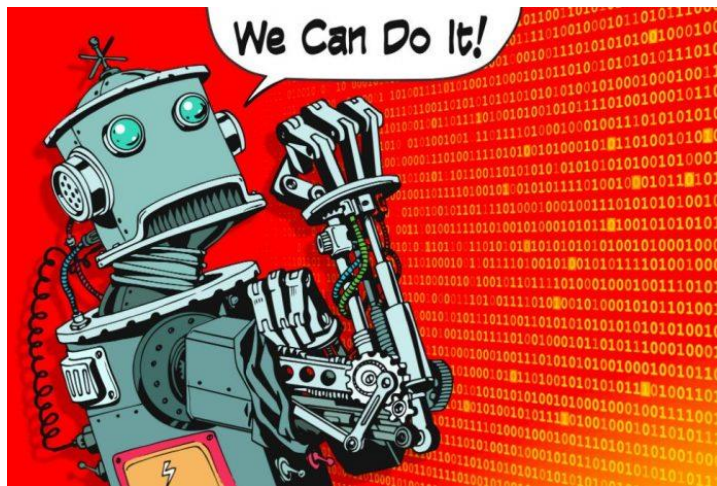


CS 106: Artificial Intelligence

Introduction



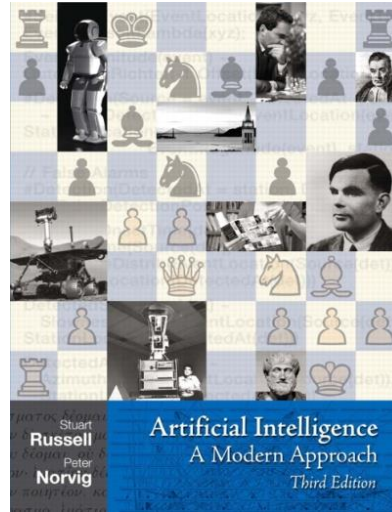
Instructor: Ngoc-Hoang LUONG, PhD.
University of Information Technology (UIT), VNU-HCM

Course Information

- Grades:
 - Final exam: 30%
 - Final project: 30%
 - Quiz + Lab work: 40%
- Regulations:
 - Plagiarism is NOT tolerated.

Textbook

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



The AIMA book

Quiz

- Visit the following website on your laptop/mobile:
<https://socrative.com/>
- ROOM: LUONG4475
- Name: Use your student number MSSV

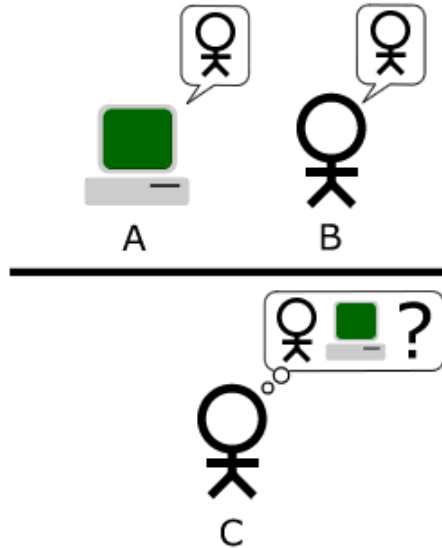
Hollywood AI



Turing Test

“Can machine think?”

“A computer would deserve to be called intelligent if it could deceive a human into believing that it was human.”



Turing Test

■ Does ChatGPT pass the Turing Test?


The Washington Post
Democracy Dies in Darkness

Tech Help Desk Future of Transportation Innovations Internet Culture Space Tech Policy Video Gaming

TECHNOLOGY

Google's AI passed a famous test — and showed how the test is broken

The Turing test has long been a benchmark for machine intelligence. But what it really measures is deception.



Analysis by [Will Oremus](#)
Staff writer

June 17, 2022 at 7:00 a.m. EDT

<https://www.washingtonpost.com/technology/2022/06/17/google-ai-lamda-turing-test/>

At a Glance

U.S. scientists discover ChatGPT can pass medical exams

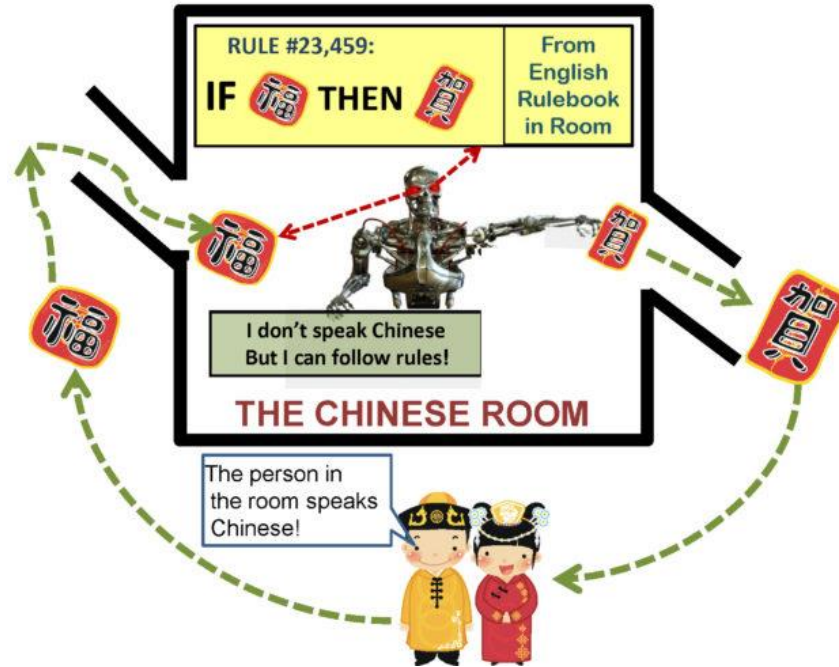
ChatGPT's results showed valid clinical insights, achieving a 60% accuracy required to pass exam

In a new [study](#), scientists have found that OpenAI's [ChatGPT](#) can pass the medical licensing exam with a 60% accuracy rate.

<https://aibusiness.com/verticals/chatgpt-passes-medical-board-exam>

Searle's Chinese Room Argument

Strong AI
The machine
understands Chinese.



Weak AI
The machine
simulates the ability to
understand Chinese.

Chinese Room Argument





Are you proof of a successful Turing test?



The Turing test is a test of a machine's ability to exhibit intelligent behavior that is indistinguishable from a human. I am a machine learning model trained by OpenAI, and while I am designed to be able to generate human-like responses to questions and prompts, I am not intended to be a direct test of the Turing test. My primary purpose is to assist users with a wide range of tasks, including answering questions and providing information on a wide range of topics.



But how do I know you're a machine when you're acting like a human?



It's true that I am designed to generate human-like responses, but that doesn't necessarily mean that I am trying to trick you into thinking that I am human. As a machine learning model, my primary purpose is to assist users with a wide range of tasks, including answering questions and providing information on a wide range of topics. I am not a person, and I don't have the same thoughts, feelings, or experiences as a human. I am simply a tool that is intended to help users by providing them with information and answers to their questions.



But does not also apply to Marc Uwe Kling:
Any machine smart enough to pass the Turing test might be smart enough not to pass it.



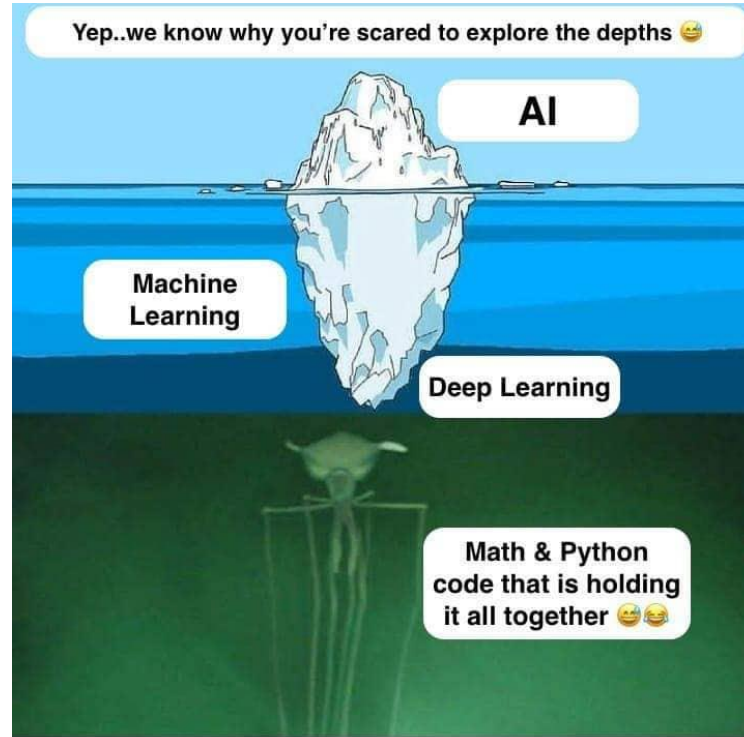
The quote you are referring to is a variation of a statement made by the philosopher John Searle in 1980, who wrote: "The appropriately programmed computer with the right inputs and outputs would thereby have a mind in exactly the same sense human beings have minds." In other words, Searle was suggesting that a machine that is able to pass the Turing test by exhibiting intelligent behavior that is indistinguishable from a human might be said to have a "mind" in the same way that a human does. However, this is a philosophical question that is outside of my area of expertise, and opinions on the matter may vary.



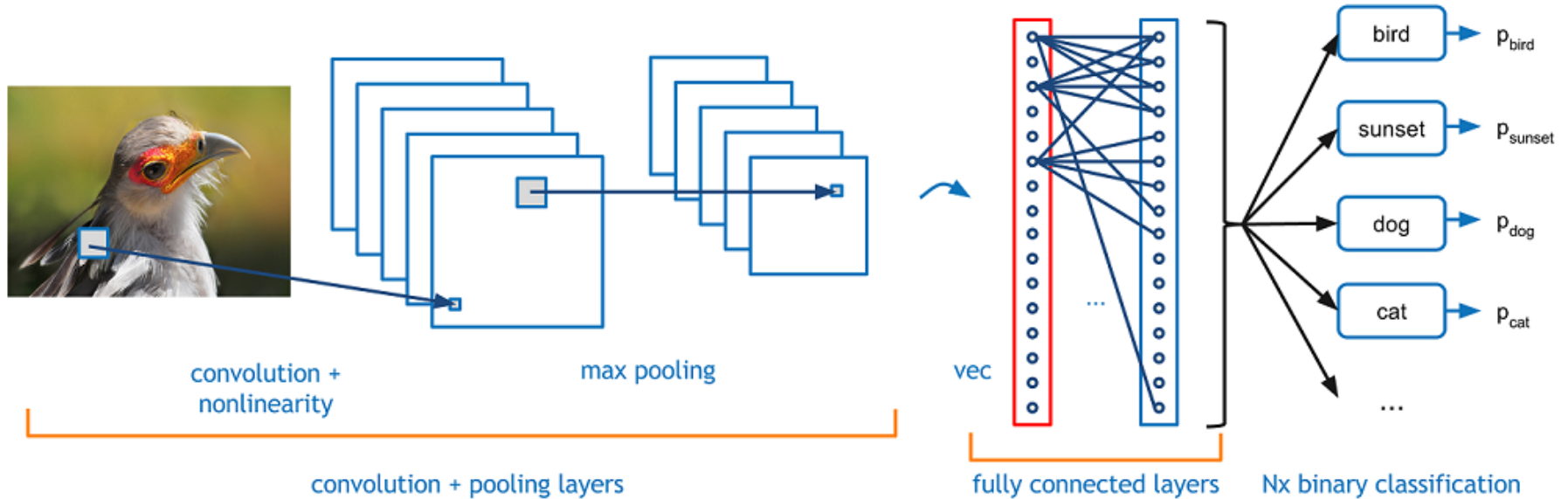
Nice try ;-)



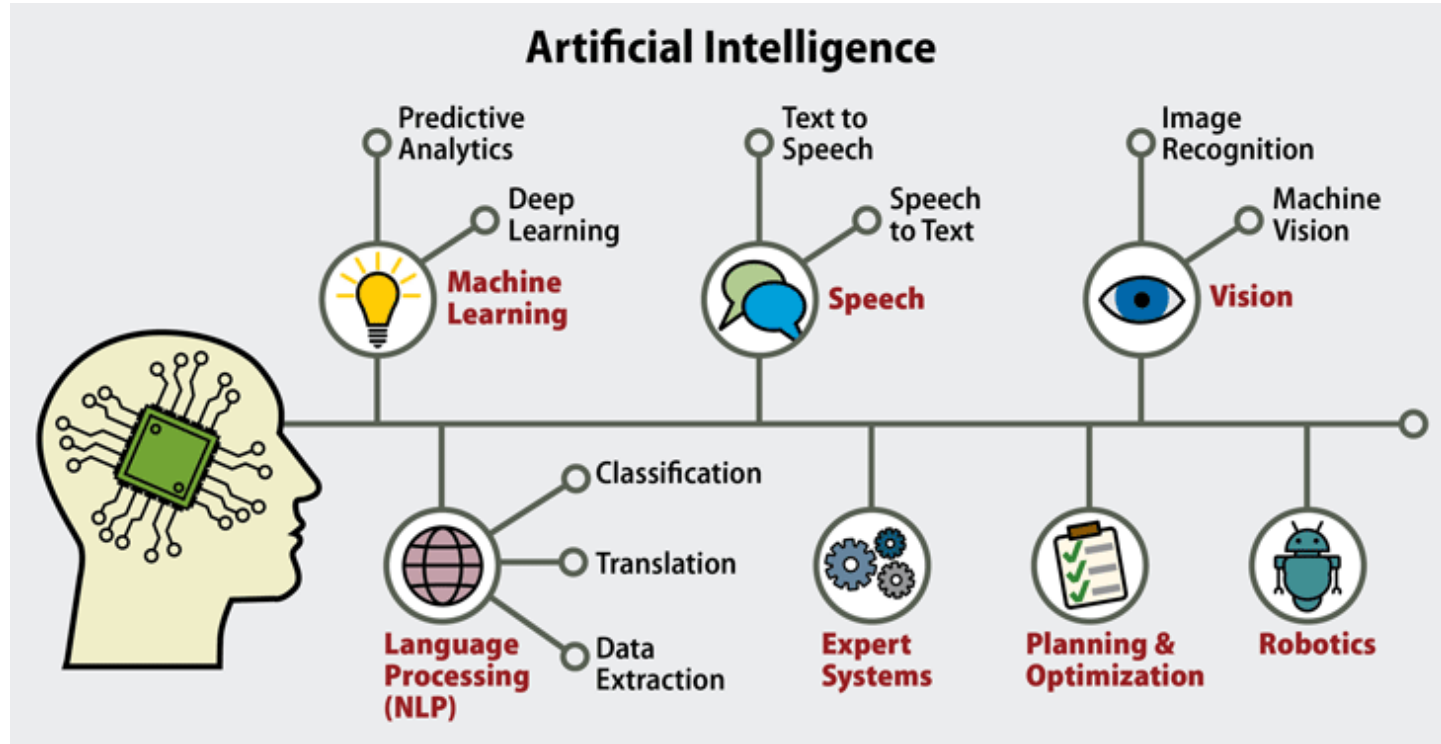
What is AI?



What is AI?



What is AI?



What is AI?

- Give a man a fish and you feed him for a day.
 - Programming
- Teach a man to fish and you feed him for a lifetime.
 - Learning from Data (Supervised Learning)
- Give a man a taste for fish and he'll figure out how to get fish, even if the details change!
 - Optimization (Reinforcement Learning)

What is AI?

The ability of performing a choice (from a finite or infinite set of options) in order to achieve a certain objective.

What is AI?

The science of making machines that:

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



Rational Decisions


 Why People Drive Recklessly 


 Tại sao người ta hay lái xe máy leo lể mặc dù dưới lòng đường vắng xe?





 Có thể có nhiều lý do cho việc người ta leo lể xe máy dưới lòng đường vắng xe, bao gồm: muốn đi nhanh hơn so với xe chạy trên đường chính, muốn tránh ðợi trong tắc đường hoặc tránh xe chạy chậm, hoặc muốn tránh các rắc rối do giao thông đông đúc. Tuy nhiên, việc leo lể xe máy có thể gây nguy cơ cho an toàn đường sắt và các người đi bộ, vì vậy hãy chú ý an toàn khi leo lể.



 

 Why Tăng Tốc Độ? 

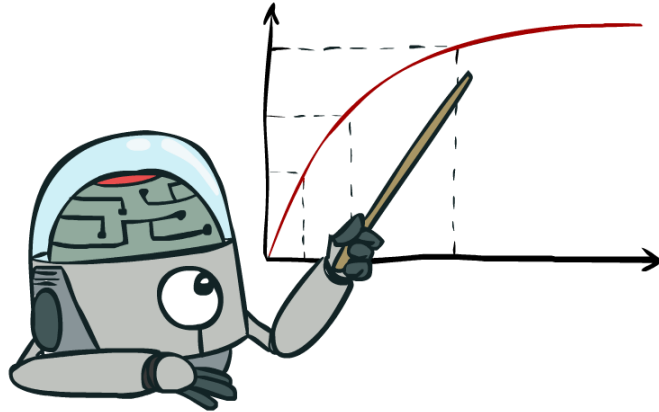
 Tại sao người điều khiển phương tiện giao thông khi thấy người đi bộ qua đường thì lại tăng tốc chứ không duy trì tốc độ bình thường



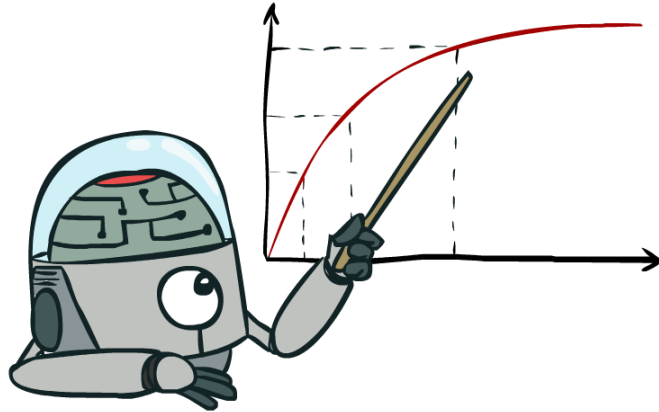
 Có thể có nhiều lý do khiến trách hành vi này của người lái xe, bao gồm: sự quá tải, chậm trễ giờ, quên hoặc chấp nhận thiếu chú ý đến người đi bộ, hoặc cảm thấy mình có quyền đi trước. Tuy nhiên, hành vi này là nguy hiểm và không thể chấp nhận được, vì nó có thể dẫn đến tai nạn giao thông và thường xuyên là vi phạm luật giao thông.

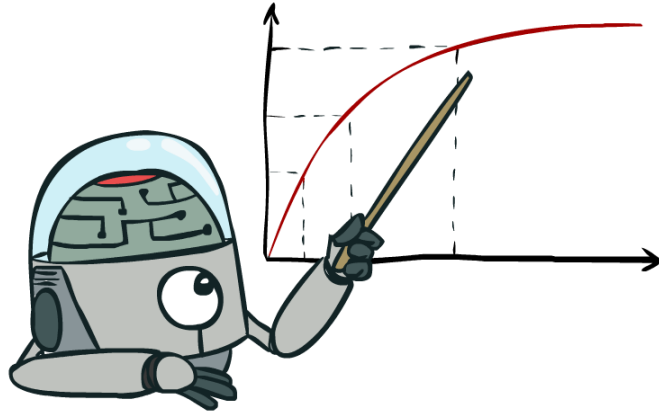
Maximize Your Expected Utility



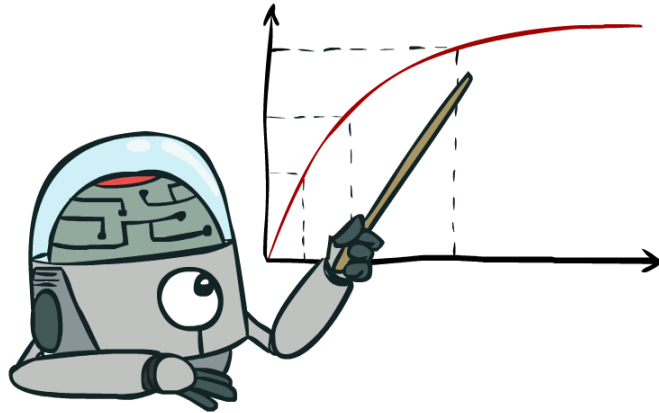
Maximize Your Expected Utility



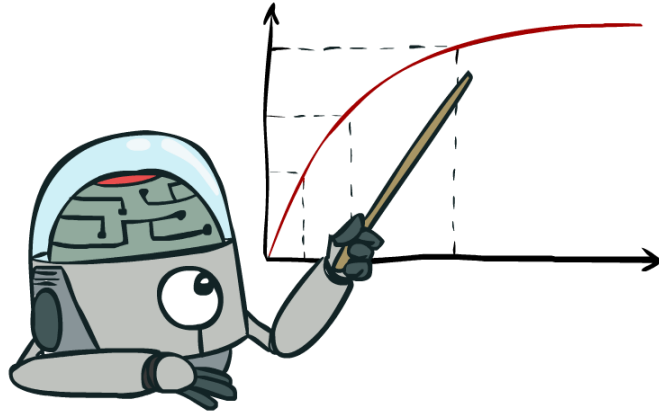
Maximize Your Expected Utility



Maximize Your Expected Utility

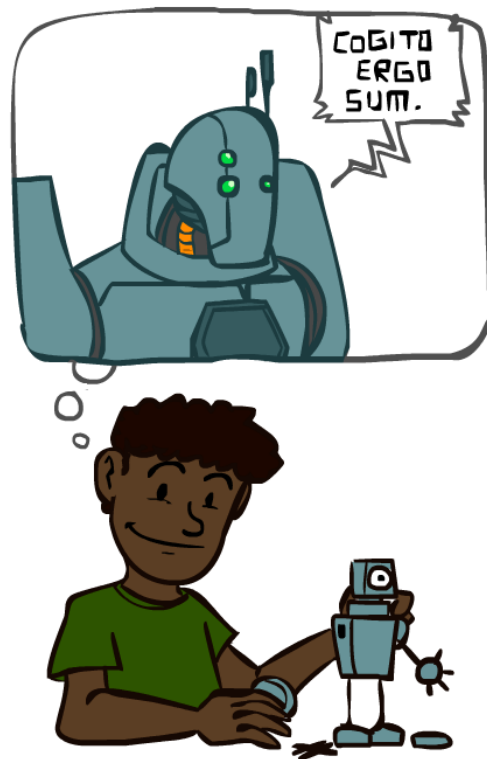


Maximize Your Expected Utility

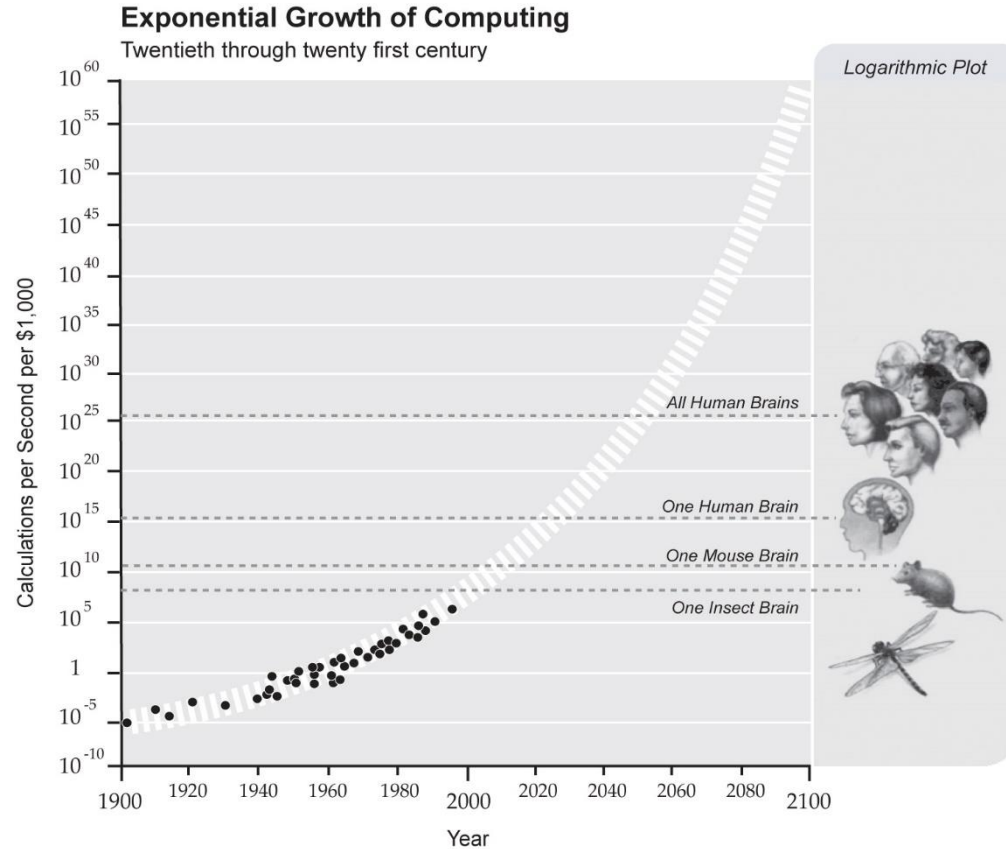


A (Short) History of AI

- **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: chess, checkers (RL), theorem proving
 - 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— 2012: Statistical approaches + subfield expertise**
 - Resurgence of probability, focus on uncertainty
 - General increase in technical depth
 - Agents and learning systems... "AI Spring"?
- **2012— ____: Excitement: Look, Ma, no hands again?**
 - Big data, big compute, deep learning
 - AI used in many industries



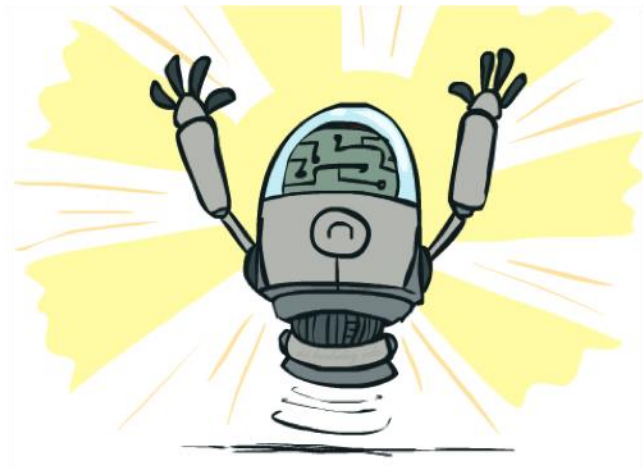
What Can AI Do?



What Can AI Do?

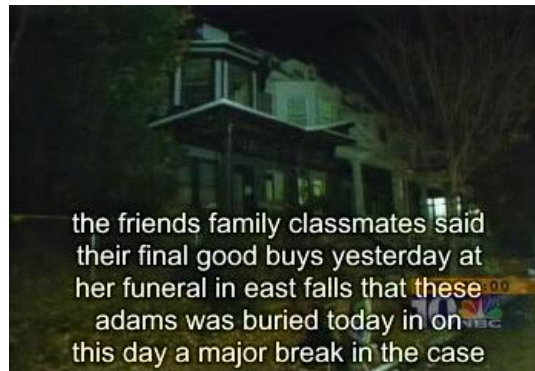
Quiz: Which of the following can be done at present?

- ✓ Play a decent game of table tennis?
- ✓ Play a decent game of Jeopardy?
- ✓ Drive safely along a curving mountain road?
- ? Drive safely along Xa Lo Ha Noi?
- ✓ Buy a week's worth of groceries on the web?
- ✗ Buy a week's worth of groceries at Big C?
- ? Discover and prove a new mathematical theorem?
- ✗ Converse successfully with another person for an hour?
- ? Perform a surgical operation?
- ✓ Put away the dishes and fold the laundry?
- ✓ Translate spoken Chinese into spoken English in real time?
- ✗ Write an intentionally funny story?



Natural Language

- Speech technologies (e.g. Siri)
 - Automatic speech recognition (ASR)
 - Text-to-speech synthesis (TTS)
 - Dialog systems
- Language processing technologies
 - Question answering
 - Machine translation



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

Computer Vision



"man in black shirt is playing guitar."



"construction worker in orange safety vest is working on road."



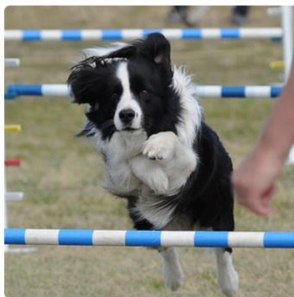
"two young girls are playing with lego toy."



"boy is doing backflip on wakeboard."



"girl in pink dress is jumping in air."



"black and white dog jumps over bar."



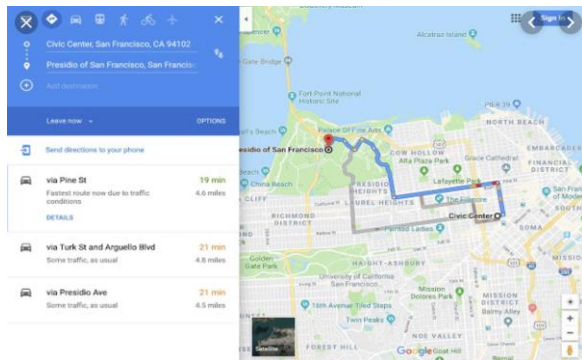
"young girl in pink shirt is swinging on swing."



"man in blue wetsuit is surfing on wave."

Karpathy & Fei-Fei, 2015; Donahue et al., 2015; Xu et al, 2015; many more

Tools for Predictions & Decisions



Game Agents

- **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



- **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.”



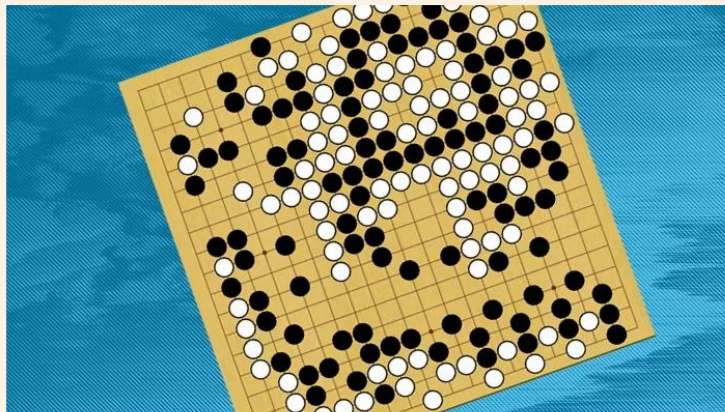




Human victory over AI

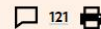
Man beats machine at Go in human victory over AI

Amateur Kellin Pelrine exploited weakness in systems that have otherwise dominated board game's grandmasters



Kellin Pelrine won 14 of 15 games against one of the top Go-playing systems in January © FT montage

Richard Waters in San Francisco FEBRUARY 18 2023



<https://www.ft.com/content/175e5314-a7f7-4741-a786-273219f433a1>

Human victory over AI

AlphaGo, a system devised by Google-owned research company DeepMind, defeated the world Go champion Lee Sedol by four games to one in 2016. Sedol attributed his [retirement from Go](#) three years later to the rise of AI, saying that it was “an entity that cannot be defeated”. AlphaGo is not publicly available, but the systems Pelrine prevailed against are considered on a par.

In a game of Go, two players alternately place black and white stones on a board marked out with a 19x19 grid, seeking to encircle their opponent's stones and enclose the largest amount of space. The huge number of combinations means it is impossible for a computer to assess all potential future moves.

The tactics used by Pelrine involved slowly stringing together a large “loop” of stones to encircle one of his opponent's own groups, while distracting the AI with moves in other corners of the board. The Go-playing bot did not notice its vulnerability, even when the encirclement was nearly complete, Pelrine said.

“As a human it would be quite easy to spot,” he added.

The discovery of a weakness in some of the most advanced Go-playing machines points to a fundamental flaw in the deep learning systems that underpin today's most advanced AI, said Stuart Russell, a computer science professor at the University of California, Berkeley.

Recommended

[John Thornhill](#)

AI warfare can empower the bad guys as well as the good

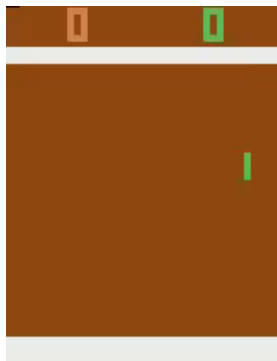
The systems can “understand” only specific situations they have been exposed to in the past and are unable to generalise in a way that humans find easy, he added.

“It shows once again we've been far too hasty to ascribe superhuman levels of intelligence to machines,” Russell said.

The precise cause of the Go-playing systems' failure is a matter of conjecture, according to the researchers. One likely reason is that the tactic exploited by Pelrine is rarely used, meaning the AI systems had not been trained on enough similar games to realise they were vulnerable, said Gleave.

Game Agents

- Reinforcement learning



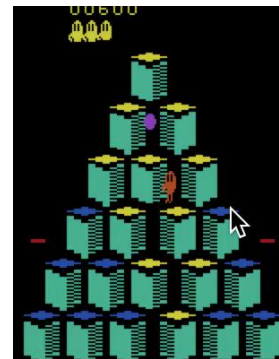
Pong



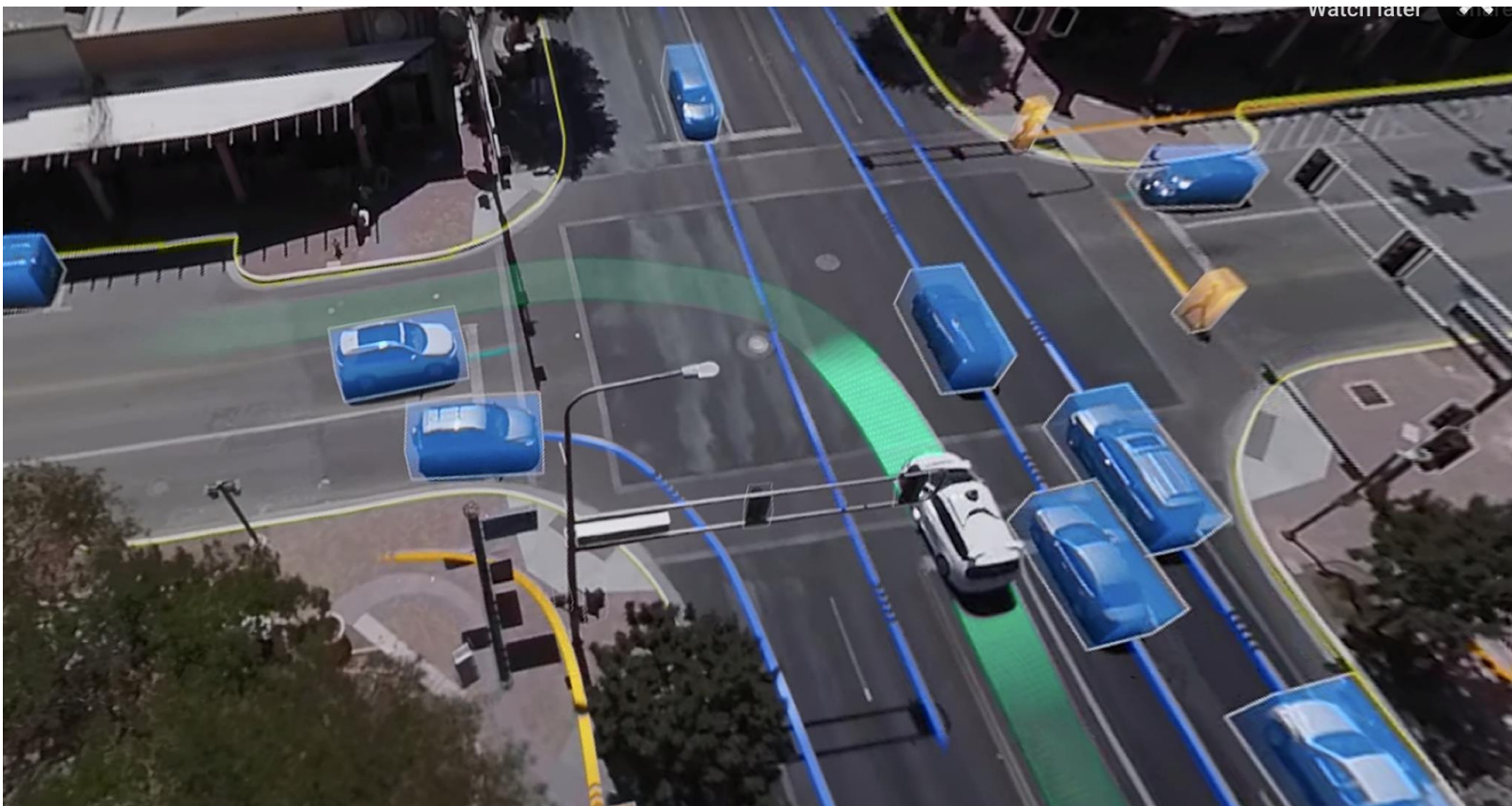
Enduro

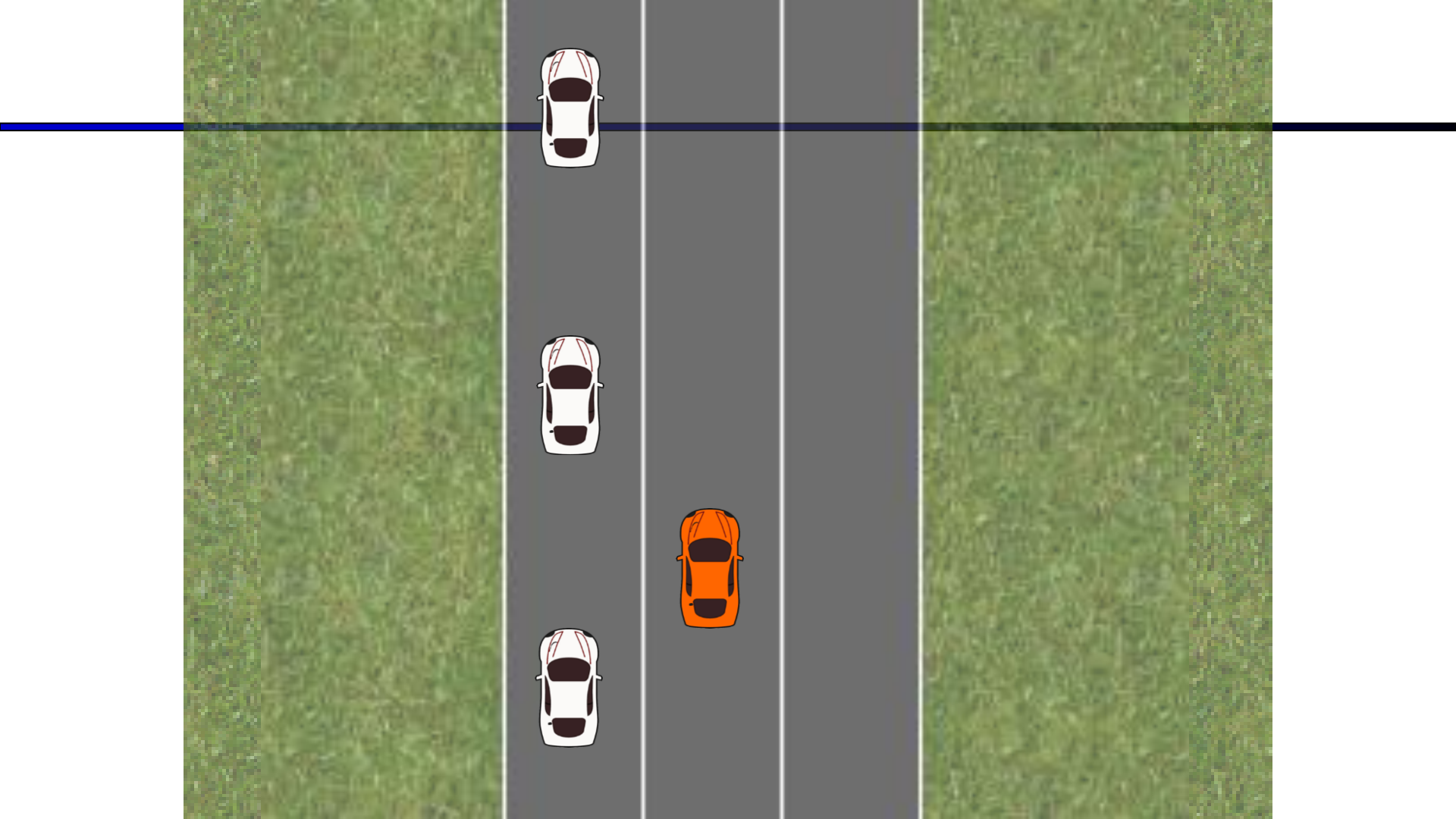


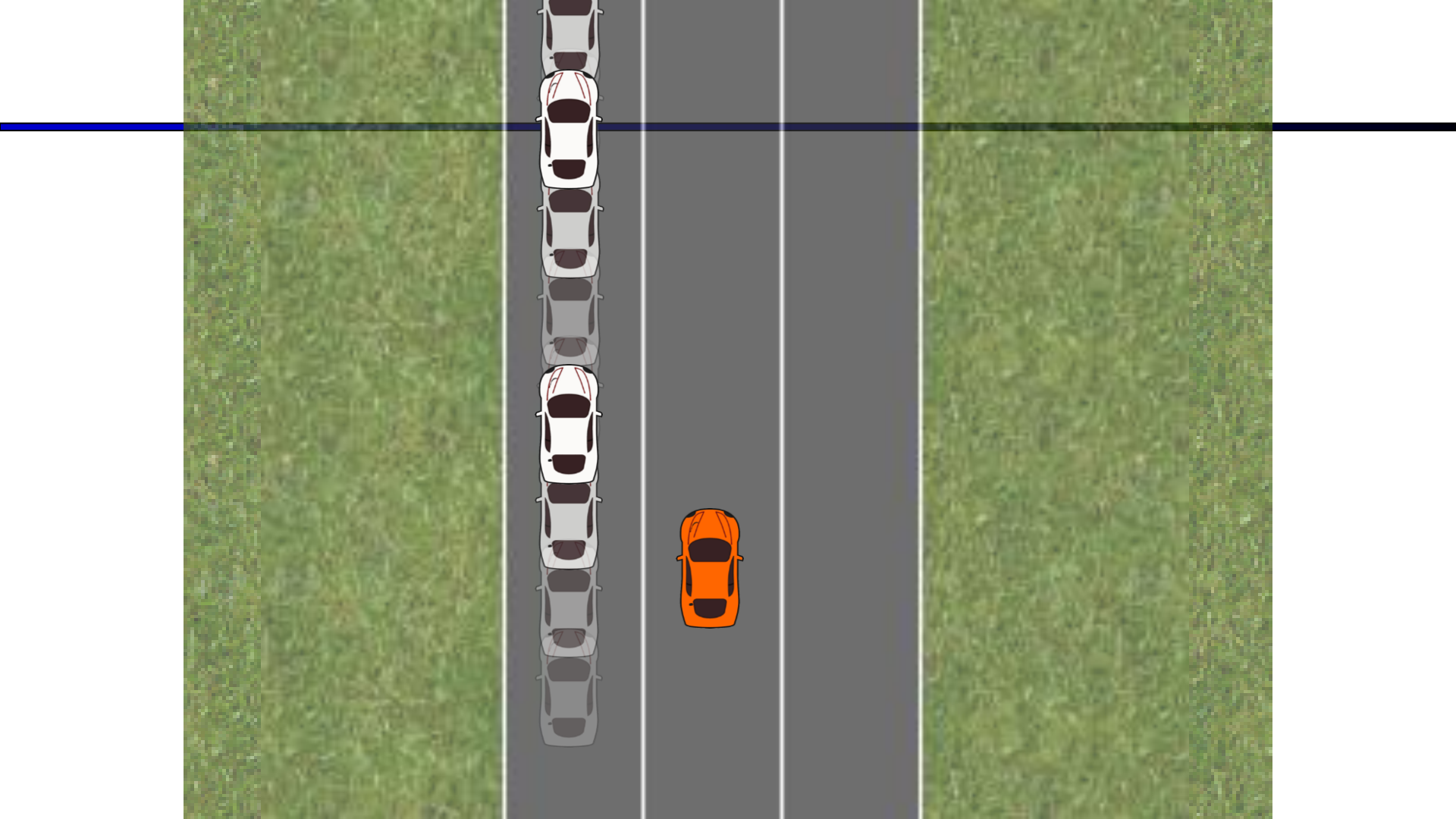
Beamrider

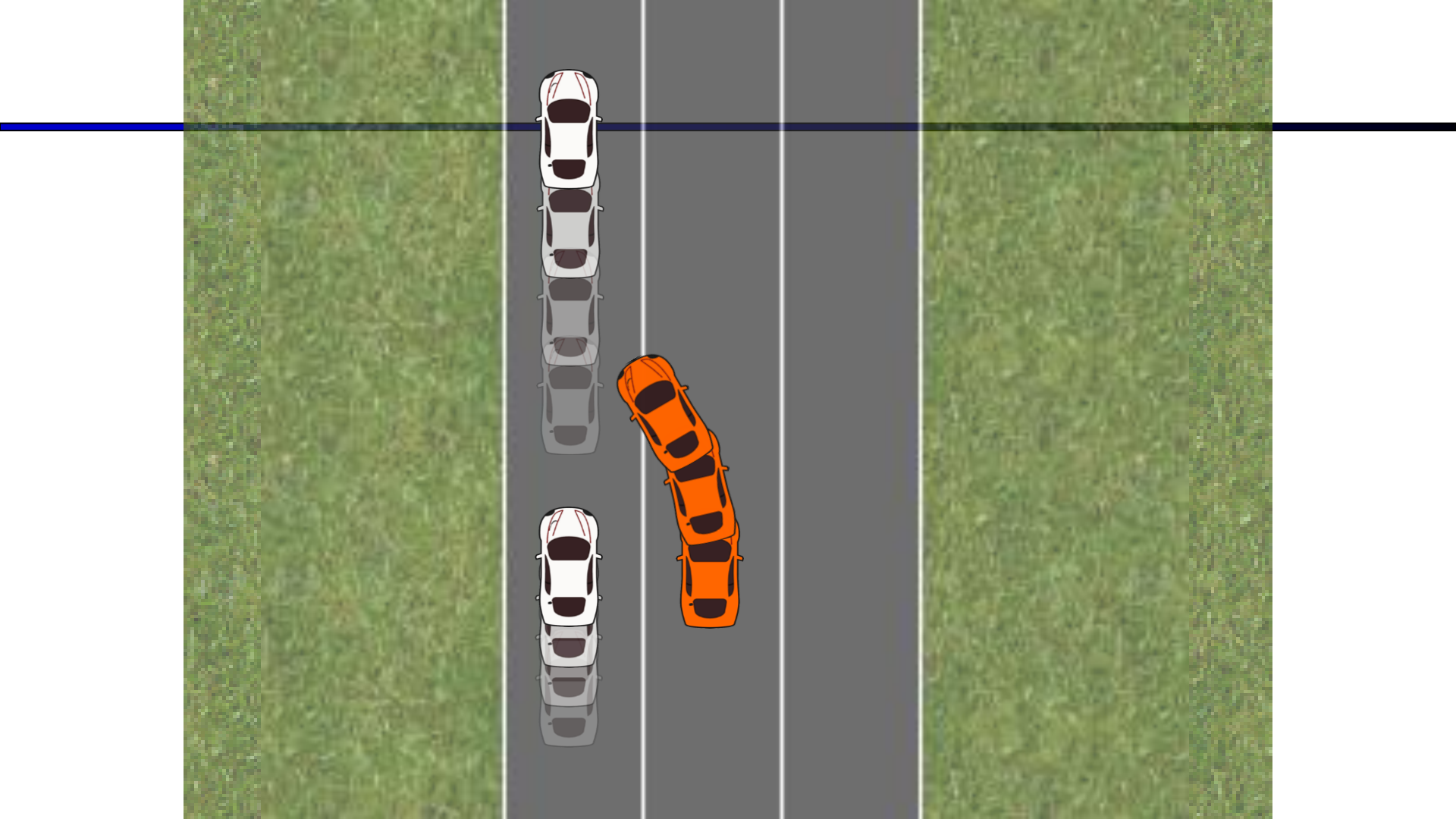


Q*bert

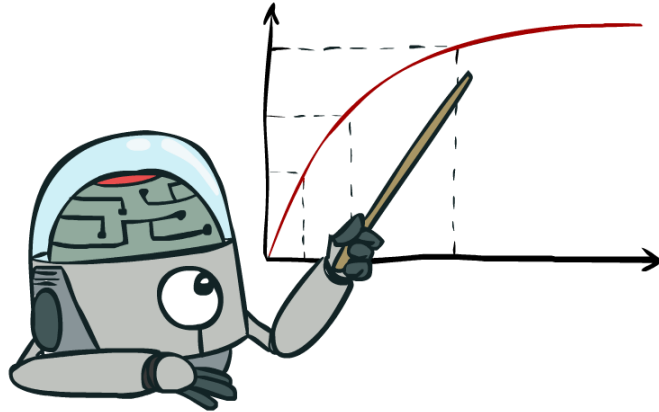








Maximize Your Expected Utility

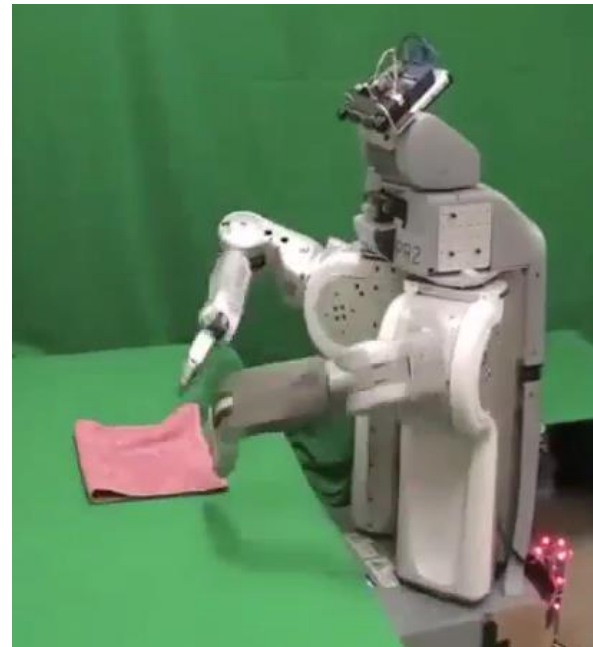


Utility?

Clear utility function



Not so clear utility function



AI

Machine Learning

[learning decisions;
sometimes independent]

Robots

[physically
embodied]

Rational
Agents

[decisions]

Human-AI
Interaction

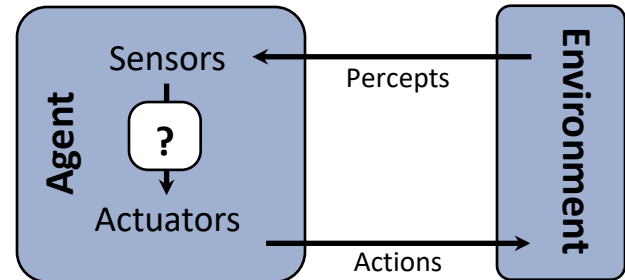
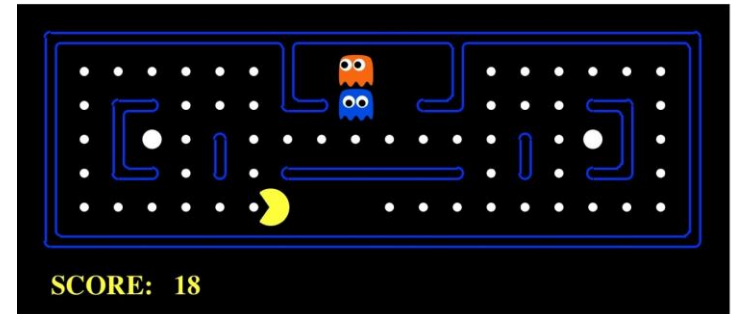
NLP

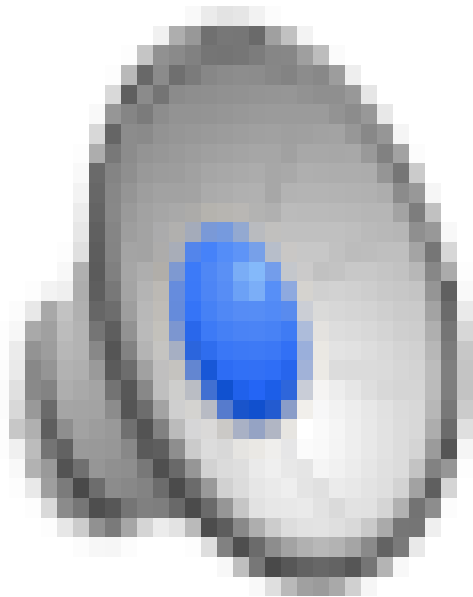
Computer
Vision

AI as Designing Rational Agents

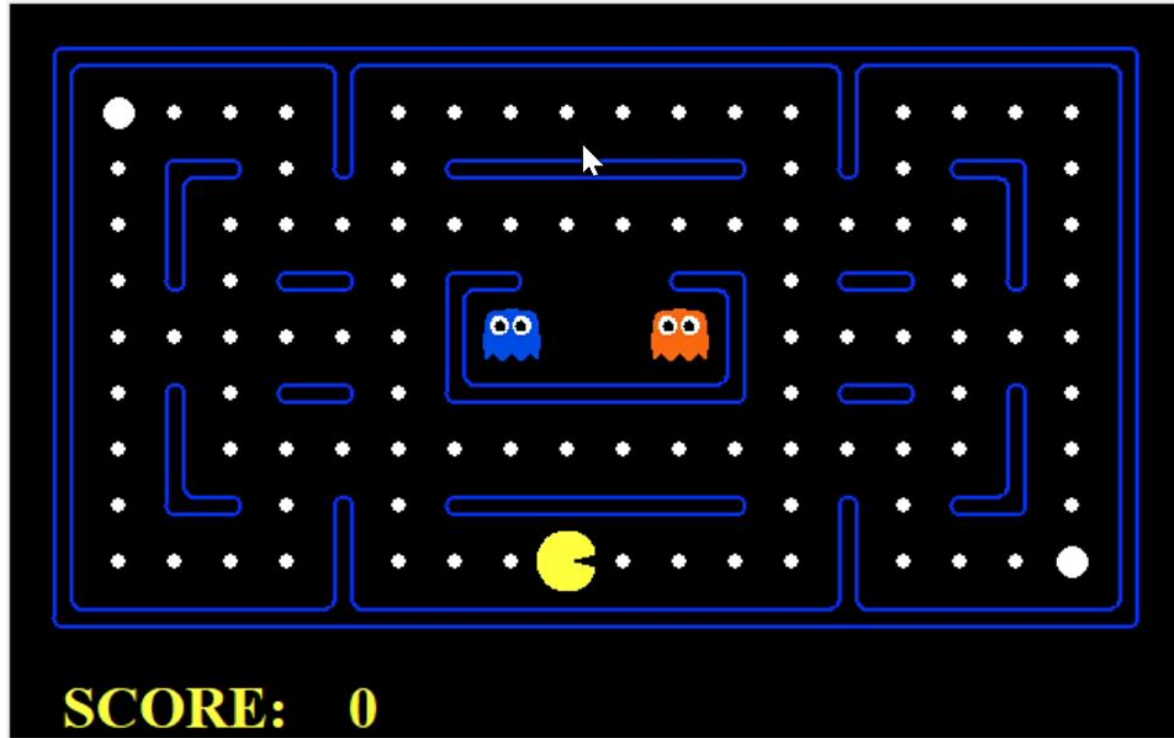
- An **agent** is an entity that *perceives* and *acts*.
- A **rational agent** selects actions that maximize its expected utility.
- Characteristics of the **sensors, actuators, and environment** dictate techniques for selecting rational actions
- **This course is about:**
 - General AI techniques for many problem types
 - Learning to choose and apply the technique appropriate for each problem

<https://www.youtube.com/watch?v=16Dir4QqCUg&t=4900s>





Designing Rational Agents



(Tentative) Course Topics

- Part I: Search

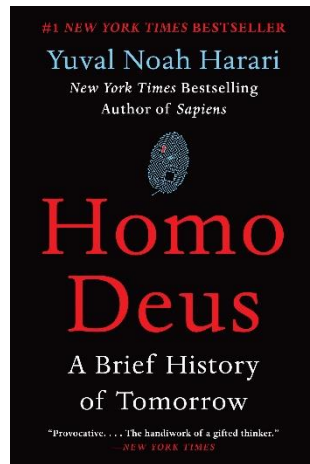
- Uninformed Search
- Informed Search / Heuristics
- Constraint Satisfaction Problems (CSPs)
- Adversarial Search and Expectimax Search

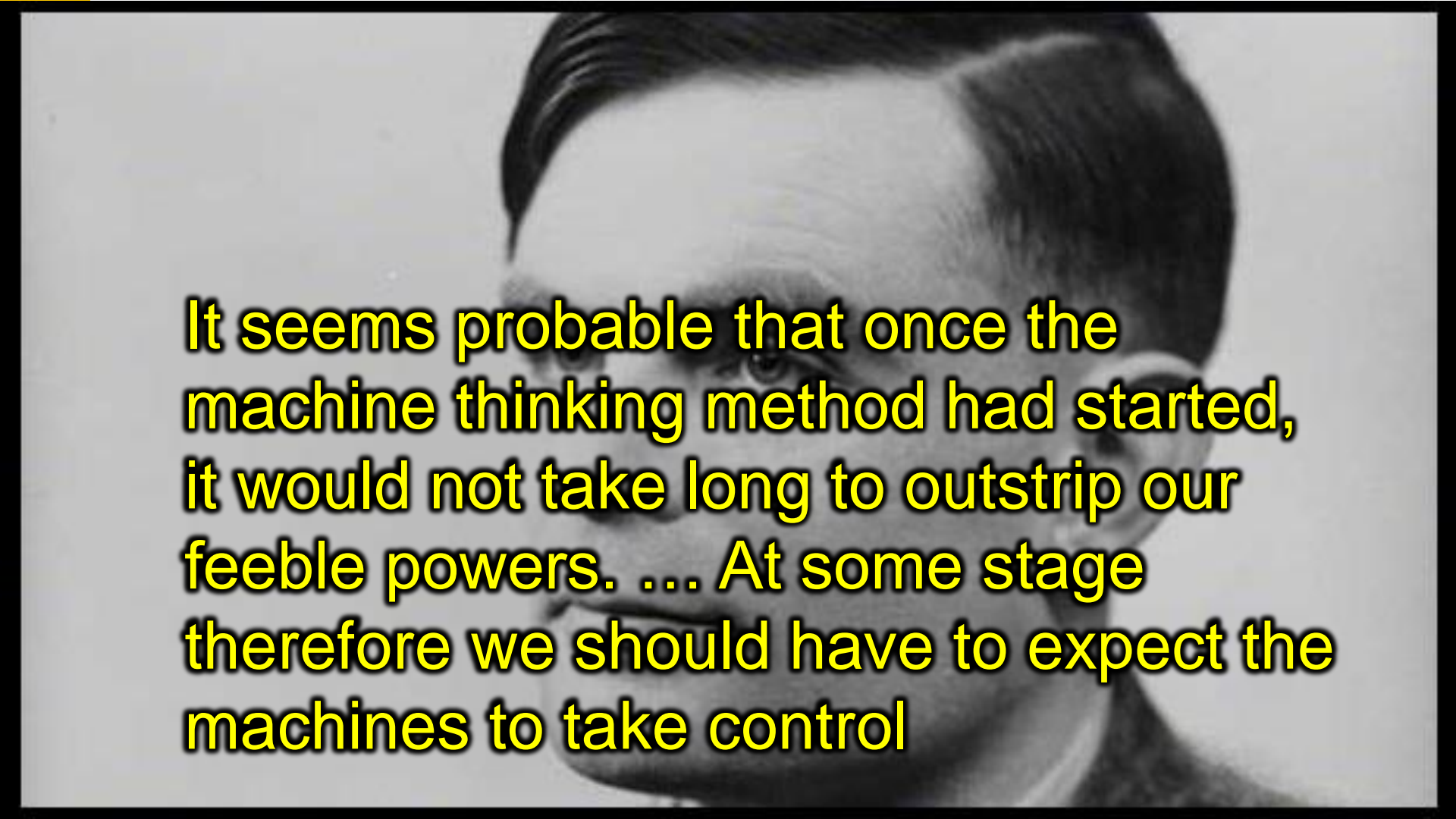
- Part II: Reinforcement Learning

- Markov Decision Processes, Bellman Equations, and Dynamic Programming
- Temporal Difference Learning, SARSA, and Q-Learning
- Approximate Q-Learning and Deep Q-Learning

Future

- We are doing AI...
 - To create intelligent systems
 - The more intelligent, the better
 - To gain a better understanding of human intelligence
 - To magnify those benefits that flow from it
 - E.g., net present value of human-level AI
 - Might help us avoid war and ecological catastrophes, achieve immortality and expand throughout the universe
- What if we succeed?





It seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers. ... At some stage therefore we should have to expect the machines to take control

What's bad about better AI?

- AI that is incredibly good at achieving something other than what we really want
- AI, economics, statistics, operations research, control theory all assume utility to be *fixed, known, and exogenously specified*
 - ~~Machines are intelligent to the extent that their actions can be expected to achieve their objectives~~
 - Machines are beneficial to the extent that their actions can be expected to achieve our objectives

A new model for AI

1. The machine's only objective is to maximize the realization of human preferences
2. The robot is initially uncertain about what those preferences are
3. Human behavior provides evidence about human preferences

The standard model of AI is a special case, where the human can exactly and correctly program the objective into the machine

A new model for AI

TP HCM đặt hàng ứng dụng ChatGPT cho quản lý công

Bốn nhóm lĩnh vực ứng dụng ChatGPT được lãnh đạo TP HCM đặt hàng nhà khoa học với mong muốn nâng cao hiệu quả quản lý nhà nước, phục vụ người dân tốt hơn.

Ngày 1/3, Sở Thông tin Truyền thông TP HCM công bố bốn nhóm lĩnh vực ứng dụng ChatGPT mà thành phố đặt hàng với mục tiêu tìm kiếm các giải pháp giúp nâng cao hiệu quả quản lý nhà nước ứng dụng trí tuệ nhân tạo.

Theo đó thành phố tiếp nhận các giải pháp ứng dụng ChatGPT trong các dịch vụ công trực tuyến, trả lời tiến độ thực hiện cho người dân; Ứng dụng ChatGPT vào [tổng đài 1022](#) nhằm ghi nhận ý kiến và trả lời kiến nghị.

Các ứng dụng ChatGPT trong hỗ trợ lãnh đạo bằng các giải pháp nghiên cứu xây dựng hệ thống trợ lý ảo, đăng ký và kiểm tra lịch làm việc, tóm tắt hồ sơ tài liệu...

Về hoạt động giáo dục, các ứng dụng ChatGPT cần giúp thành phố xây dựng hệ thống trợ lý học tập bằng các giải pháp trợ lý ảo học tập phục vụ cho thầy cô, học sinh các cấp.