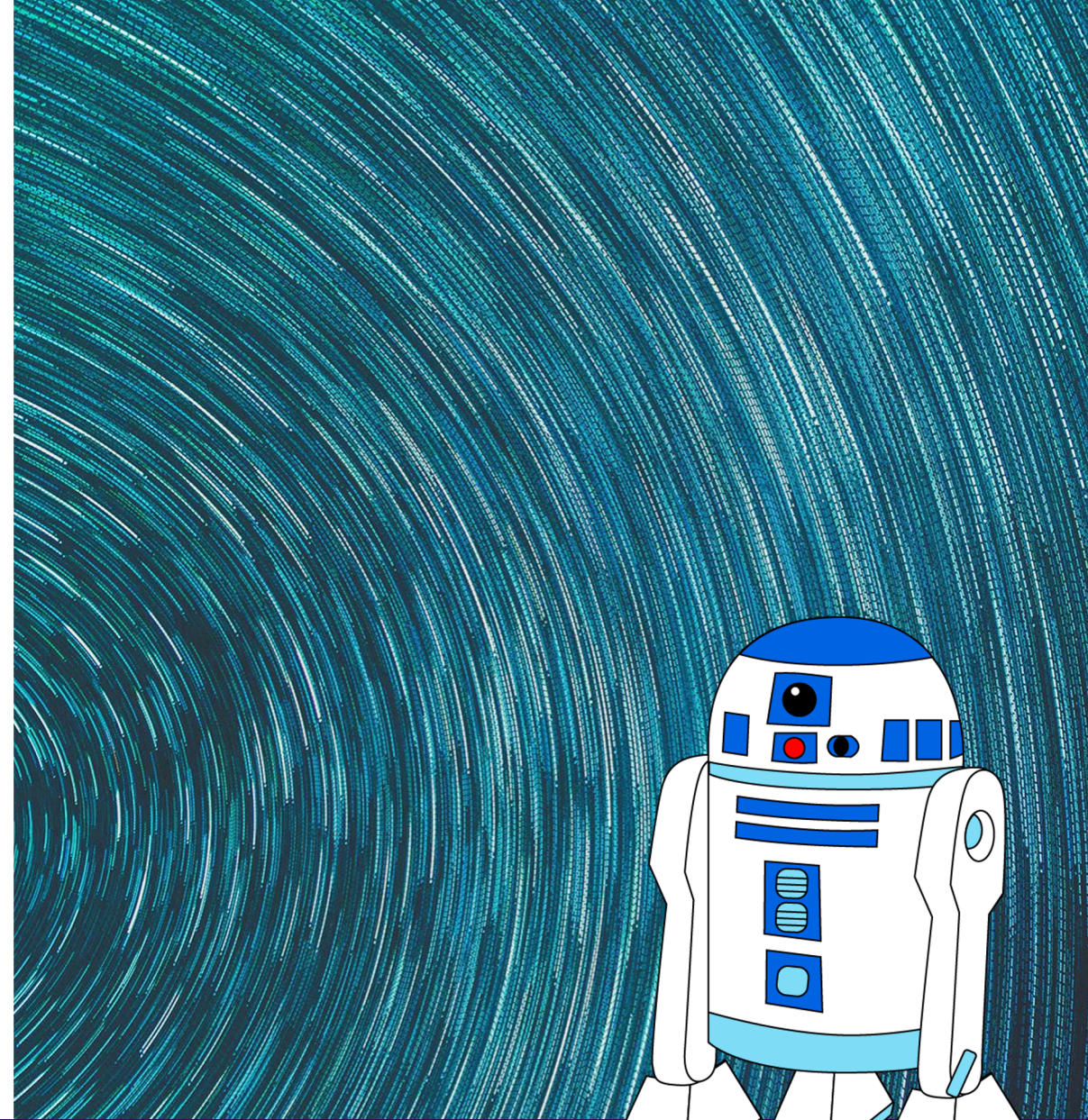
Your task is to decide at what point (if at all) the text you see begins to be written by a computer. Please click the blue button **as soon as** you are confident. Don't be surprised if some excerpts are entirely human-written. You will be awarded 5 points if you guess correctly, with decreasing amount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

Human-Written Prompt:

The world isn't always compliant with our wants.

CIS 421/521:
ARTIFICIAL INTELLIGENCE

Rational Agents



Four views of Artificial Intelligence

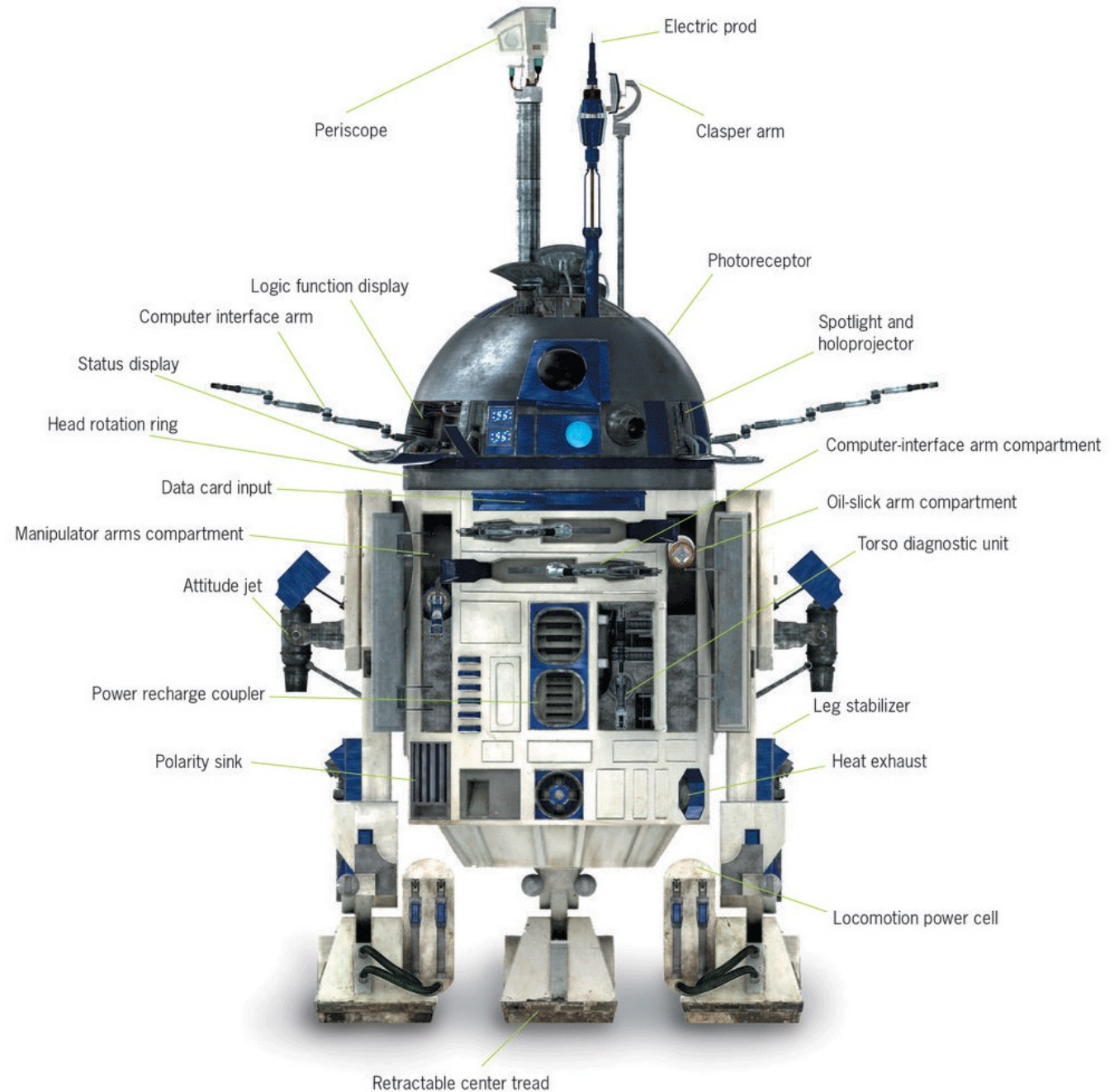
Thinking humanly	Thinking rationally
Acting humanly	Acting rationally

This course is about effective programming techniques for designing rational agents

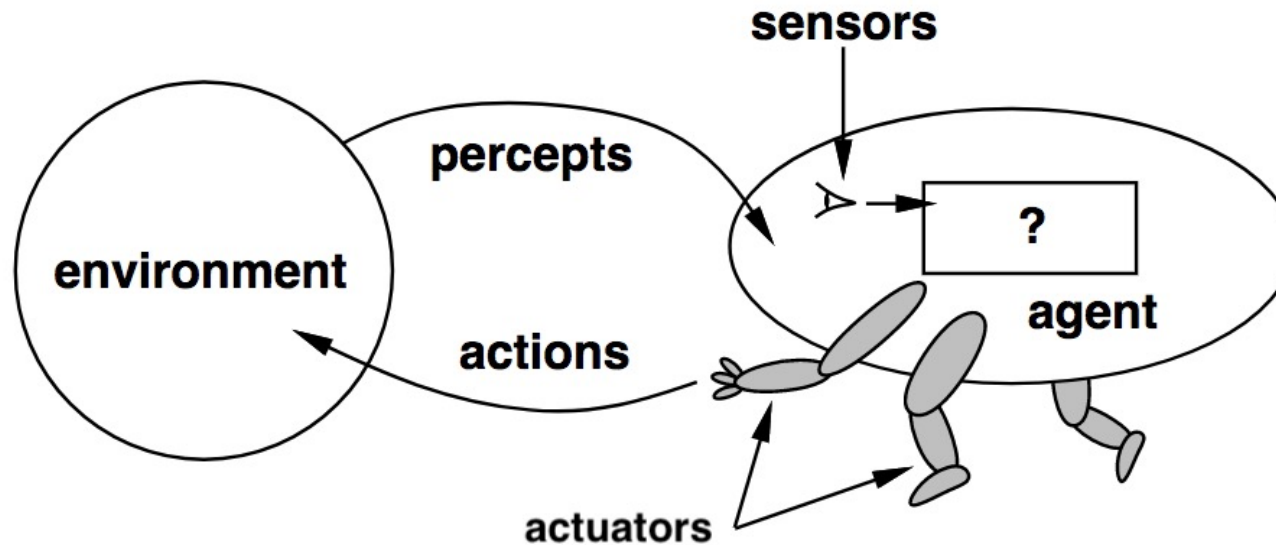
Agents

An **agent** is anything that **perceives** its environment through **sensors** and can **act** on its environment through **actuators**

A **percept** is the agent's perceptual inputs at any given instance.



Agents and environments



An agent is specified by an *agent function* $f:P \rightarrow a$ that maps a sequence of percept vectors P to an action a from a set A :

$$P=[p_0, p_1, \dots, p_t]$$

$$A=\{a_0, a_1, \dots, a_k\}$$

abstract
mathematical
description

Agent function & program

The *agent program* runs on the physical *architecture* to produce f

- $agent = architecture + program$

“Easy” solution: a giant table that maps every possible sequence P to an action a

- One problem: exponential in length of P

Agents

An *agent* is anything that can be viewed as

- *perceiving* its *environment* through *sensors* and
- *acting* upon that environment through *actuators*

Human agent:

- Sensors: eyes, ears, ...
- Actuators: hands, legs, mouth, ...

Robotic agent:

- Sensors: cameras and infrared range finders
- Actuators: various motors

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



Rational Behavior

Rational Agent

Let's try to define "rational agent".

A **rational agent** is an agent that perceives its environment and and behaves rationally

Rational behavior: doing the right thing

Obviously doing the right thing is better than doing the wrong thing, but *what does it mean to do the right thing?*

In Philosophy

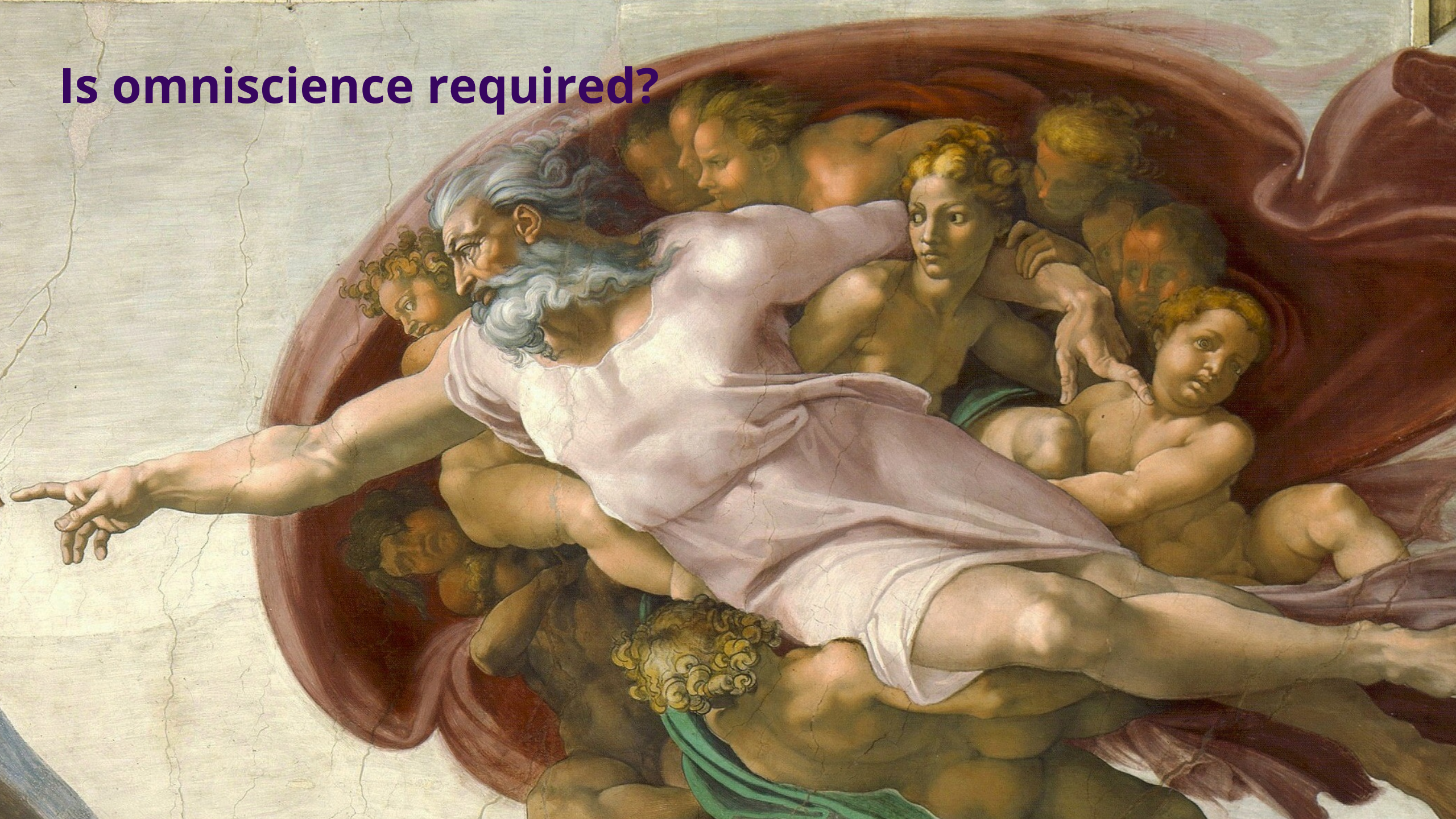
Moral philosophy has developed different notions of "the right thing".

AI is usually concerned with **Consequentialism**.

We evaluate an agent's behavior by its consequences.



Is omniscience required?



In Economics

A BEHAVIORAL MODEL OF RATIONAL CHOICE

Herbert A. Simon*

Summary: A model is proposed for the description of rational choice by organisms of limited computational ability.

Rational choice theory is a framework for understanding social and economic behavior.

The basic premise is that aggregate social behavior results from the behavior of individual actors, each of whom is making their individual decisions.

It assumes that individuals have preferences and choose the alternative that they prefer.



Herbert Simon

Performance measure

How do we know if an agent is acting rationally?

- Informally, we expect that it will do the right thing in all circumstances.

How do we know if it's doing the right thing?

We define a **performance measure**:

- An objective criterion for success of an agent's behavior
- given the evidence provided by the percept sequence.

Performance measure examples

Performance measures for a vacuum-cleaner agent might include things like:

- +1 point for each clean square at time T
- +1 for cleaning a square, -1 for each move
- -1000 for more than k dirty squares



Rule of thumb for performance measures

It is better to design performance measures according to **what you want to be achieved** in the environment, **rather than how you think the agent should behave**.

For example, what might happen if we said

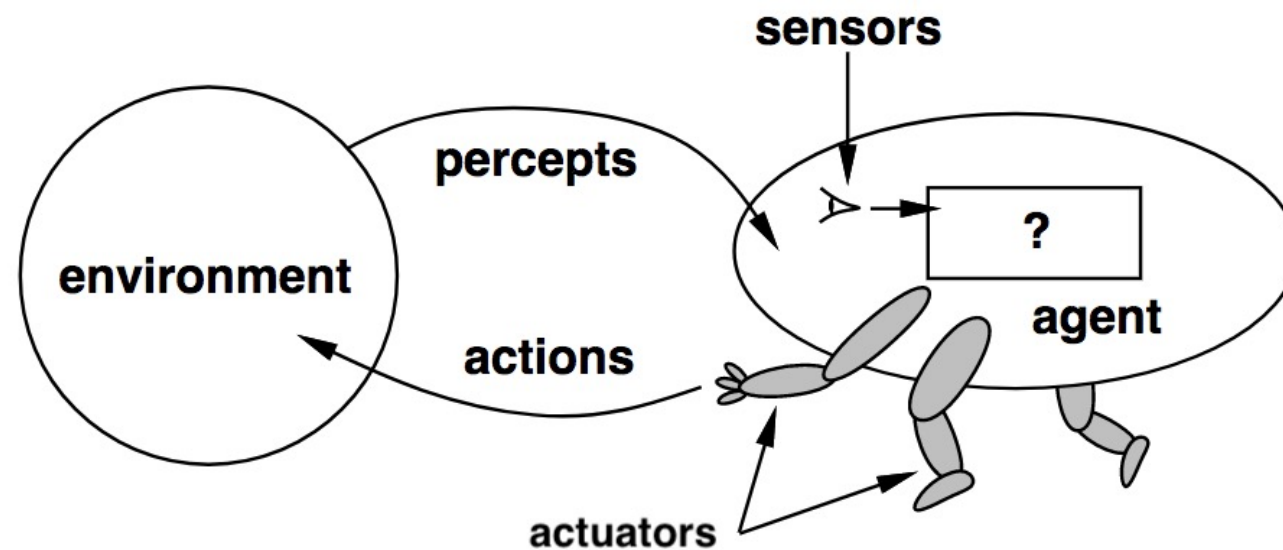
- +1 point for each time the robot cleans a square
- instead of
- +1 point for each clean square at time T



Rational agents

Rational Agent:

- For each possible percept sequence P
- a rational agent selects an action a
- to *maximize* its *performance measure*



Rationality is *not* omniscience

Ideal agent: maximizes *actual* performance, but needs to be *omniscient*.

- Usually impossible.....
 - But consider tic-tac-toe agent...
- Rationality \neq Guaranteed Success

Caveat: *computational limitations make complete rationality unachievable*

→ design best *program* for given machine resources



Expected value

Rational Agent (initial definition):

- For each possible percept sequence P ,
- a rational agent selects an action a
- to maximize its performance measure

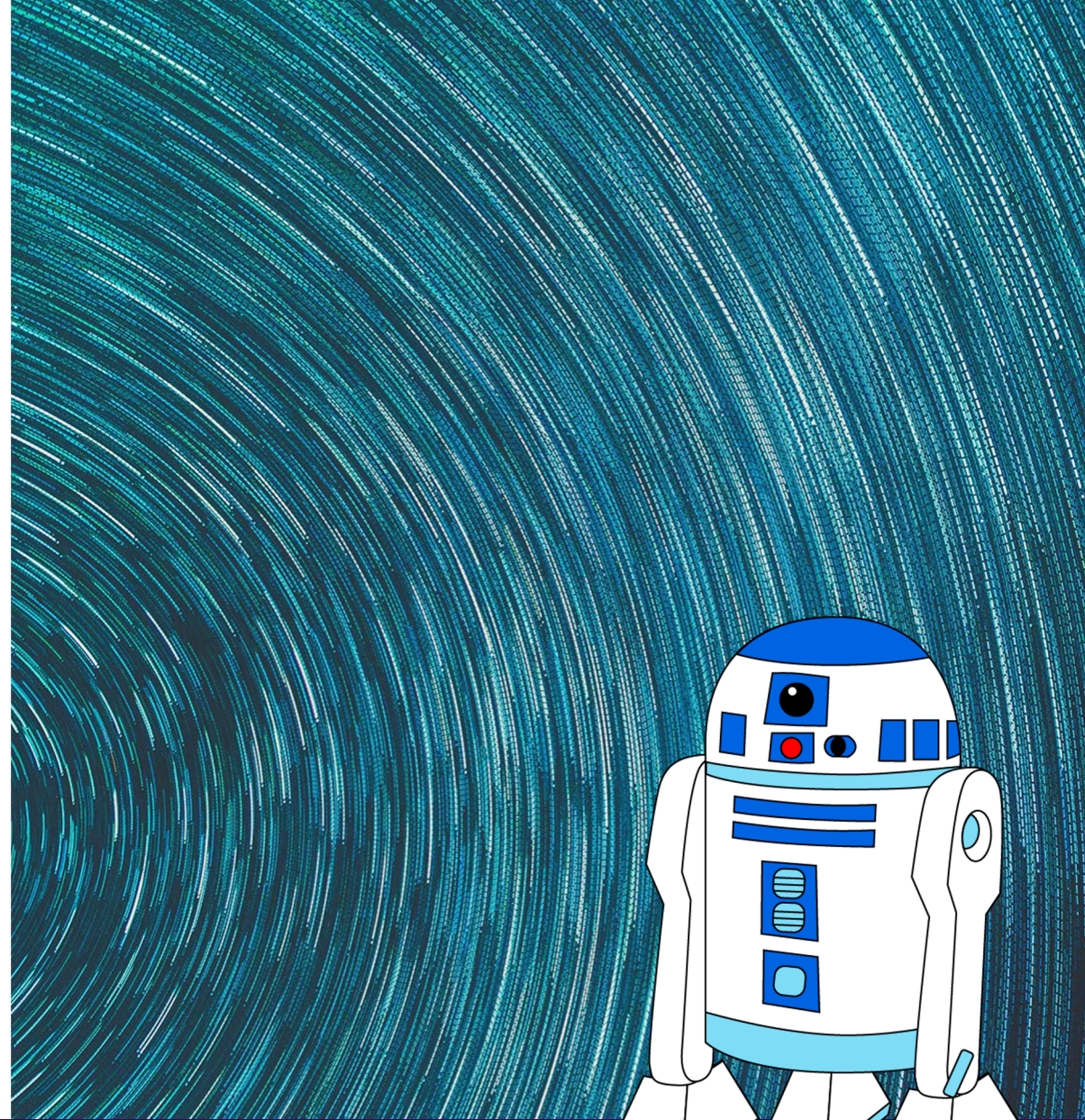
Rational Agent (revised definition):

- For each possible percept sequence P ,
- a rational agent selects an action a
- that maximizes the **expected value** of its performance measure

***It doesn't have to know
what the actual outcome
will be.***

CIS 421/521:
ARTIFICIAL INTELLIGENCE

Task Environments



Task environments

To design a rational agent we need to specify a *task environment*

- a problem specification for which the agent is a solution

PEAS: to specify a task environment

- *P*erformance measure
- *E*nvironment
- *A*ctuators
- *S*ensors



PEAS: Specifying an automated taxi driver

*P*erformance measure:

- ?

*E*nvironment:

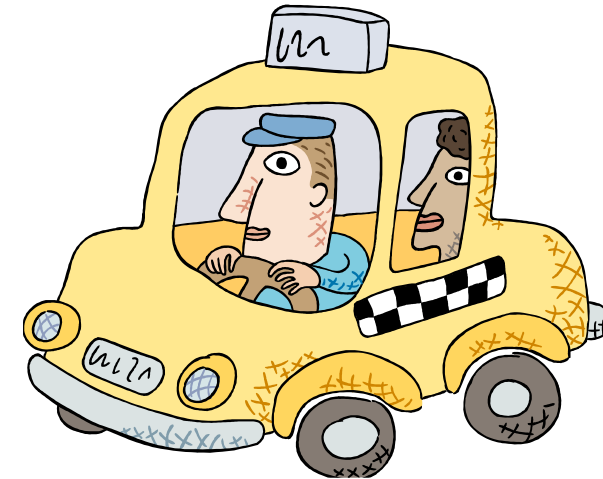
- ?

*A*ctuators:

- ?

*S*ensors:

- ?



PEAS: Specifying an automated taxi driver

Performance measure:

- safe, fast, legal, comfortable, maximize profits

Environment:

- roads, other traffic, pedestrians, customers

Actuators:

- steering, accelerator, brake, signal, horn

Sensors:

- cameras, LiDAR, speedometer, GPS





<https://www.today.com/video/amazon-adebuts-new-package-delivery-drone-61414981780>

PEAS: Amazon Prime Air

*P*erformance measure:

- ?

*E*nvironment:

- ?

*A*ctuators:

- ?

*S*ensors:

- ?



The rational agent designer's goal

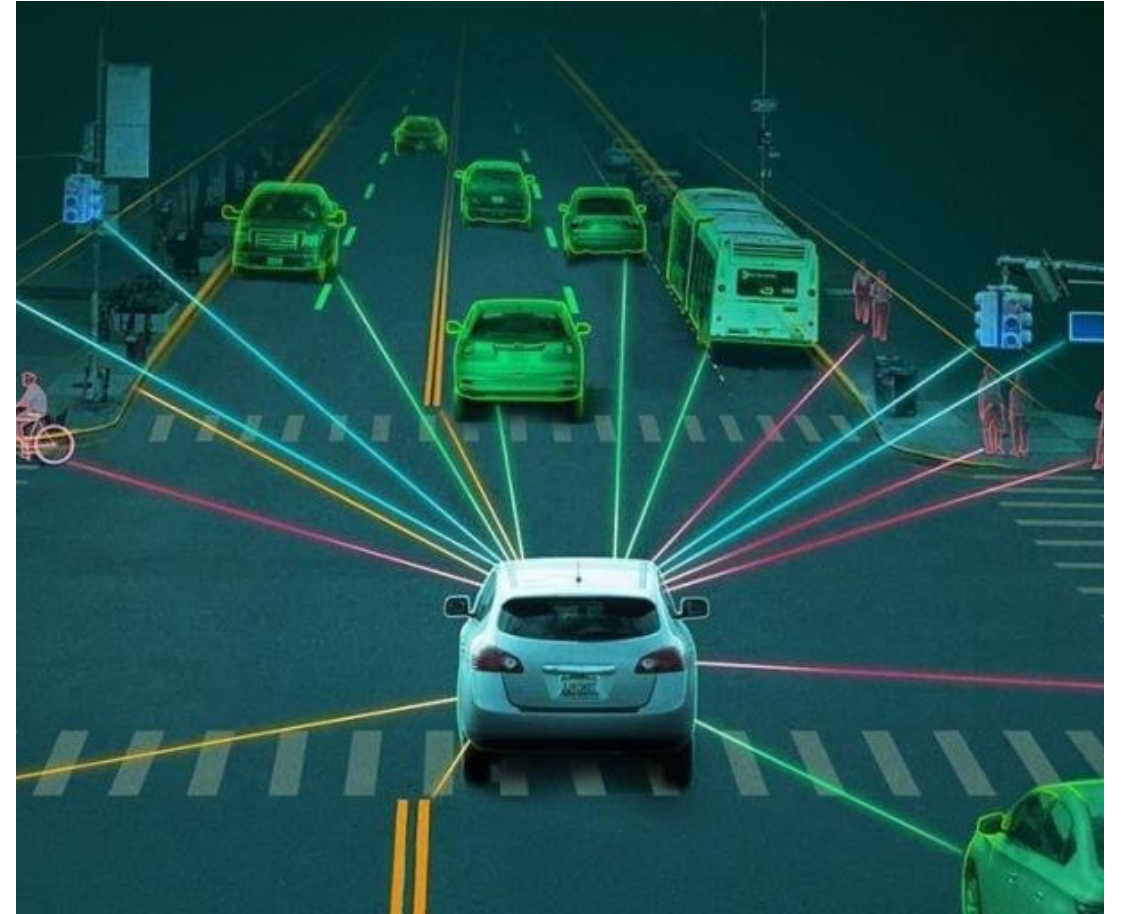
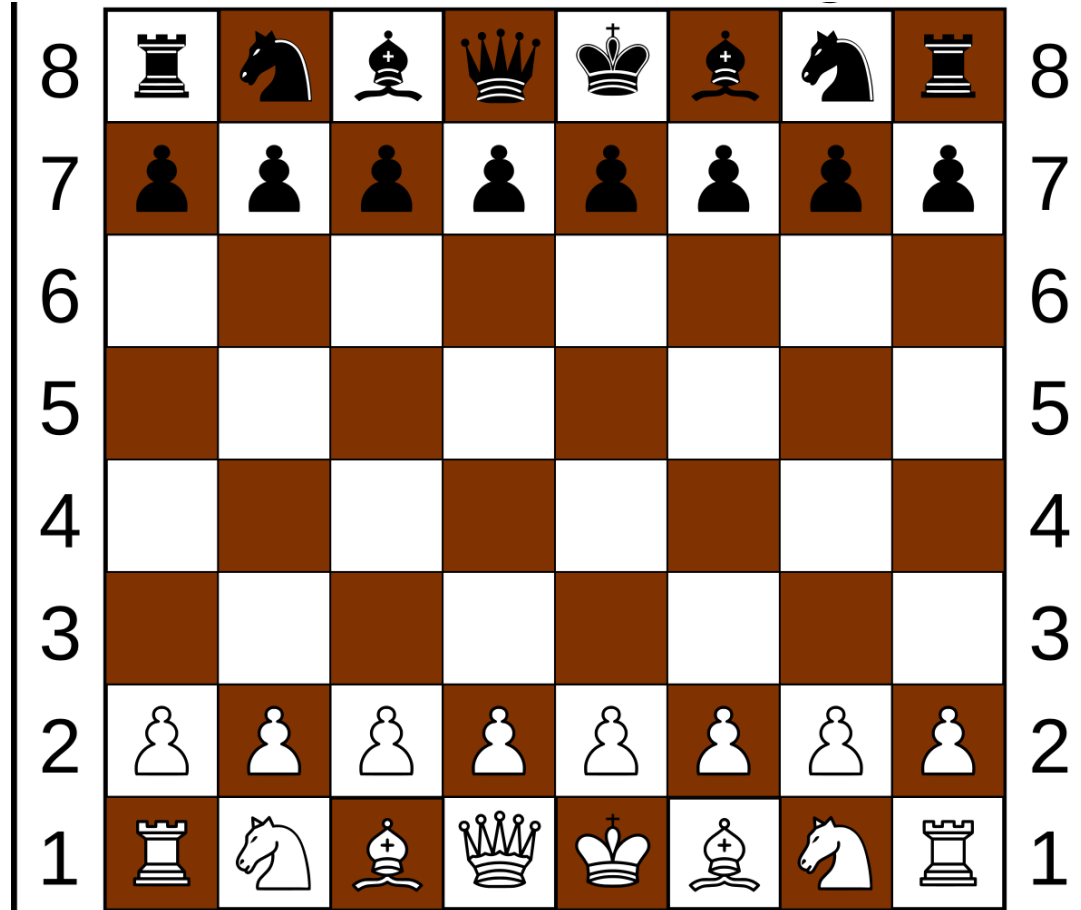
Goal of AI practitioner who designs rational agents:
given a *PEAS* task environment,

1. Construct *agent function* f that maximizes the expected value of the performance measure,
2. Design an *agent program* that implements f on a particular architecture

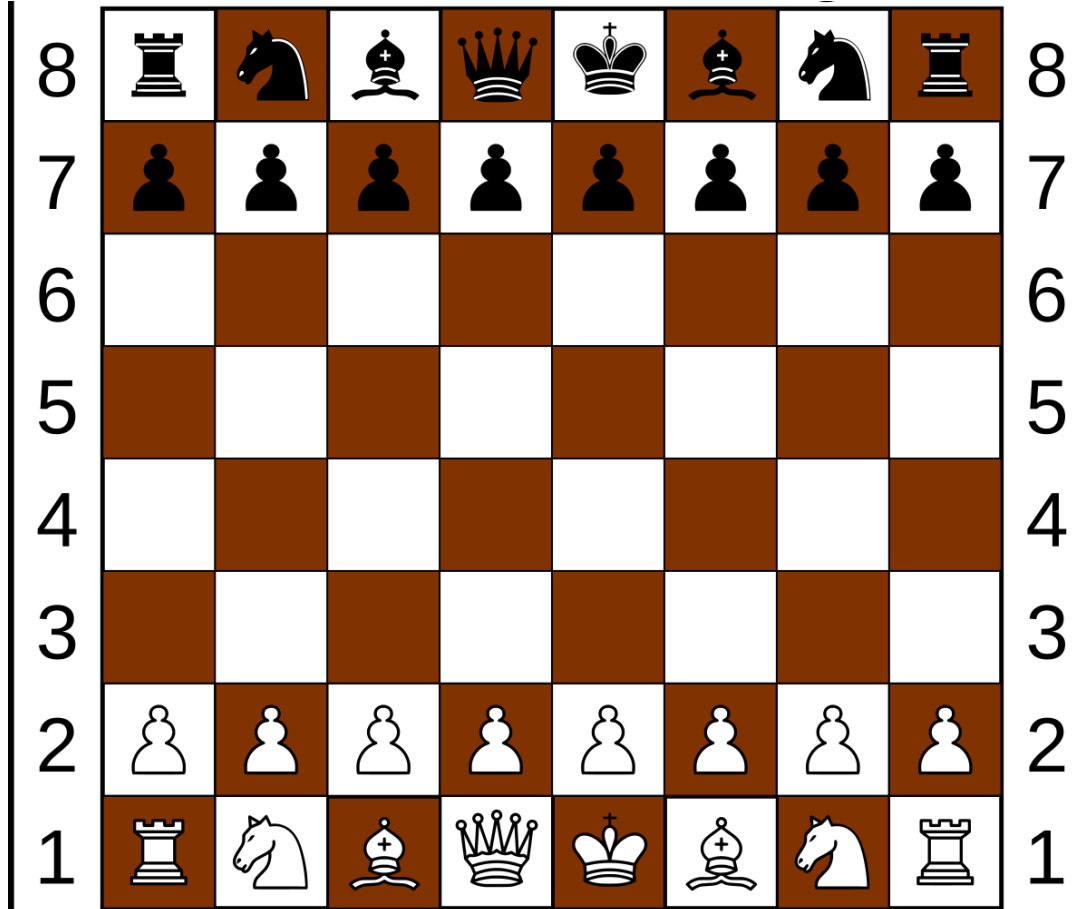
abstract
mathematical
description

concrete
implementation

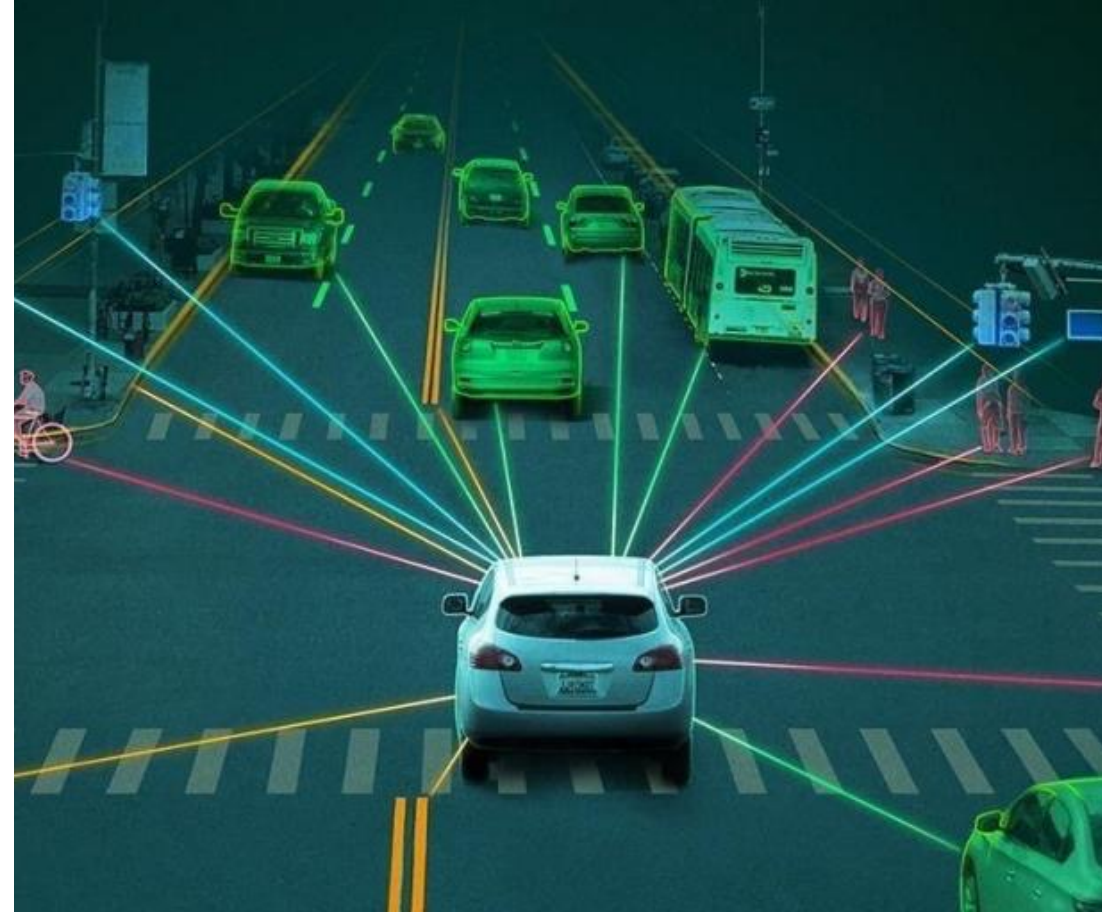
Fully Observable v. Partially Observable



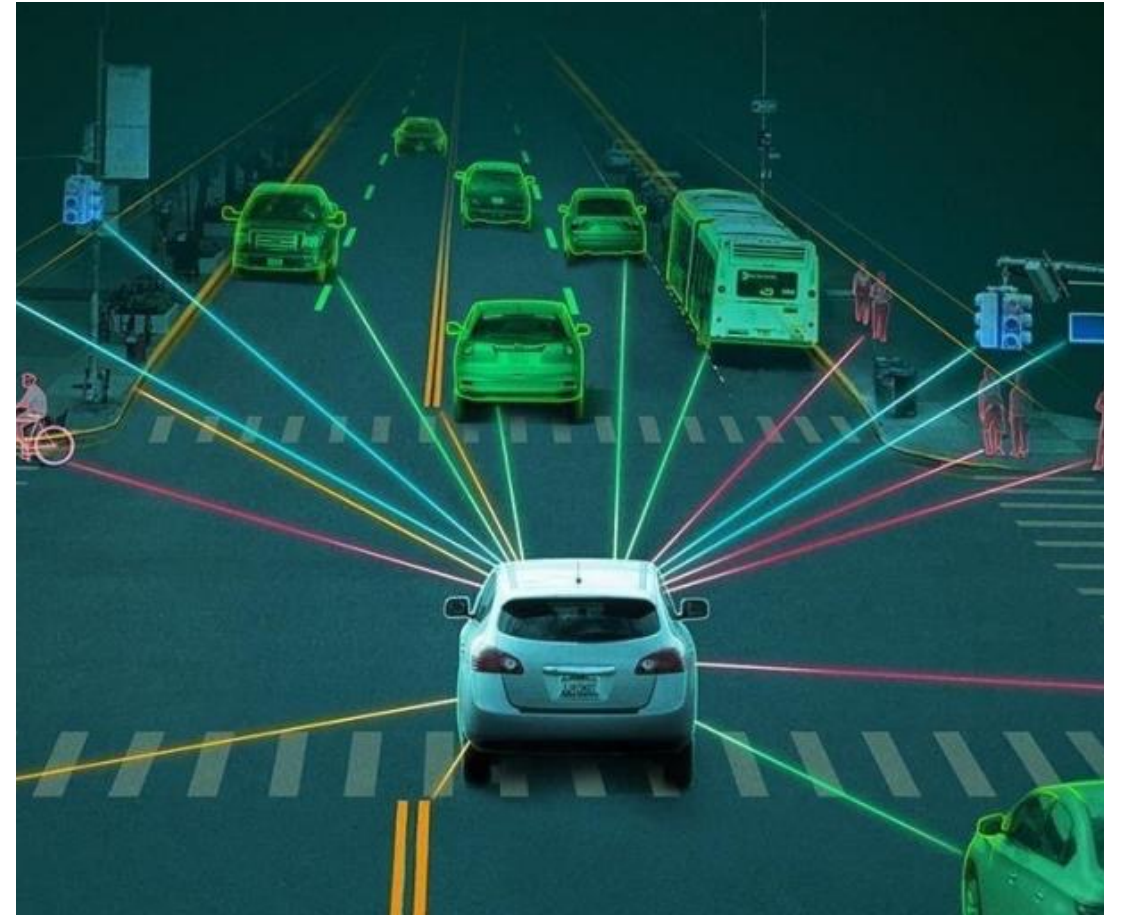
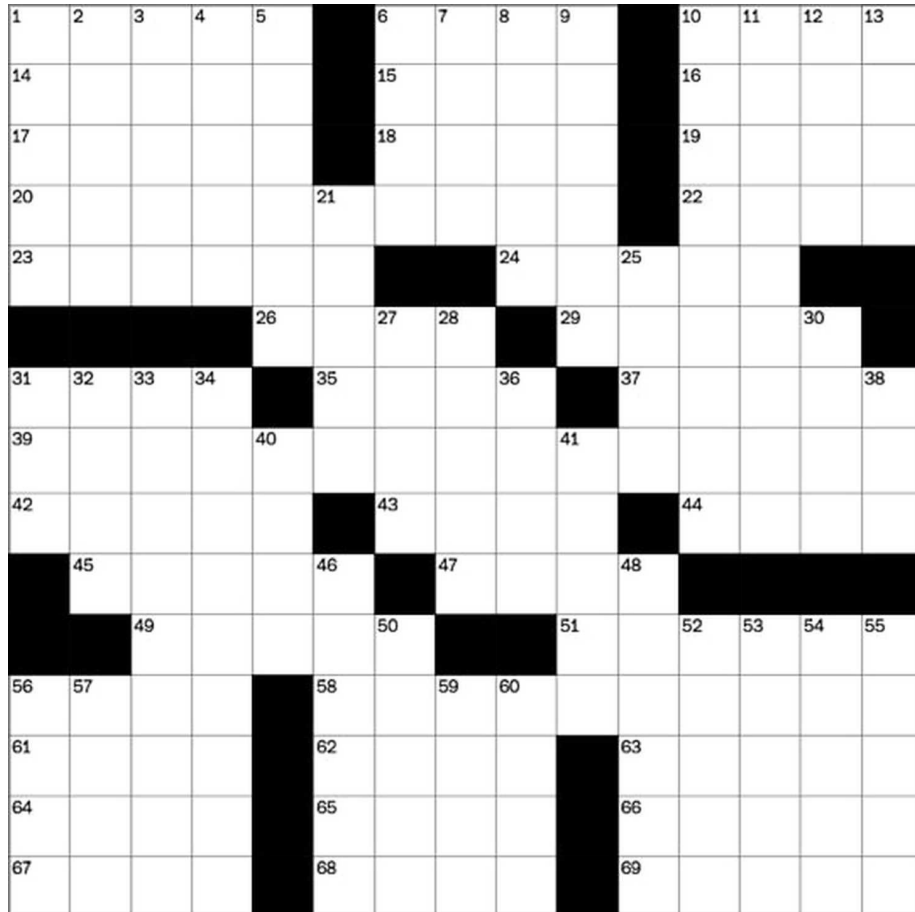
Deterministic v. Nondeterministic v. Stochastic



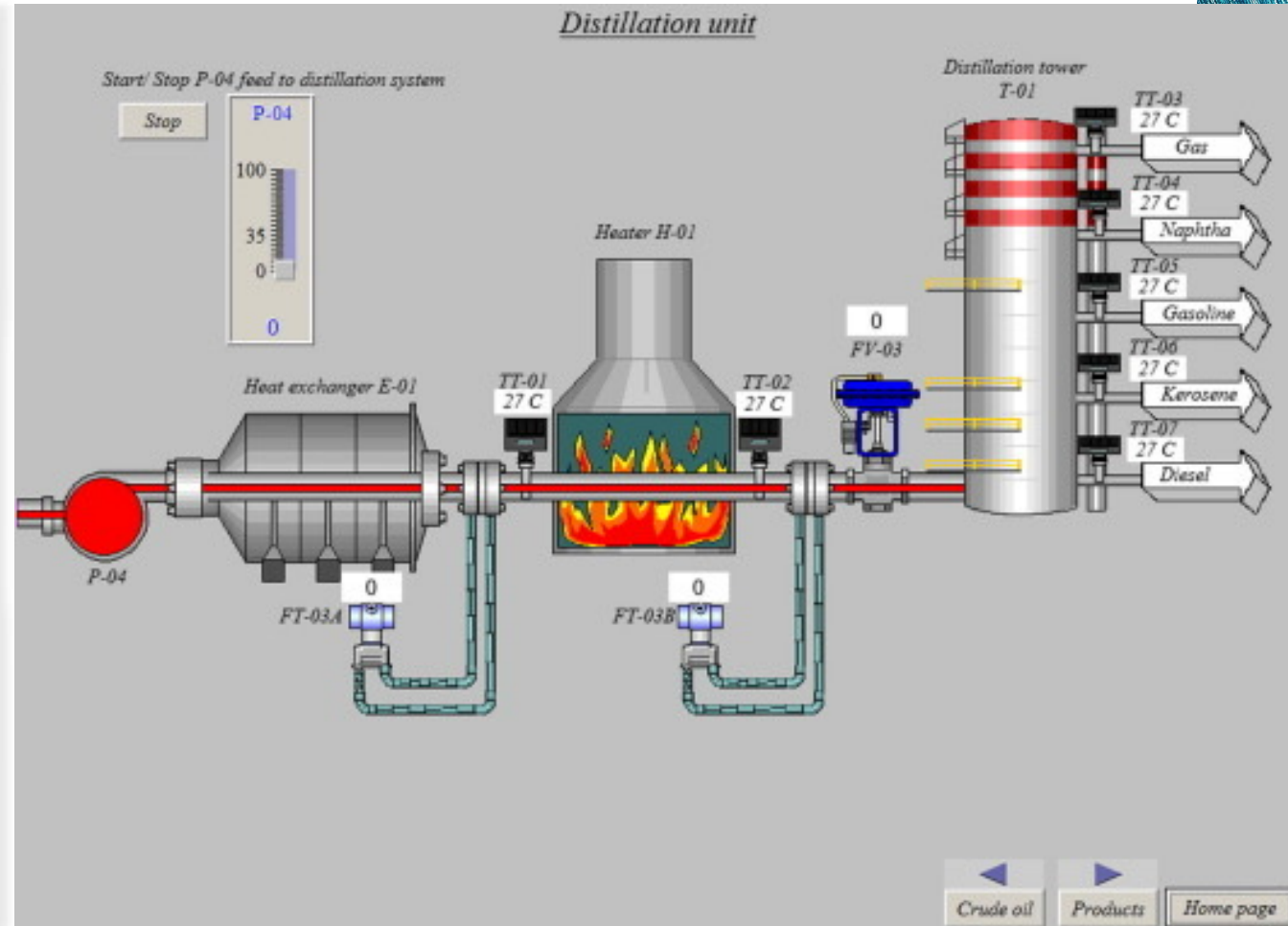
Episodic v. Sequential



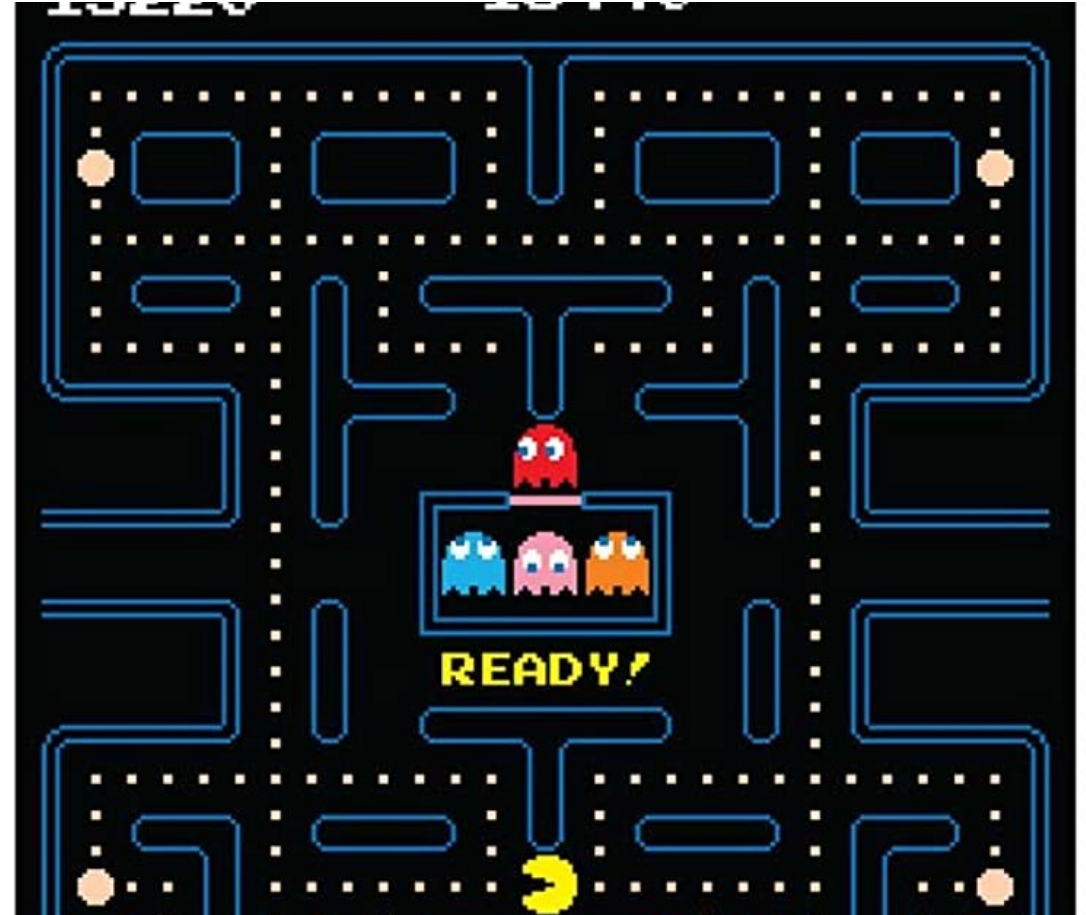
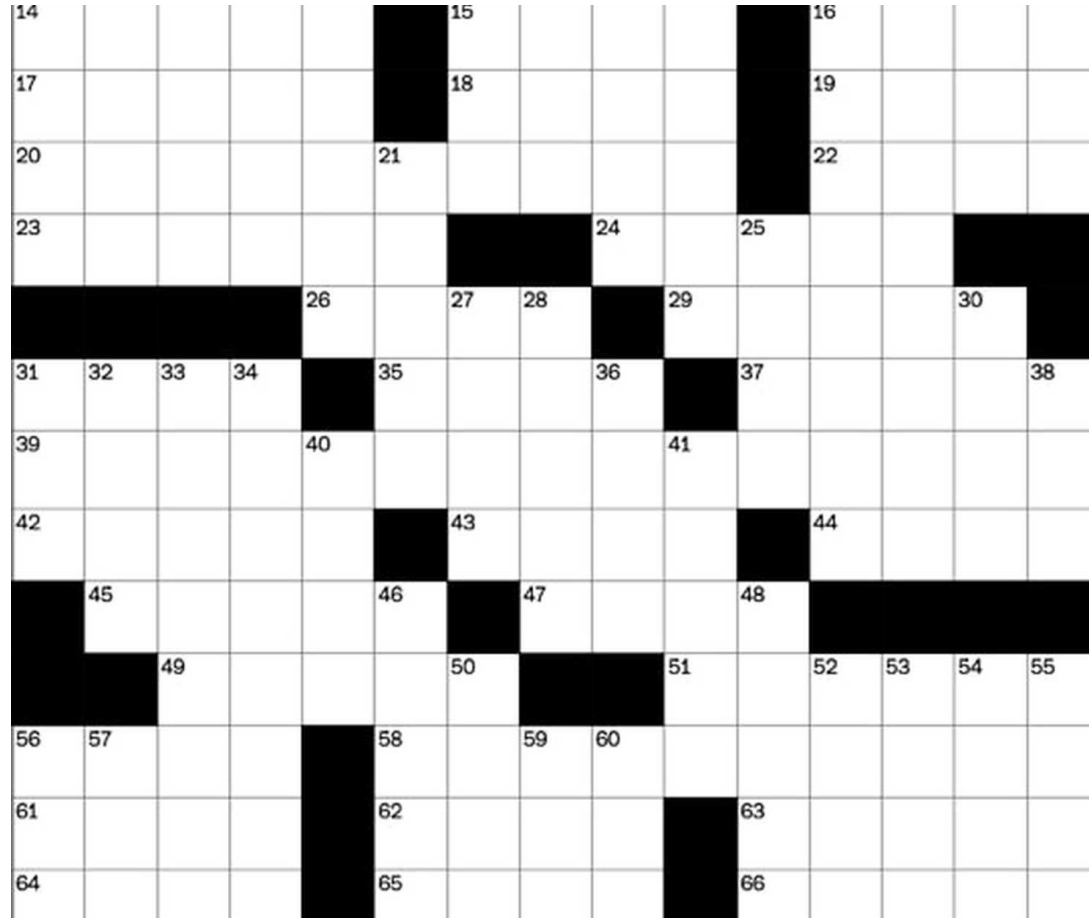
Static v. Dynamic



Discrete v. Continuous



Single Agent v. Multi Agent



When should something be considered an agent?

If we're talking about a self driving taxi, when should we consider something part of the environment versus another agent?

For instance, a telephone pole is part of the environment, but a car might be another agent.

When something behavior can best be described as having its own performance measure, then we should consider it to be an agent.

The Hardest Environment

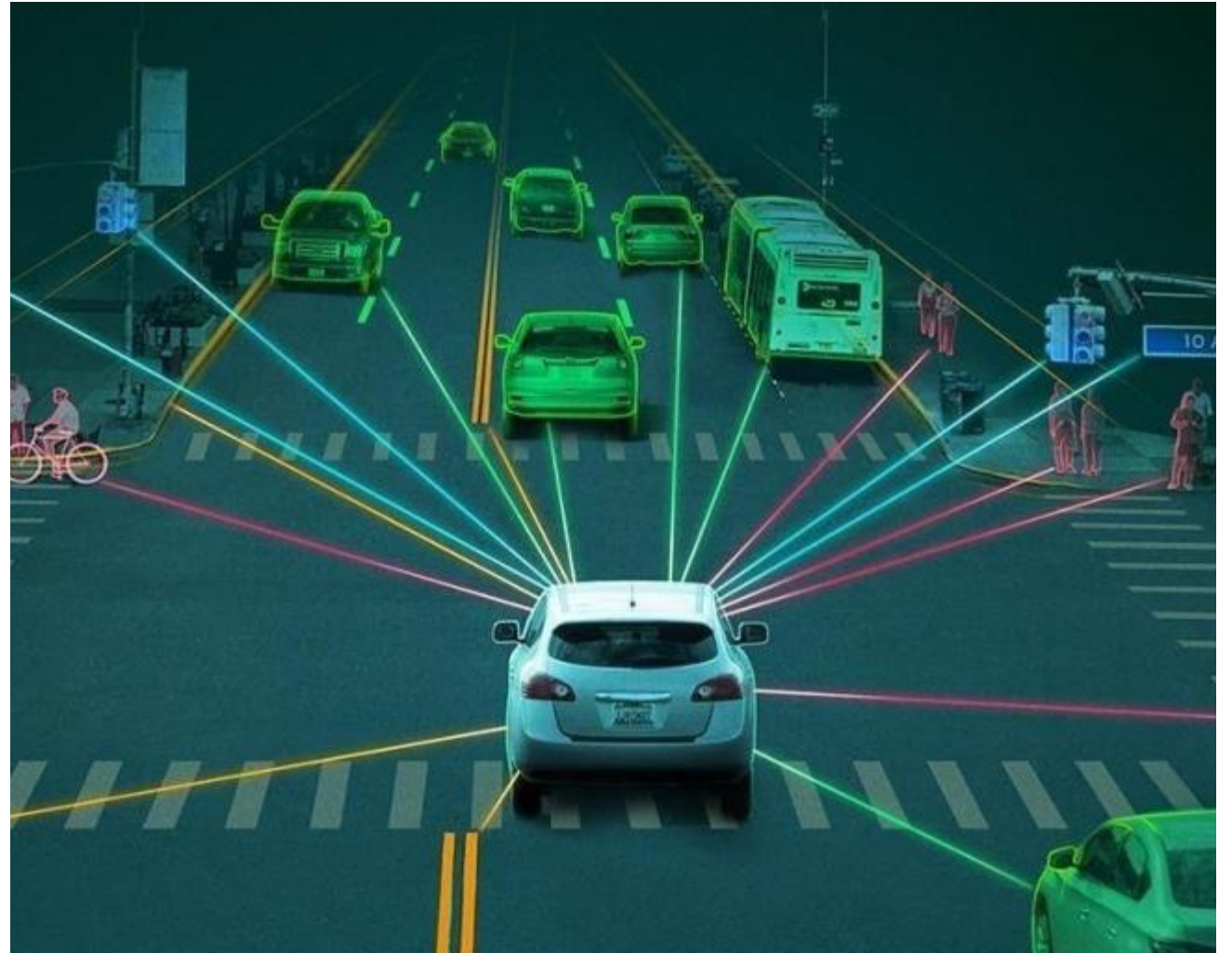
The hardest case is

Continuous

Partially Observable

Stochastic

Multiagent



Environment Restrictions for Now

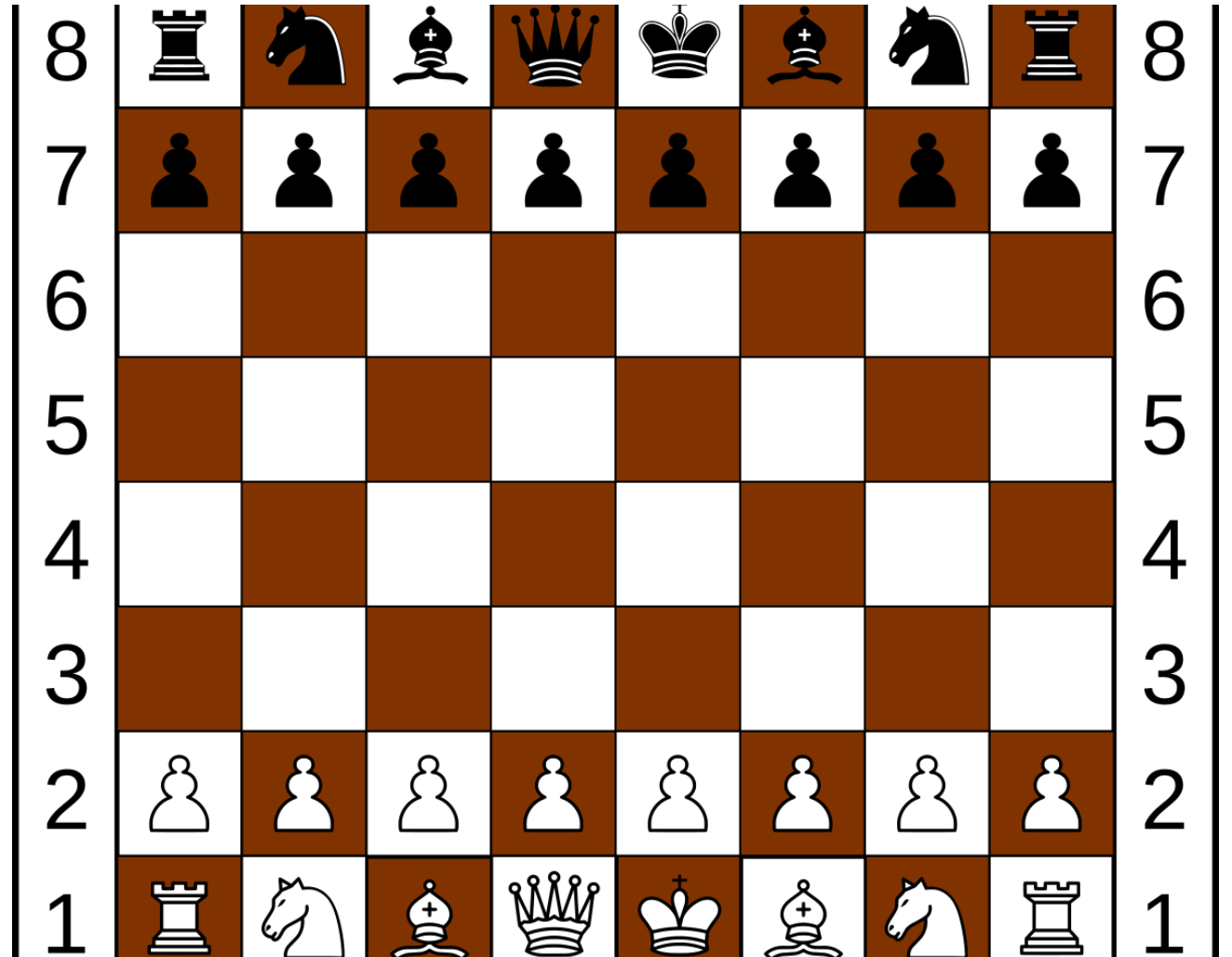
We will assume environment is

Static

Fully Observable

Deterministic

Discrete

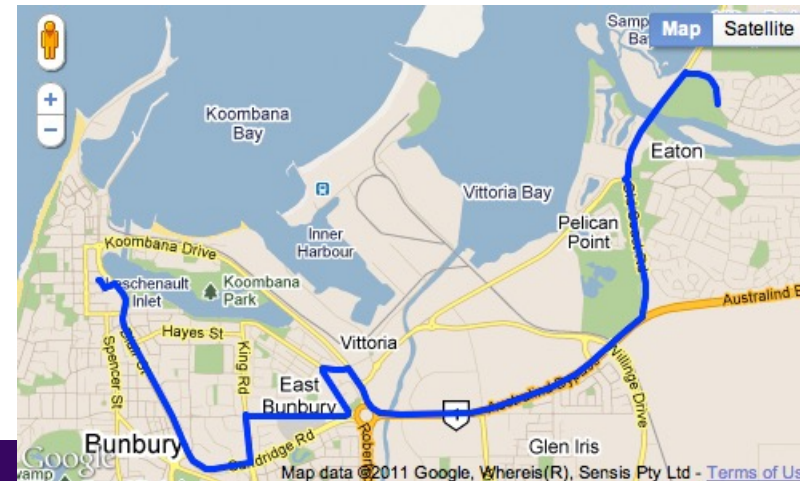


Reflex agents v. Problem solving agents

A simple reflex agent is one that selects an action based on the current percept, and ignores the rest of the percept history.



A problem-solving agent must plan ahead. It will consider a sequence of actions that form a path to a goal state. The computational process that it undertakes is called search.



Problem Solving Agents & Problem Formulation

AIMA 3.1-3.2

Example search problem: 8-puzzle



Formulate *goal*

- Pieces to end up in order as shown...

7	2	4
5		6
8	3	1

Start State

	1	2
3	4	5
6	7	8

Goal State

Formulate *search problem*

- **States:** configurations of the puzzle (9! configurations)
- **Actions:** Move one of the movable pieces (≤ 4 possible)
- **Performance measure:** minimize total moves

Find *solution*

- Sequence of pieces moved: 3,1,6,3,1,...