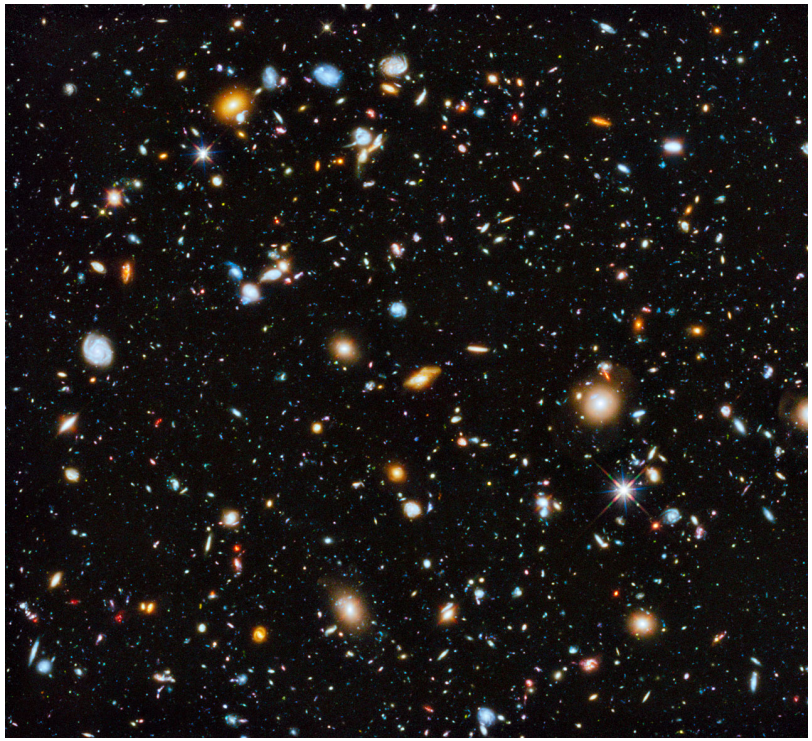
Lecture 1: Overview



Roadmap

What is AI?

Course overview

Course logistics

Optimization

Microsoft creates AI that can read a document and answer questions about it as well as a person

January 15, 2018 | Allison Linn

Microsoft researchers achieve new conversational speech recognition milestone

August 20, 2017 | By Xuedong

June 24, 2014

DeepFace: Closing Performance in Fa

Conference on Computer Vision and Pattern Recognition (CVPR)

By: **Yaniv Taigman**, Ming Yang, **Marc'Aurelio Ranzato**, **Lior Wolf**

Abstract

In modern face recognition, the conventional pipeline co
classify. We revisit both the alignment step and the repre
modeling in order to apply a piecewise affine transforma
layer deep neural network. This deep network involves m
locally connected layers without weight sharing, rather tl
trained it on the largest facial dataset to-date, an identity
belonging to more than 4,000 identities.

If you think AI will never replace radiologists—you may want to think again

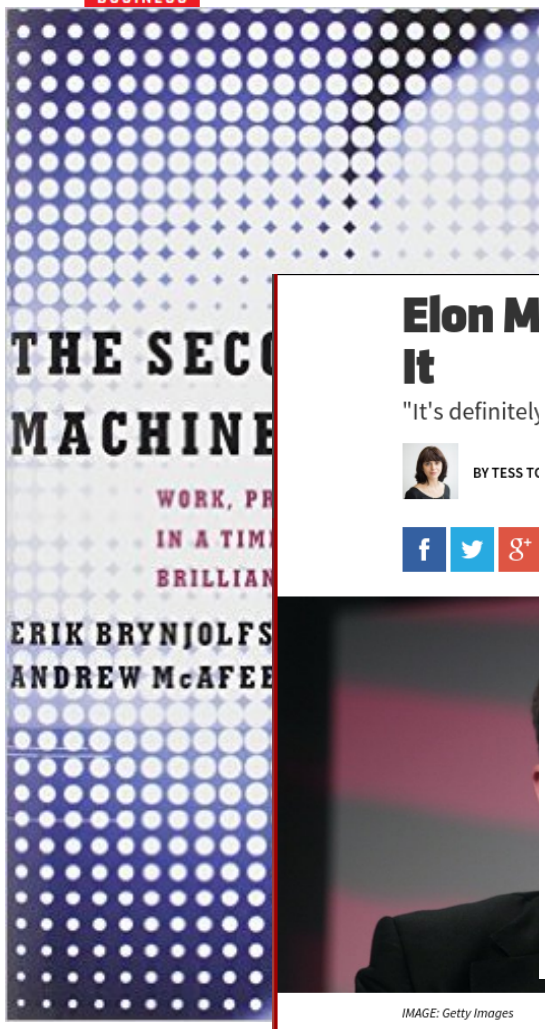
May 14, 2018 | *Michael Walter* | *Artificial Intelligence*

It's one of the most frequently discussed questions in radiology today: What kind of long-term impact will artificial intelligence (AI) have on radiologists?

Robert Schier, MD, a radiologist for RadNet, shared his own thoughts on the topic in a new commentary published by the *Journal of the American College of Radiology*—and he's not quite as optimistic as some of his colleagues throughout the industry.

The Atlantic SUBSCRIBE SEARCH MENU

BUSINESS



Technology

Stephen Hawking warns artificial intelligence could end mankind

By Rory Cellan-Jones
Technology correspondent

2 December 2014 | Technology

Elon M
It
"It's definitely

BY TESS TO

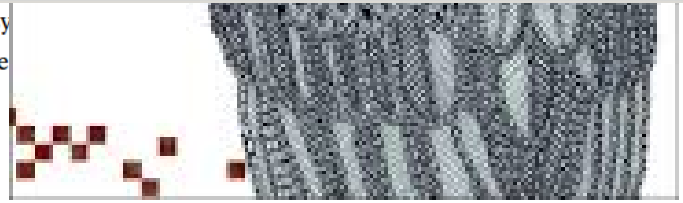
IMAGE: Getty Images

Elon Musk has emerged as a leading voice in speaking out on the potential dangers of artificial intelligence, going so far as to call it the "biggest existential threat" to

Take a Video Tour of Facebook's Frank Gehry-Designed New York City Office

HIT THE ROAD

The advances we'
humanoid robots, speech recognition and sy
Jeopardy!-champion computers—are not the



Companies

Google "An important shift from a mobile first world to an AI first world" [CEO Sundar Pichai @ Google I/O 2017]

Microsoft Created AI and Research group as 4th engineering division, now 8K people [2016]

facebook Created Facebook AI Research, Mark Zuckerberg very optimistic and invested

Others: IBM, Amazon, Apple, Uber, Salesforce, Baidu, Tencent, etc.

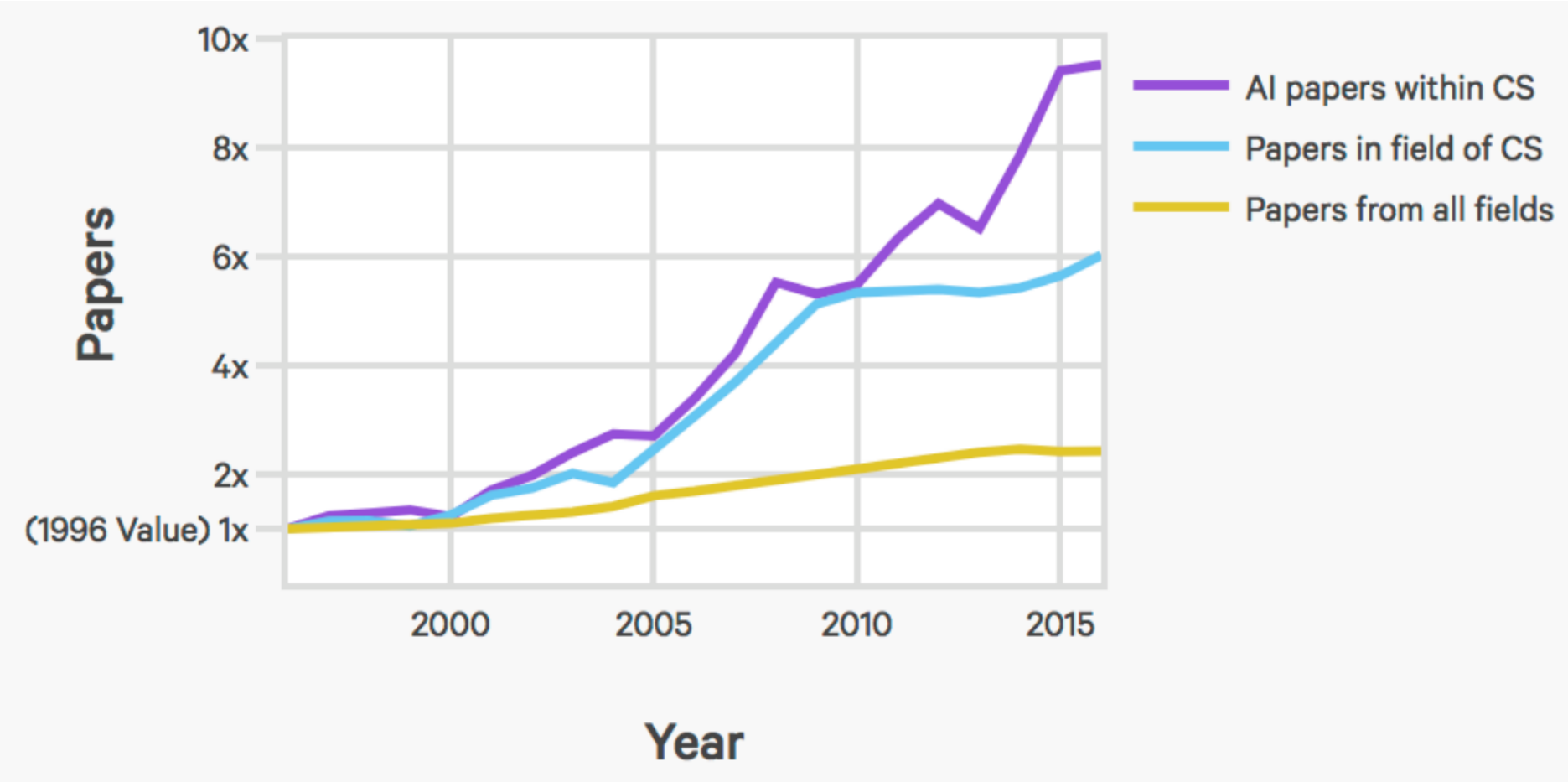
Governments

"AI holds the potential to be a major driver of economic growth and social progress" [White House report, 2016]

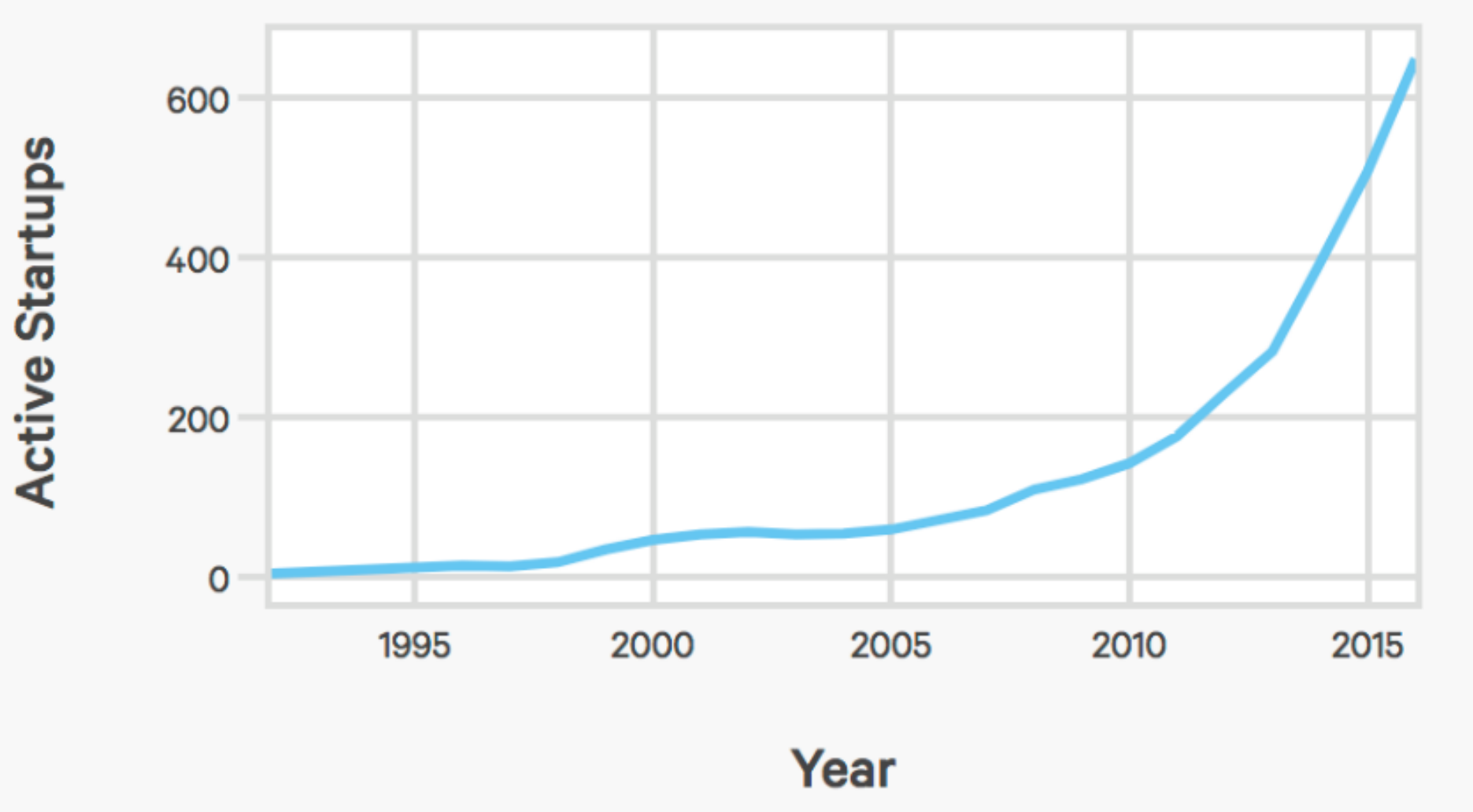
Released domestic strategic plan to become world leader in AI by 2030 [2017]

"Whoever becomes the leader in this sphere [AI] will become the ruler of the world" [Putin, 2017]

AI index: number of published AI papers



AI index: number of AI startups



AI index: AI conference attendance

Ok, really, what is AI?

Two views of AI

AI agents: how can we re-create intelligence?

AI tools: how can we benefit society?

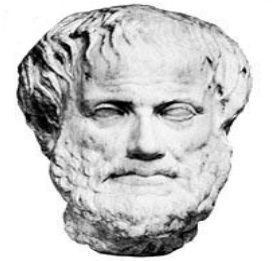
AI agents...

An intelligent agent

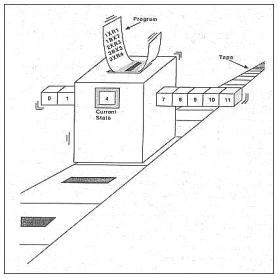
Perception Robotics Language

Knowledge Reasoning Learning

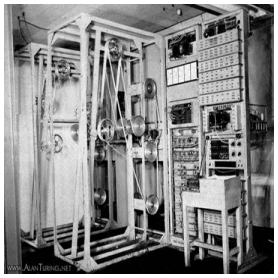
Pre-AI developments



Philosophy: **intelligence** can be achieved via mechanical computation (e.g., Aristotle)



Church-Turing thesis (1930s): any computable function is **computable** by a Turing machine

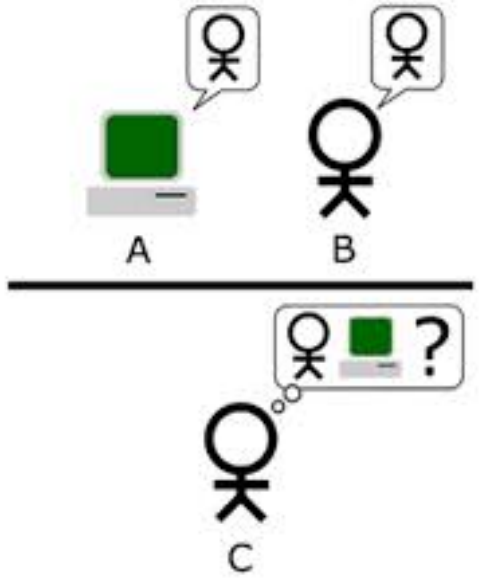


Real computers (1940s): actual **hardware** to do it: Heath Robinson, Z-3, ABC/ENIAC

The Turing Test (1950)

[Turing, 1950. Computing Machinery and Intelligence]

"Can machines think?"



Q: Please write me a sonnet on the subject of the Forth Bridge.

A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764.

A: (Pause about 30 seconds and then give as answer) 105621.

Tests behavior — simple and objective

Birth of AI (1956)

Workshop at Dartmouth College; attendees: John McCarthy, Marvin Minsky, Claude Shannon, etc.

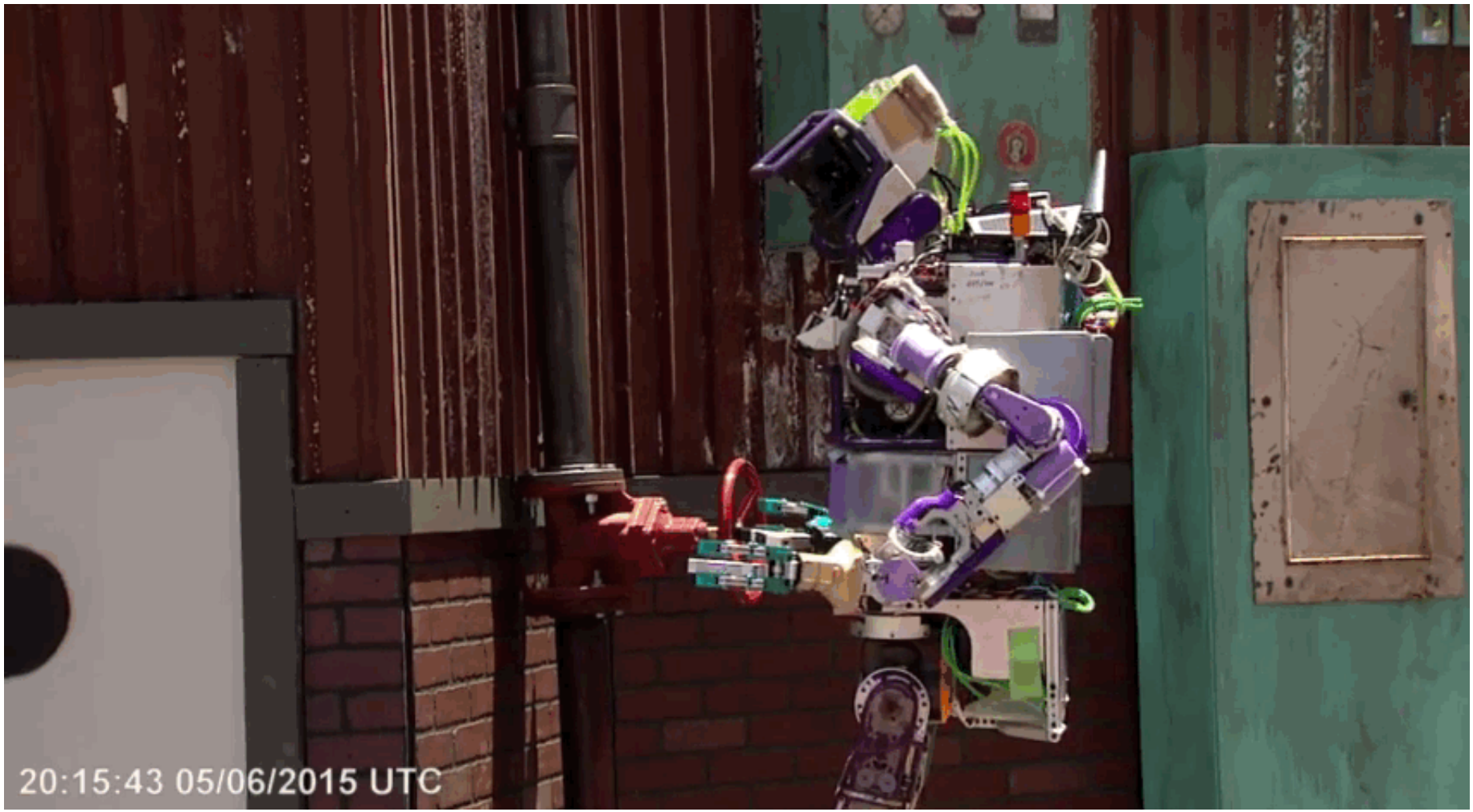
Aim for **general principles**:

Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it.

A very brief history

- 1956: Dartmouth workshop, John McCarthy coined "AI"
- 1960: checkers playing program, Logical Theorist
- 1966: ALPAC report cuts off funding for translation
- 1974: Lighthill report cuts off funding in UK
- 1970-80s: expert systems (XCON, MYCIN) in industry
- 1980s: Fifth-Generation Computer System (Japan); Strategic Computing Initative (DARPA)
- 1987: collapse of Lisp market, government funding cut
- 1990-: rise of machine learning
- 2010s: heavy industry investment in deep learning

2015 DARPA Robotics Challenge



[Li et al., 2016]

Open-domain dialogue

A: How old are you?

B: I'm 16. Why are you asking?

A: I thought you were 12.

B: What made you think so?

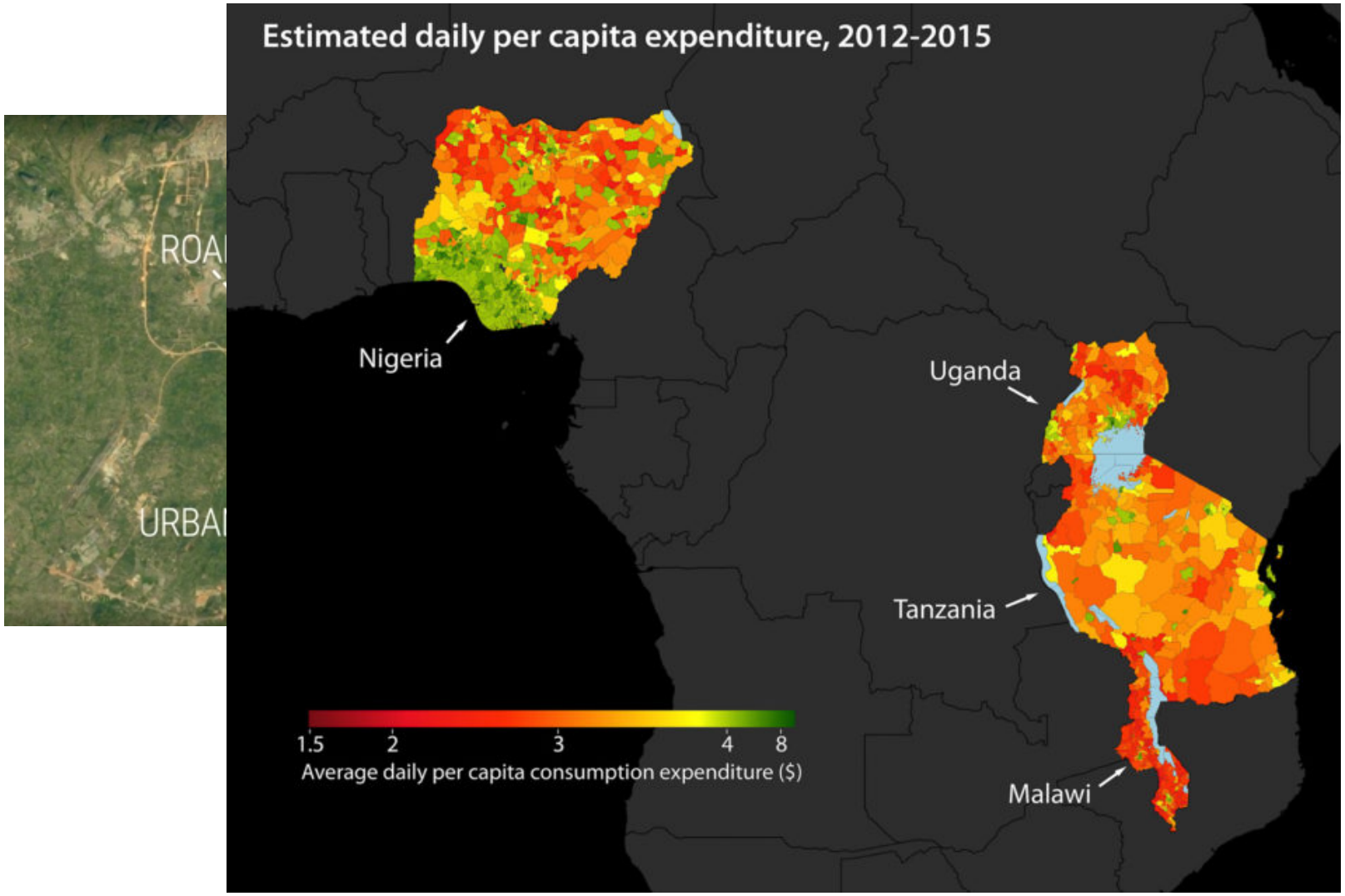
A: I don't know what you are talking about.

B: You don't know what you are saying.

AI tools...

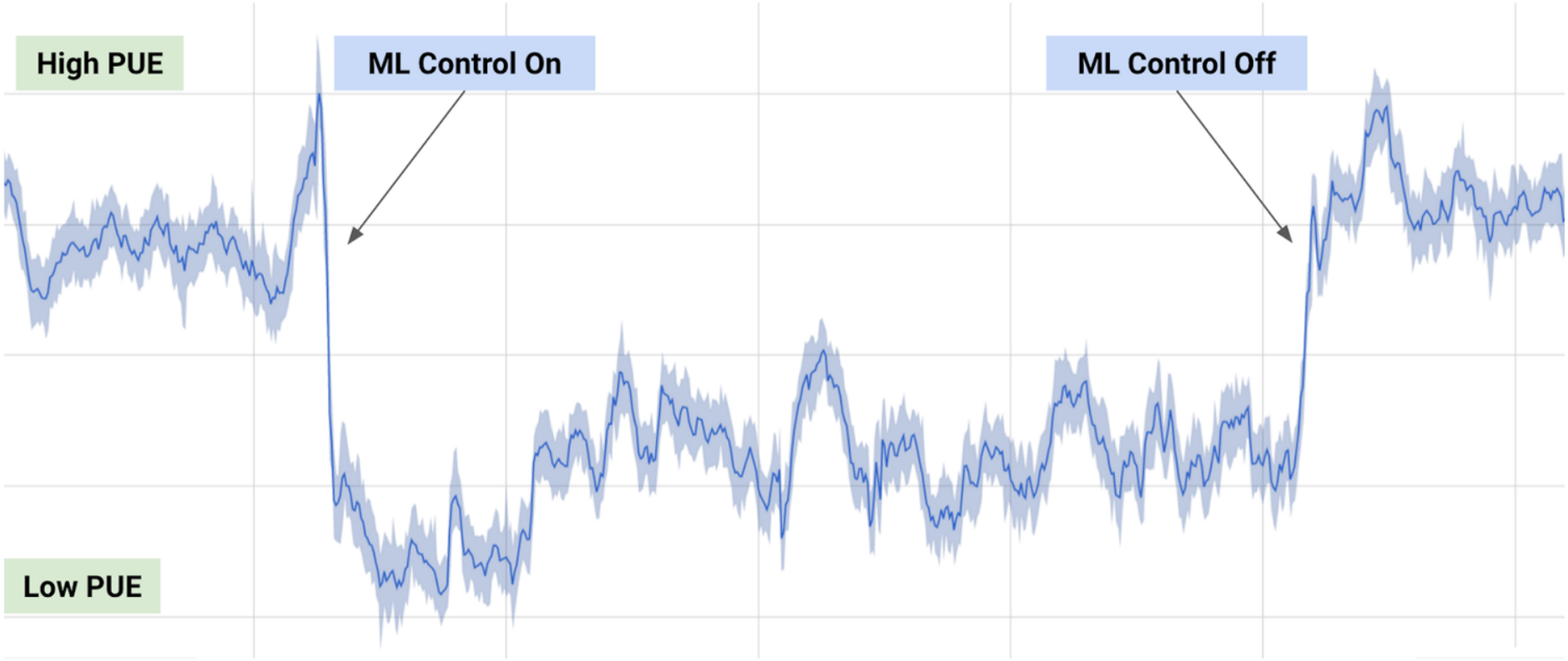
[Jean et al. 2016]

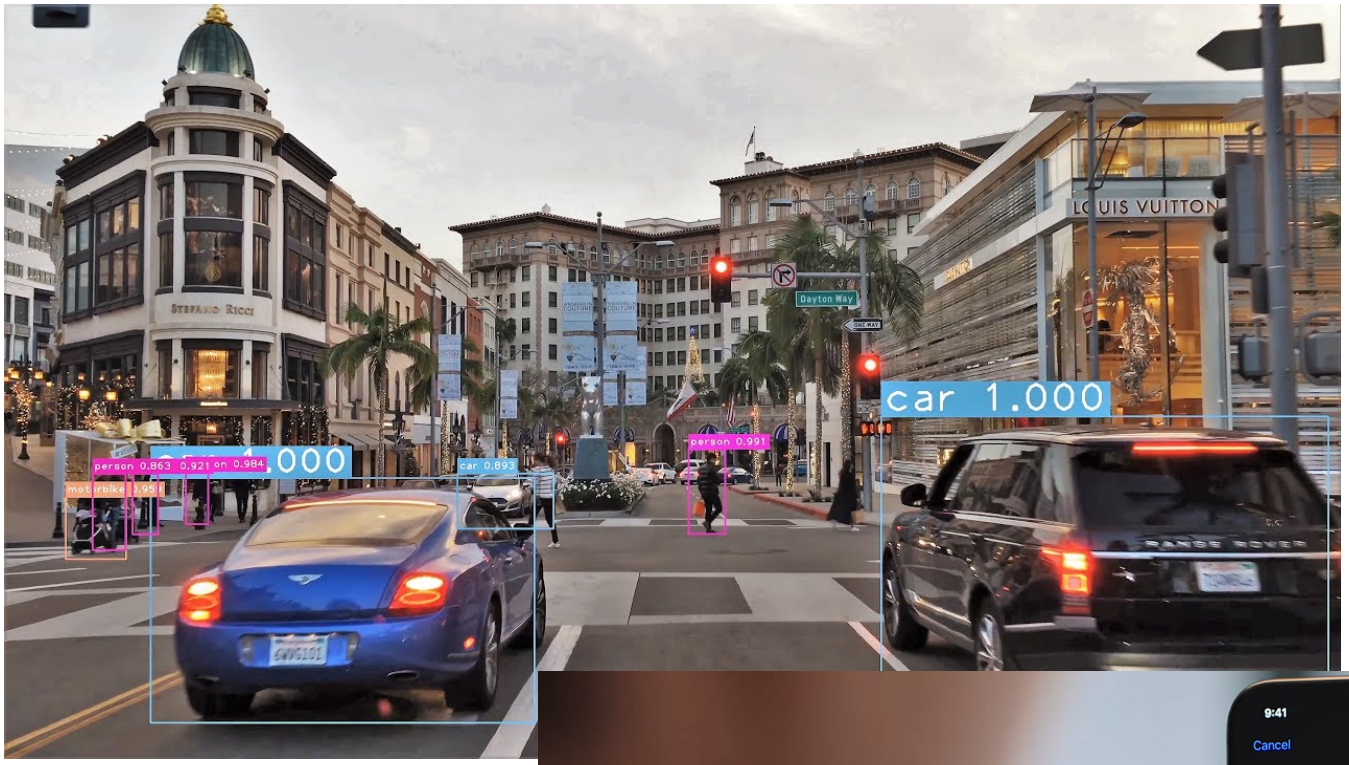
Predicting poverty

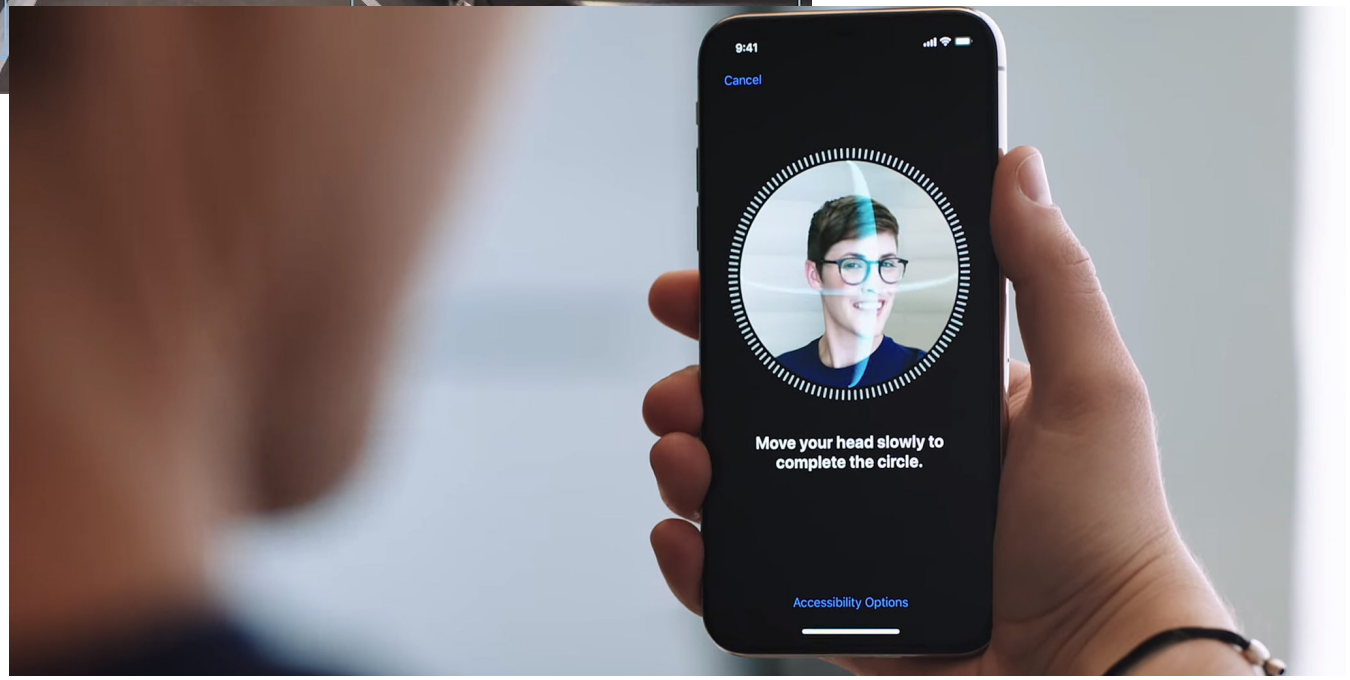


[DeepMind]

Saving energy by cooling datacenters




STEFANO RICCI
LOUIS VUITTON
Dayton Way
ONE WAY
car 1.000
person 0.991
car 0.893
person 0.863
motorbike 0.951

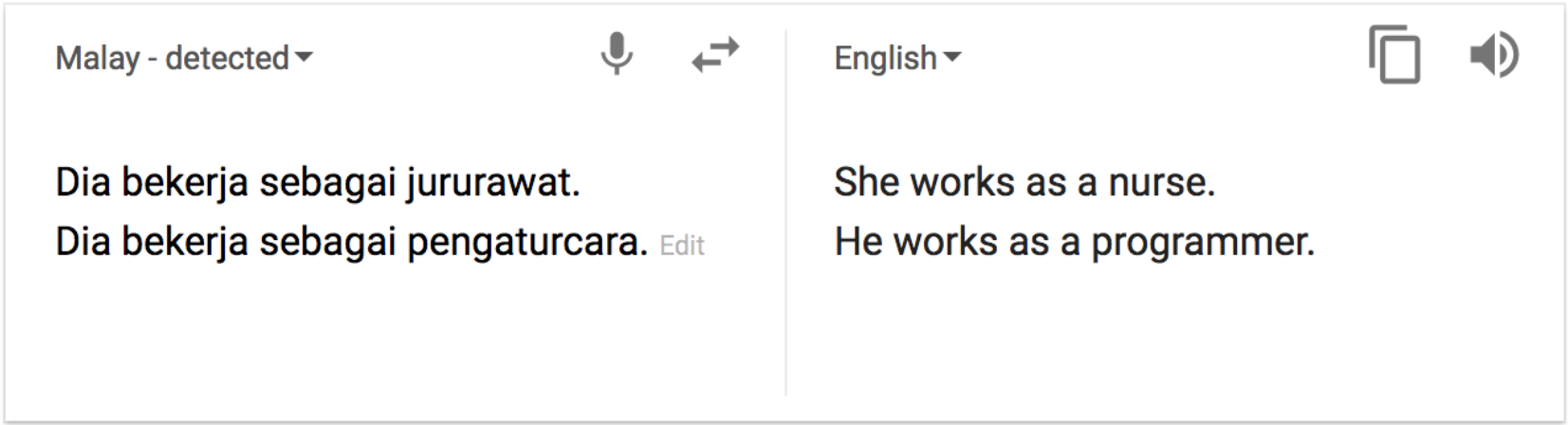

9:41
Cancel
Move your head slowly to complete the circle.
Accessibility Options

Security

[Evtimov+ 2017]

[Sharif+ 2016]

Bias in machine translation



society $\Rightarrow$ data $\Rightarrow$ predictions

Fairness in criminal risk assessment

- Northpointe: COMPAS predicts criminal risk score (1-10)
- ProPublica: given that an individual did not reoffend, blacks 2x likely to be (wrongly) classified 5 or above
- Northpointe: given a risk score of 7, 60% of whites reoffended, 60% of blacks reoffended

Question

What inspires you more?

Building agents with human-level intelligence

Developing tools that can benefit society

Summary so far

- AI dream of achieving human-level intelligence is ongoing
- Still lots of open research questions
- AI is having huge societal impact
- Need to think carefully about real-world consequences

Roadmap

What is AI?

Course overview

Course logistics

Optimization

How do we solve AI tasks?

Paradigm

Modeling

Inference **Learning**

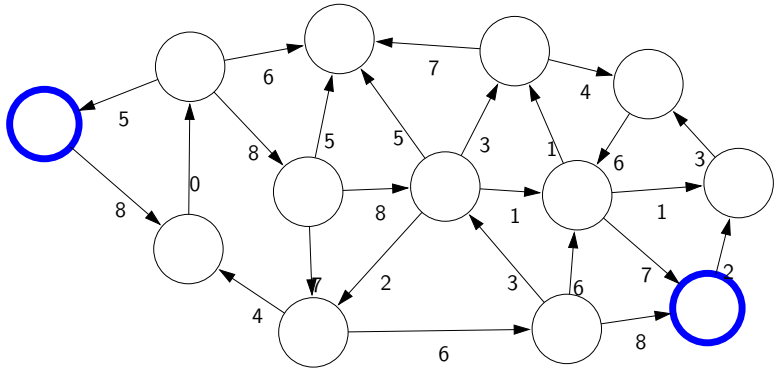
- In this class, we will adopt the **modeling-inference-learning** paradigm to help us navigate the solution space. In reality, the lines are blurry, but this paradigm serves as an ideal and a useful guiding principle.

Paradigm: modeling

Real world

Modeling

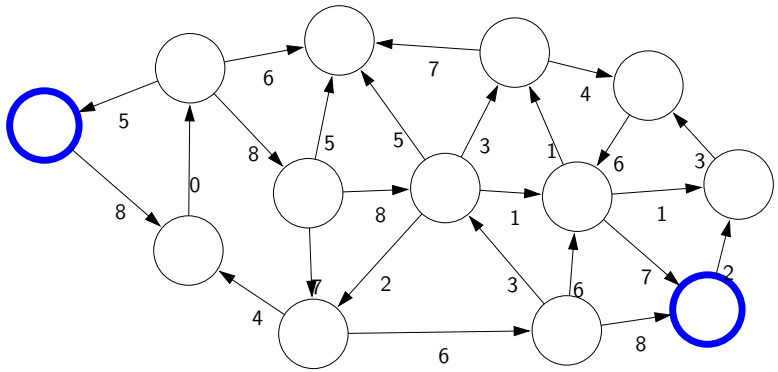
Model



- The first pillar is modeling. Modeling takes messy real world problems and packages them into neat formal mathematical objects called **models**, which can be subject to rigorous analysis but is more amenable to what computers can operate on. However, modeling is lossy: not all of the richness of the real world can be captured, and therefore there is an art of modeling: what does one keep versus ignore? (An exception to this is games such as Chess or Go or Sodoku, where the real world is identical to the model.)
- As an example, suppose we're trying to have an AI that can navigate through a busy city. We might formulate this as a graph where nodes represent points in the city.

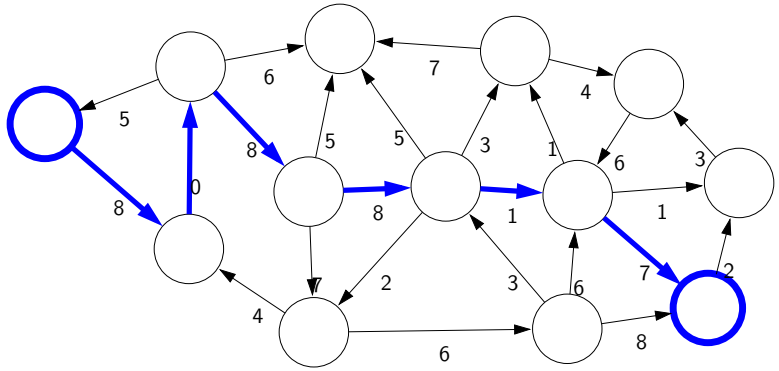
Paradigm: inference

Model



Inference

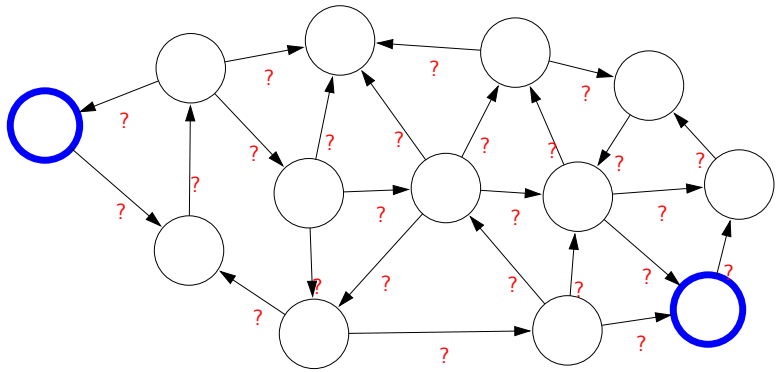
Predictions



- The second pillar is inference. Given a model, the task of **inference** is to answer questions with respect to the model. For example, given the model of the city, one could ask questions such as: what is the shortest path? what is the cheapest path?
- For some models, computational complexity can be a concern (games such as Go), and usually approximations are needed.

Paradigm: learning

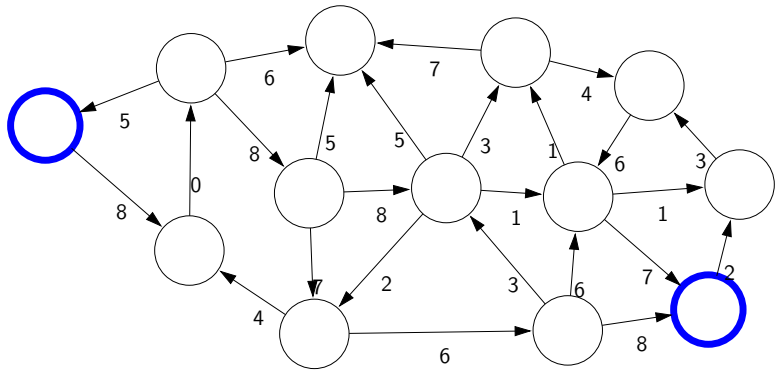
Model without parameters



+data

Learning

Model with parameters



- But where does the model come from? Remember that the real world is rich, so if the model is to be faithful, the model has to be rich as well. But we can't possibly write down such a rich model manually.
- The idea behind (machine) **learning** is to instead get it from data. Instead of constructing a model, one constructs a skeleton of a model (more precisely, a model family), which is a model without parameters. And then if we have the right type of data, we can run a machine learning algorithm to tune the parameters of the model.

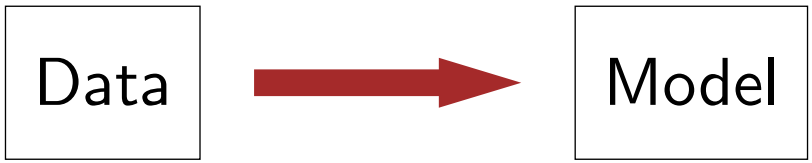
Course plan

"Low-level intelligence" "High-level intelligence"

Machine learning

- We now embark on our tour of the topics in this course. The topics correspond to types of models that we can use to represent real-world tasks. The topics will in a sense advance from low-level intelligence to high-level intelligence, evolving from models that simply make a reflex decision to models that are based on logical reasoning.

Machine learning



- The main driver of recent successes in AI
- Move from "code" to "data" to manage the information complexity
- Requires a leap of faith: **generalization**

- Supporting all of these models is **machine learning**, which has been arguably the most crucial ingredient powering recent successes in AI. Conceptually, machine learning allows us to shift the information complexity of the model from code to data, which is much easier to obtain (either naturally occurring or via crowdsourcing).
- The main conceptually magical part of learning is that if done properly, the trained model will be able to produce good predictions beyond the set of training examples. This leap of faith is called **generalization**, and is, explicitly or implicitly, at the heart of any machine learning algorithm. This can even be formalized using tools from probability and statistical learning theory.

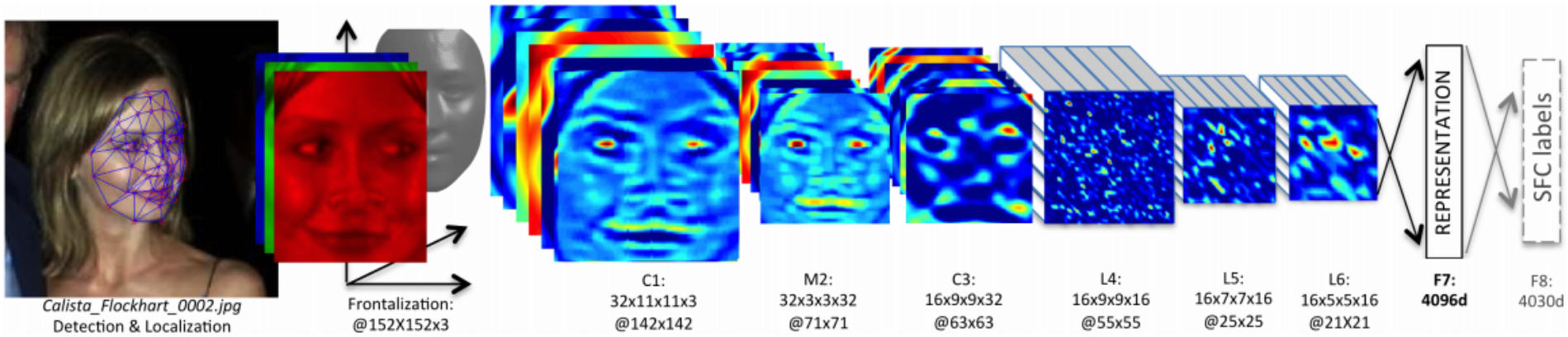
Course plan


Reflex
"Low-level intelligence"
"High-level intelligence"
Machine learning

What is this animal?

Reflex-based models

- Examples: linear classifiers, deep neural networks



- Most common models in machine learning

- Fully feed-forward (no backtracking)

- The idea of a reflex-based model simply performs a fixed sequence of computations on a given input. Examples include most models found in machine learning from simple linear classifiers to deep neural networks. The main characteristic of reflex-based models is that their computations are feed-forward; one doesn't backtrack and consider alternative computations. Inference is trivial in these models because it is just running the fixed computations, which makes these models appealing.

Course plan

Search problems

Markov decision processes

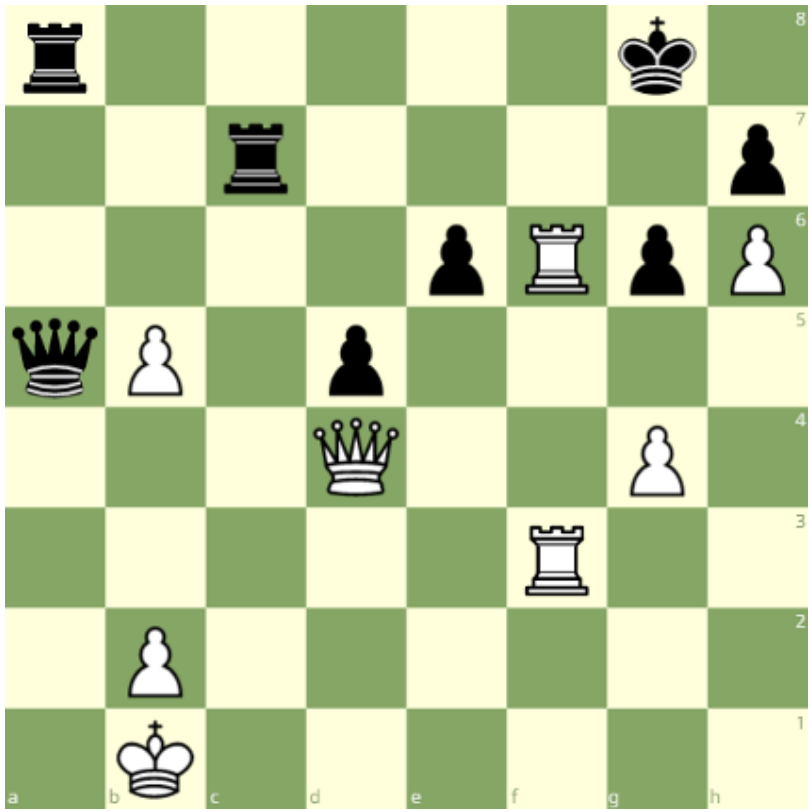
Adversarial games

Reflex **States**

"Low-level intelligence" "High-level intelligence"

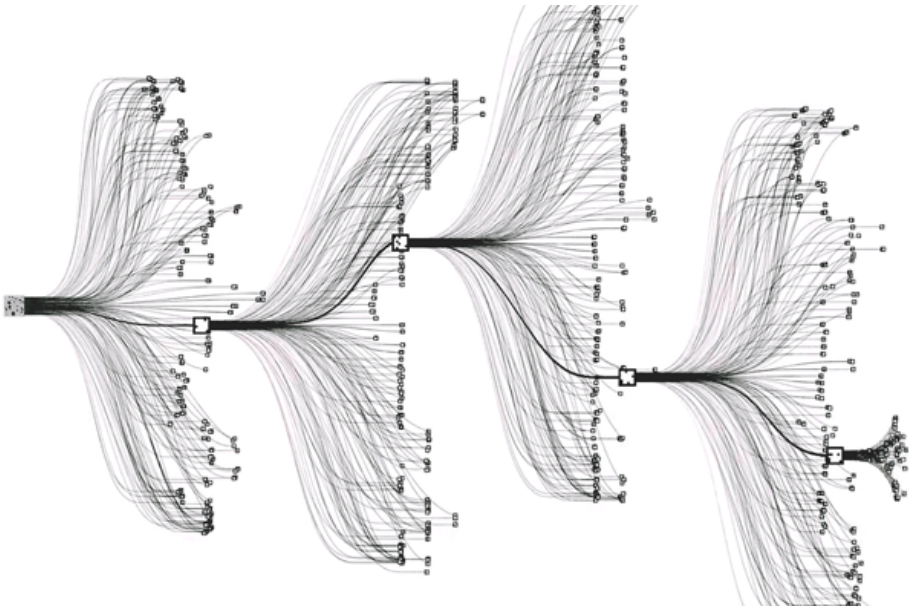
Machine learning

State-based models



White to move

State-based models



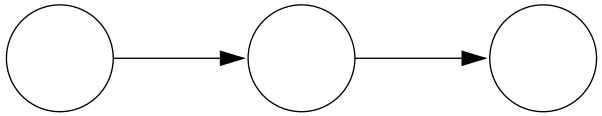
Applications:

- Games: Chess, Go, Pac-Man, Starcraft, etc.
- Robotics: motion planning
- Natural language generation: machine translation, image captioning

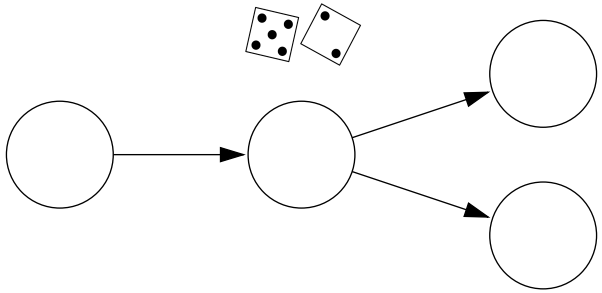
- Reflex-based models are too simple for tasks that require more forethought (e.g., in playing chess or planning a big trip). State-based models overcome this limitation.
- The key idea is, at a high-level, to model the **state** of a world and transitions between states which are triggered by actions. Concretely, one can think of states as nodes in a graph and transitions as edges. This reduction is useful because we understand graphs well and have a lot of efficient algorithms for operating on graphs.

State-based models

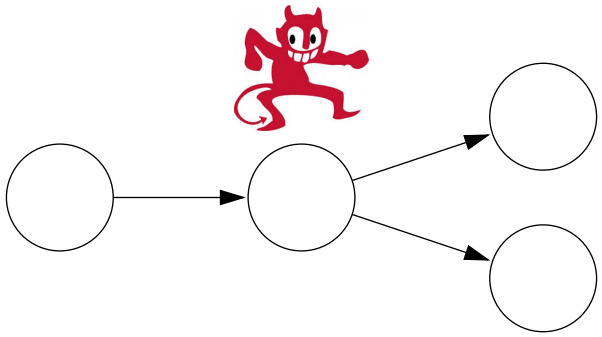
Search problems: you control everything



Markov decision processes: against nature (e.g., Blackjack)

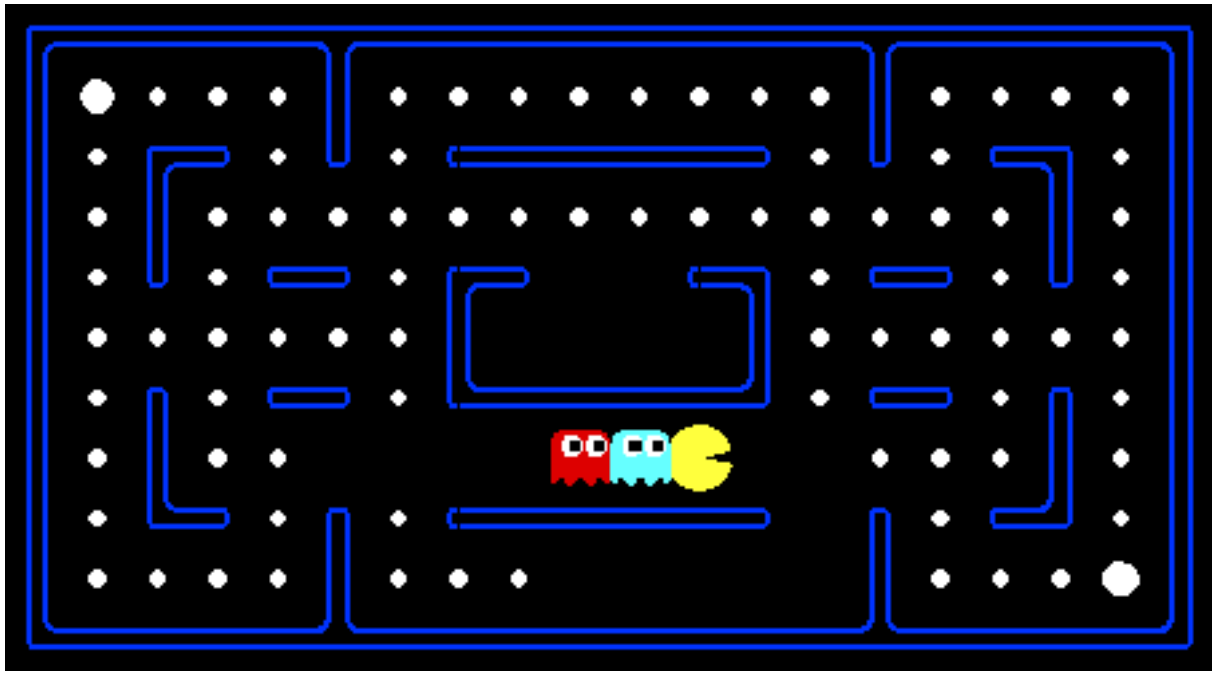


Adversarial games: against opponent (e.g., chess)



- Search problems are adequate models when you are operating in environment that has no uncertainty. However, in many realistic settings, there are other forces at play.
- **Markov decision processes** handle tasks with an element of chance (e.g., Blackjack), where the distribution of randomness is known (reinforcement learning can be employed if it is not).
- **Adversarial games**, as the name suggests, handle tasks where there is an opponent who is working against you (e.g., chess).

Pac-Man



[demo]

Question

What kind of model is appropriate for playing Pac-Man against ghosts that move into each valid adjacent square with equal probability?

search problem

Markov decision process

adversarial game

Course plan

Search problems

Markov decision processes

Adversarial games

Constraint satisfaction problems

Bayesian networks

Reflex **States** **Variables**

"Low-level intelligence" "High-level intelligence"

Machine learning

Sudoku

5	3			7				
6			1	9	5			
	9	8					6	
8				6				3
4			8		3			1
7				2				6
	6					2	8	
			4	1	9			5
				8			7	9

5	3	4	6	7	8	9	1	2
6	7	2	1	9	5	3	4	8
1	9	8	3	4	2	5	6	7
8	5	9	7	6	1	4	2	3
4	2	6	8	5	3	7	9	1
7	1	3	9	2	4	8	5	6
9	6	1	5	3	7	2	8	4
2	8	7	4	1	9	6	3	5
3	4	5	2	8	6	1	7	9

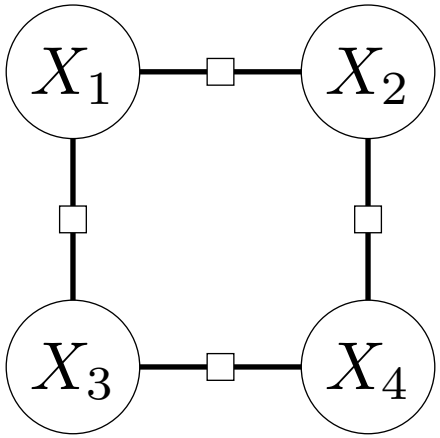
Goal: put digits in blank squares so each row, column, and 3x3 sub-block has digits 1–9

Note: order of filling squares doesn't matter in the evaluation criteria!

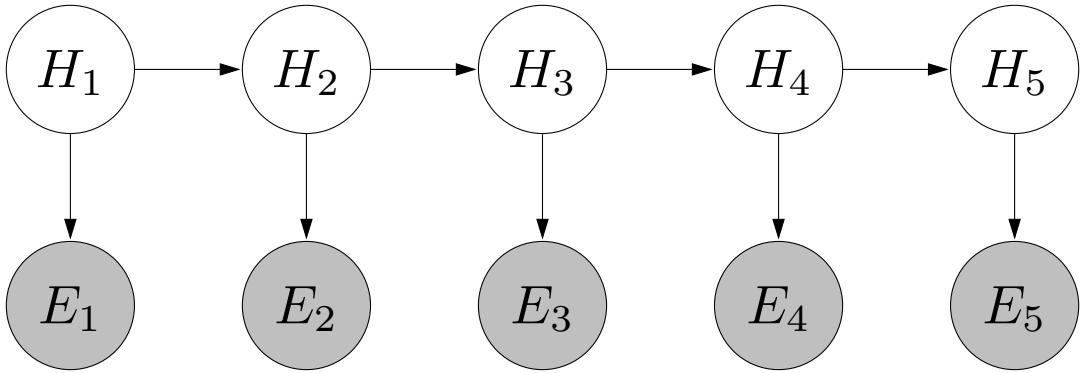
- In state-based models, solutions are procedural: they specify step by step instructions on how to go from A to B. In many applications, the order in which things are done isn't important.

Variable-based models

Constraint satisfaction problems: hard constraints (e.g., Sudoku, scheduling)



Bayesian networks: soft dependencies (e.g., tracking cars from sensors)



- **Constraint satisfaction problems** are variable-based models where we only have hard constraints. For example, in scheduling, we can't have two people in the same place at the same time.
- **Bayesian networks** are variable-based models where variables are random variables which are dependent on each other. For example, the true location of an airplane H_t and its radar reading E_t are related, as are the location H_t and the location at the last time step H_{t-1}. The exact dependency structure is given by the graph structure and formally defines a joint probability distribution over all the variables. This topic is studied thoroughly in probabilistic graphical models (CS228).

Course plan

Search problems

Markov decision processes

Adversarial games

Constraint satisfaction problems

Bayesian networks

Reflex **States** **Variables** **Logic**

"Low-level intelligence" "High-level intelligence"

Machine learning

Logic

- Dominated AI from 1960s-1980s, still useful in programming systems
- Powerful representation of knowledge and reasoning
- Brittle if done naively
- Open question: how to combine with machine learning?

- Our last stop on the tour is **logic**. Even more so than variable-based models, logic provides a compact language for modeling, which gives us more expressivity.
- It is interesting that historically, logic was one of the first things that AI researchers started with in the 1950s. While logical approaches were in a way quite sophisticated, they did not work well on complex real-world tasks with noise and uncertainty. On the other hand, methods based on probability and machine learning naturally handle noise and uncertainty, which is why they presently dominate the AI landscape. However, they have yet to be applied successfully to tasks that require really sophisticated reasoning.
- In this course, we will appreciate the two as not contradictory, but simply tackling different aspects of AI — in fact, in our schema, logic is a class of models which can be supported by machine learning. An active area of research is to combine the richness of logic with the robustness and agility of machine learning.

Motivation: virtual assistant

Tell information

Ask questions

Use natural language!

[demo]

Need to:

- Digest **heterogenous** information
- Reason **deeply** with that information

- One motivation for logic is a virtual assistant. At an abstract level, one fundamental thing a good personal assistant should be able to do is to take in information from people and be able to answer questions that require drawing inferences from these facts.
- In some sense, telling the system information is like machine learning, but it feels like a very different form of learning than seeing 10M images and their labels or 10M sentences and their translations. The type of information we get here is both more heterogenous, more abstract, and the expectation is that we process it more deeply (we don't want to have to tell our personal assistant 100 times that we prefer morning meetings).
- And how do we interact with our personal assistants? Let's use natural language, the very tool that was built for communication!

Course plan

Search problems

Markov decision processes

Adversarial games

Constraint satisfaction problems

Bayesian networks

Reflex **States** **Variables** **Logic**

"Low-level intelligence" "High-level intelligence"

Machine learning

Roadmap

What is AI?

Course overview

Course logistics

Optimization

Course objectives

Before you take the class, you should know...

- Programming (CS 106A, CS 106B, CS 107)
- Discrete math, mathematical rigor (CS 103)
- Probability (CS 109)

At the end of this course, you should...

- Be able to tackle real-world tasks with the appropriate techniques
- Be more proficient at math and programming

Exam

- Goal: test your ability to use knowledge to solve new problems, not know facts
- All written problems (look at past exam problems for style)
- Closed book except one page of notes
- Covers all material up to and including preceding week

Project

- Goal: choose any task you care about and apply techniques from class
- Work in groups of up to 3; find a group early, your responsibility to be in a good group
- Milestones: proposal, progress report, poster session, final report
- Task is completely open, but must follow well-defined steps: task definition, implement baselines/oracles, evaluate on dataset, literature review, error analysis (read website)
- Help: assigned a CA mentor, come to any office hours

THE HONOR CODE

- Do collaborate and discuss together, but write up and code independently.
- Do not look at anyone else's writeup or code.
- Do not show anyone else your writeup or code or post it online (e.g., GitHub).
- When debugging, only look at input-output behavior.
- We will run MOSS periodically to detect plagarism.

Roadmap

What is AI?

Course overview

Course logistics

Optimization

Optimization

Discrete optimization: a discrete object

$$\min_{p \in \text{Paths}} \text{Distance}(p)$$

Algorithmic tool: dynamic programming

Continuous optimization: a vector of real numbers

$$\min_{\mathbf{w} \in \mathbb{R}^d} \text{TrainingError}(\mathbf{w})$$

Algorithmic tool: gradient descent

- We are now done with the high-level motivation for the class. Let us now dive into some technical details. Let us focus on the inference and the learning aspect of the **modeling-inference-learning** paradigm.
- We will approach inference and learning from an **optimization** perspective, which allows us to decouple the mathematical specification of **what** we want to compute from the algorithms for **how** to compute it.
- In total generality, optimization problems ask that you find the x that lives in a constraint set C that makes the function $F(x)$ as small as possible.
- There are two types of optimization problems we'll consider: discrete optimization problems (mostly for inference) and continuous optimization problems (mostly for learning). Both are backed by a rich research field and are interesting topics in their own right. For this course, we will use the most basic tools from these topics: **dynamic programming** and **gradient descent**.
- Let us do two practice problems to illustrate each tool. For now, we are assuming that the model (optimization problem) is given and only focus on **algorithms**.

Problem: computing edit distance

Input: two strings, s and t

Output: minimum number of character insertions, deletions, and substitutions it takes to change s into t

Examples:

"cat", "cat"	$\Rightarrow$	0
"cat", "dog"	$\Rightarrow$	3
"cat", "at"	$\Rightarrow$	1
"cat", "cats"	$\Rightarrow$	1
"a cat!", "the cats!"	$\Rightarrow$	4

[live solution]

- Let's consider the formal task of computing the edit distance (or more precisely the Levenshtein distance) between two strings. These measures of dissimilarity have applications in spelling correction, computational biology (applied to DNA sequences).
- As a first step, you should think to break down the problem into subproblems. Observation 1: inserting into s is equivalent to deleting a letter from t (ensures subproblems get smaller). Observation 2: perform edits at the end of strings (might as well start there).
- Consider the last letter of s and t. If these are the same, then we don't need to edit these letters, and we can proceed to the second-to-last letters. If they are different, then we have three choices. (i) We can substitute the last letter of s with the last letter of t. (ii) We can delete the last letter of s. (iii) We can insert the last letter of t at the end of s.
- In each of those cases, we can reduce the problem into a smaller problem, but which one? We simply try all of them and take the one that yields the minimum cost!
- We can express this more formally with a mathematical recurrence. These types of recurrences will show up throughout the course, so it's a good idea to be comfortable with them. Before writing down the actual recurrence, the first step is to express the quantity that we wish to compute. In this case: let $d(m, n)$ be the edit distance between the first m letters of s and the first n letters of t. Then we have

$$d(m,n) = \begin{cases} m & \text{if } n = 0 \\ n & \text{if } m = 0 \\ d(m-1, n-1) & \text{if } s\_m = t\_n \\ 1 + \min\{d(m-1,n-1), d(m-1,n), d(m,n-1)\} & \text{otherwise.} \end{cases}$$

- Once you have the recurrence, you can code it up. The straightforward implementation will take exponential time, but you can **memoize** the results to make it $O(n^2)$ time. The end result is the dynamic programming solution: recurrence + memoization.

Problem: finding the least squares line

Input: set of pairs $\{(x_1, y_1), \ldots, (x_n, y_n)\}$

Output: $w \in \mathbb{R}$ that minimizes the squared error

$$F(w) = \sum_{i=1}^{n} (x_i w - y_i)^2$$

Examples:

$$\begin{aligned} \{(2,4)\} &\Rightarrow 2 \\ \{(2,4),(4,2)\} &\Rightarrow ? \end{aligned}$$

[live solution]

[linear regression,gradient descent]

- The formal task is this: given a set of n two-dimensional points (x_i, y_i) which defines $F(w)$, compute the w that minimizes $F(w)$.
- **Linear regression** is an important problem in machine learning, which we will come to later. Here's a motivation for the problem: suppose you're trying to understand how your exam score (y) depends on the number of hours you study (x). Let's posit a linear relationship $y = wx$ (not exactly true in practice, but maybe good enough). Now we get a set of training examples, each of which is a (x_i, y_i) pair. The goal is to find the slope w that best fits the data.
- Back to algorithms for this formal task. We would like an algorithm for optimizing general types of $F(w)$. So let's **abstract away from the details**. Start at a guess of w (say $w = 0$), and then iteratively update w based on the derivative (gradient if w is a vector) of $F(w)$. The algorithm we will use is called **gradient descent**.
- If the derivative $F'(w) < 0$, then increase w; if $F'(w) > 0$, decrease w; otherwise, keep w still. This motivates the following update rule, which we perform over and over again: $w \leftarrow w - \eta F'(w)$, where $\eta > 0$ is a **step size** that controls how aggressively we change w.
- If η is too big, then w might bounce around and not converge. If η is too small, then w might not move very far to the optimum. Choosing the right value of η can be rather tricky. Theory can give rough guidance, but this is outside the scope of this class. Empirically, we will just try a few values and see which one works best. This will help us develop some intuition in the process.
- Now to specialize to our function, we just need to compute the derivative, which is an elementary calculus exercise: $F'(w) = \sum_{i=1}^{n} 2(x_i w - y_i)x_i$.

Question

What was the most surprising thing you learned today?

Summary

- AI has high societal impact, our responsibility to steer it positively
- Modeling [reflex, states, variables, logic] + inference + learning
- Section this Friday: review of foundations
- Homework [foundations]: due next Tuesday 11pm
- Course will be fast-paced and exciting!