Lecture 0: Introduction to AI

Required Resources

https://aima.cs.berkeley.edu/

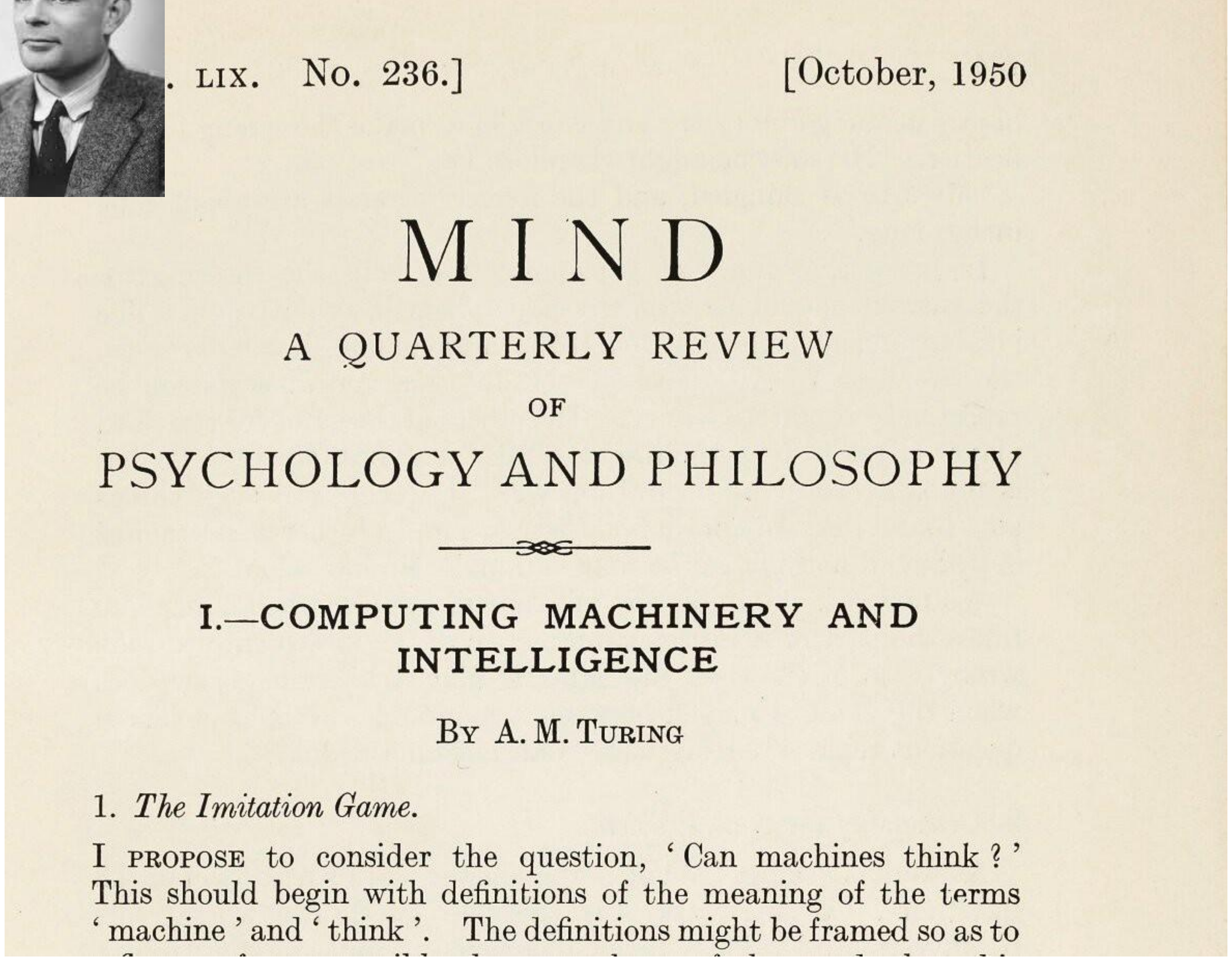
Roadmap

AI history

Ethics and responsibility

Course content

objective specification



. LIX. No. 236.] [October, 1950

MIND

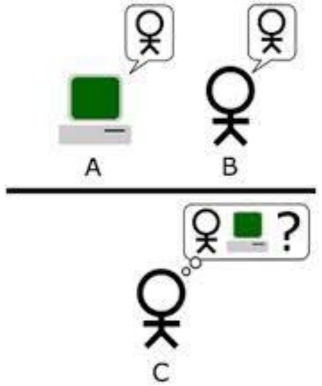
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

1. *The Imitation Game.*

I propose to consider the question, ‘Can machines think ?’ This should begin with definitions of the meaning of the terms ‘machine’ and ‘think’. The definitions might be framed so as to



Many people think that a very abstract activity, like the playing of chess, would be best. It can also be maintained that it is best to provide the machine with the best ***sense organs*** *that money can buy, and then teach it to understand and speak English. This process could follow the normal* ***teaching of a child****. Things would be pointed out and named, etc. Again I do not know what the right answer is, but I think both approaches should be tried.*

1956

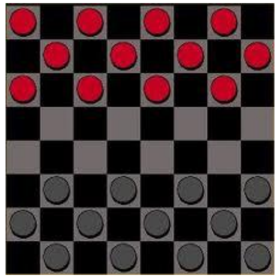
Birth of AI

1956: John McCarthy organized workshop at Dartmouth College

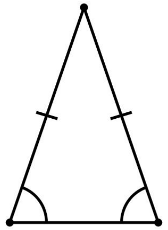
Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it.

general principles

Birth of AI, early successes



Checkers (1952): Samuel's program learned weights and played at strong amateur level



Problem solving (1955): Newell & Simon's Logic Theorist: prove theorems in Principia Mathematica using search + heuristics; later, General Problem Solver (GPS)

Overwhelming optimism...

Machines will be capable, within twenty years, of doing any work a man can do.

—Herbert Simon

Within 10 years the problems of artificial intelligence willbe substantially solved.

—Marvin Minsky

I visualize a time when we will be to robots what dogs are to humans, and I'm rooting for the machines.

—Claude Shannon

...underwhelming results

Example: machine translation

The spirit is willing but the flesh is weak.

↓

(Russian)

↓

The vodka is good but the meat is rotten.

1966: ALPAC report cut off government funding for MT, first AI winter

Implications of early era

Problems:

- Limited computation: search space grew exponentially, outpacing hardware
- Limited information: complexity of AI problems (number of words, objects, concepts in the world)

Useful contributions (John McCarthy):

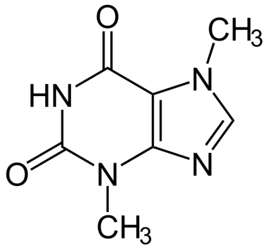
- Lisp
- Garbage collection
- Time-sharing

Knowledge-based systems (70-80s)

Expert systems: elicit specific domain knowledge from experts in form of rules:

if [premises] then [conclusion]

Knowledge-based systems (70-80s)



DENDRAL: infer molecular structure from mass spectrometry

MYCIN: diagnose blood infections, recommend antibiotics

XCON: convert customer orders into parts specification

Knowledge-based systems

Wins:

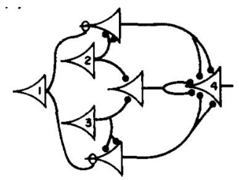
- Knowledge helped both the information and computation gap
- First **real application** that impacted industry Problems:
- Deterministic rules couldn't handle the **uncertainty** of the real world
- Rules quickly became too **complex** to create and maintain

A number of people have suggested to me that large programs like the SHRDLU program for understanding natural language represent a kind of dead end in AI programming. Complex interactions between its components give the program much of its power, but at the same time they present a formidable obstacle to understanding and extending it. In order to grasp any part, it is necessary to understand how it fits with other parts, presents a dense mass, with no easy footholds. Even having written the program, I find it near the limit of what I can keep in mind at once. — Terry Winograd

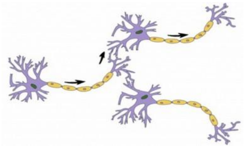
1987: Collapse of Lisp machines and second AI winter

1943

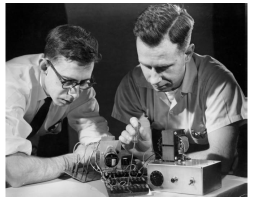
Artificial neural networks



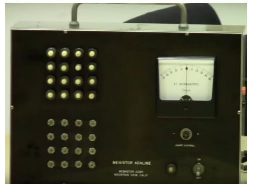
1943: artificial neural networks, relate neural circuitry and mathematical logic (Mc- Culloch/Pitts)



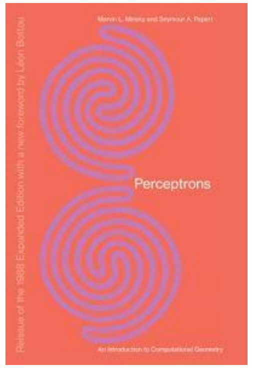
1949: ”cells that fire together wire together” learning rule (Hebb)



1958: Perceptron algorithm for linear classifiers (Rosenblatt)

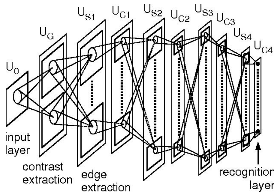


1959: ADALINE device for linear regression (Widrow/Hoff)

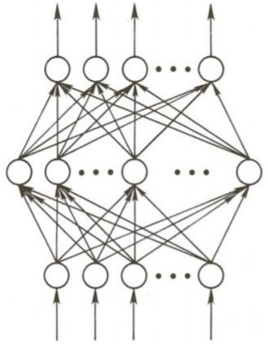


1969: Perceptrons book showed that linear models could not solve XOR, killed neural nets research (Minsky/Papert)

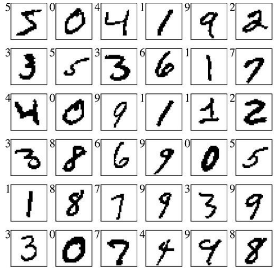
Revival of connectionism



1980: Neocognitron, a.k.a. convolutional neural networks for images (Fukushima)

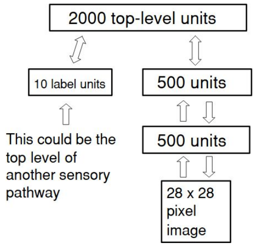


1986: popularization of backpropagation for training multi-layer networks (Rumel- hardt, Hinton, Williams)



1989: applied convolutional neural networks to recognizing handwritten digits for USPS (LeCun)

Deep learning

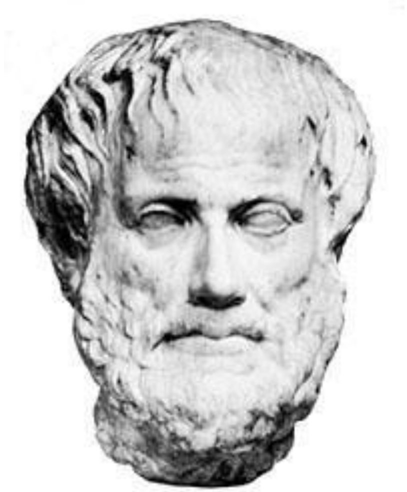


2006: unsupervised layerwise pre-training of deep networks (Hinton et al.)

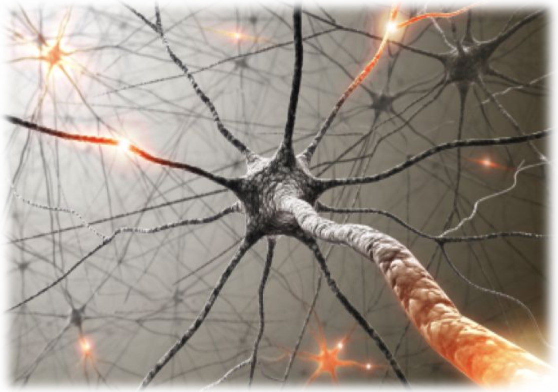
2012: AlexNet obtains huge gains in object recognition; transformed computer vision community overnight

2016: AlphaGo uses deep reinforcement learning, defeat world champion Lee Sedol in Go

Two intellectual traditions



symbolic AI

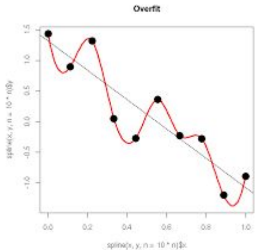


neural AI

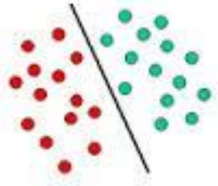
Food for thought: deep philosophical differences, but deeper connections (McCulloch/Pitts, AlphaGo)?

1801

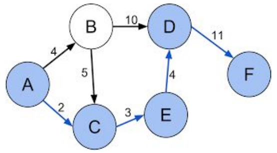
Early ideas from outside AI



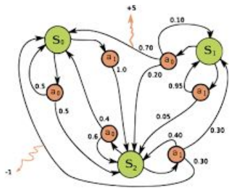
1801: linear regression (Gauss, Legendre)



1936: linear classification (Fisher)

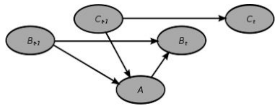


1956: Uniform cost search for shortest paths (Dijkstra)

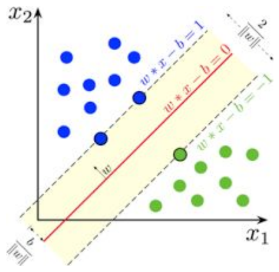


1957: Markov decision processes (Bellman)

Statistical machine learning

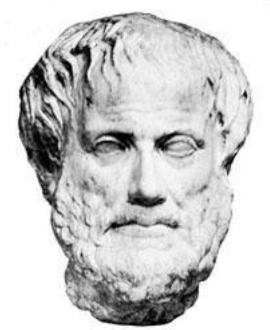


1985: Bayesian networks (Pearl)

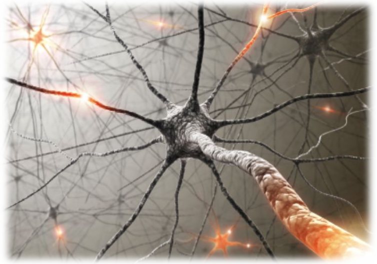


1995: Support vector machines (Cortes/Vapnik)

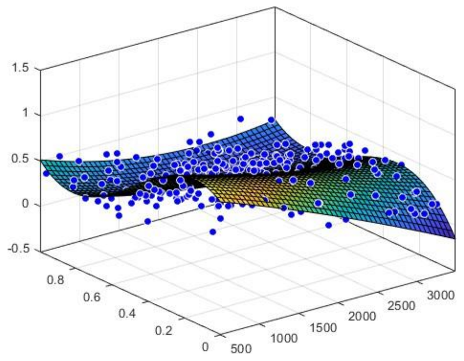
Three intellectual traditions



symbolic AI



neural AI



statistical AI

Further reading

Wikipedia article: https://en.wikipedia.org/wiki/History_of_artificial_intelligence

Encyclopedia of Philosophy article: https://plato.stanford.edu/entries/artificial-intelligence

Turing's Computing Machinery and Intelligence: https://www.csee.umbc.edu/courses/471/papers/turing.pdf

History and Philosophy of Neural Networks: https://research.gold.ac.uk/10846/1/Bishop-2014.pdf

Roadmap

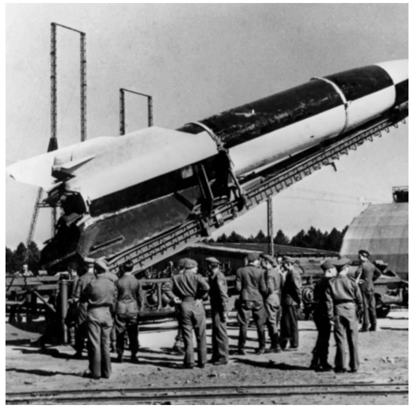
AI history

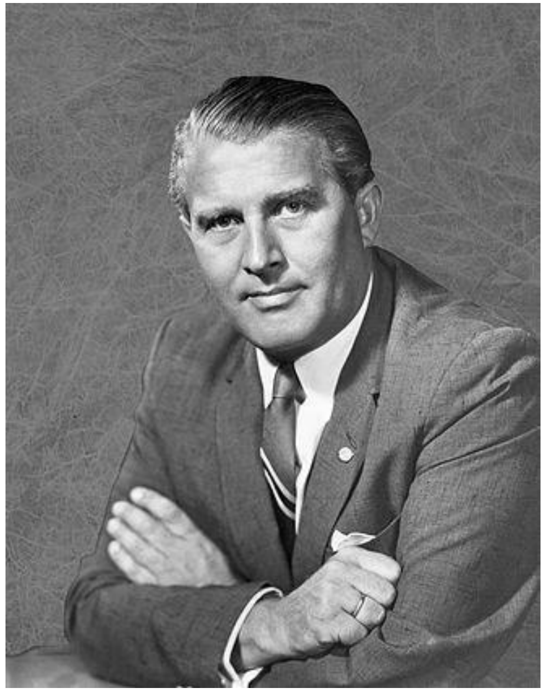
Ethics and responsibility

Course content

Why care about responsibility?

Isn’t technology value-neutral?





Wernher von Braun

”Once the rockets are up,
Who cares where they come down? That’s not my department,”

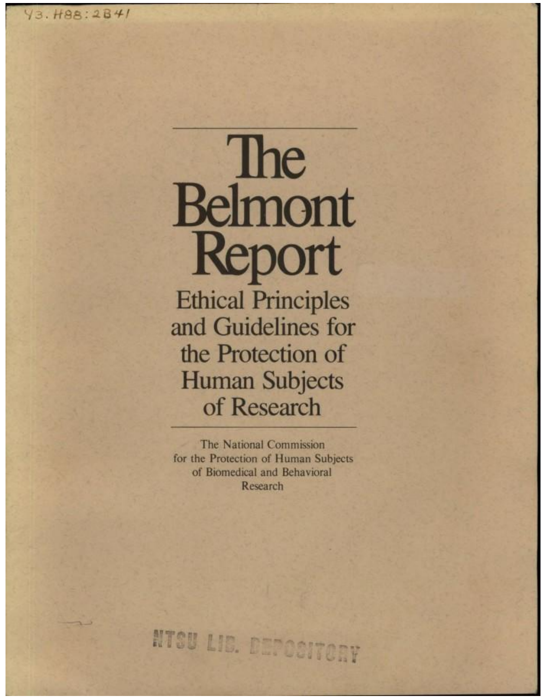
Says Wernher von Braun.

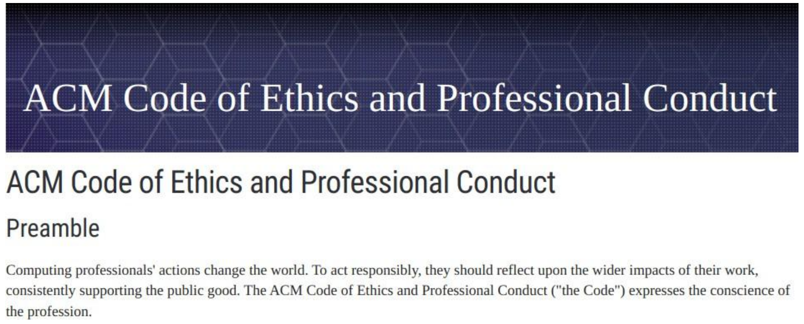
Lyrics: Tom Lehrer

Goal of responsibility

Goal: ensure AI is developed to benefit and not harm society

High-level principles: respect for persons, don't do harm




ACM Code of Ethics and Professional Conduct

ACM Code of Ethics and Professional Conduct

Preamble

Computing professionals' actions change the world. To act responsibly, they should reflect upon the wider impacts of their work, consistently supporting the public good. The ACM Code of Ethics and Professional Conduct ("the Code") expresses the conscience of the profession.


Microsoft AI principles

We put our responsible AI principles into practice through the Office of Responsible AI (ORA), the AI, Ethics, and Effects in Engineering and Research (Aether) Committee, and Responsible AI Strategy in Engineering (RAISE). The Aether Committee advises our leadership on the challenges and opportunities presented by AI innovations. ORA sets our rules and governance processes, working closely with teams across the company to enable the effort. RAISE is a team that enables the implementation of Microsoft responsible AI rules across engineering groups.

Key question: how to operationalize these principles?

Intent versus impact

← Impact →

← Intent →	positive	negative
good	**beneficial applications** healthcare, education, science	**accidents** robustness, fairness, security
bad	?	**misuse** spam, fraud, disinformation

Visual assistive technology

Disinformation

Others: spam, fraud, personal attacks

Dual-use technology

Definition: a dual use technology is one that can be used both to **benefit** and to **harm**.

Examples:

rockets
nuclear power
gene editing
social networks
AI

Levels of abstraction

deep learning

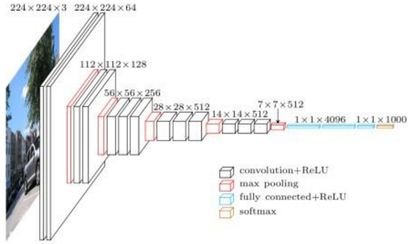


image generation

deepfakes

autonomous weapons

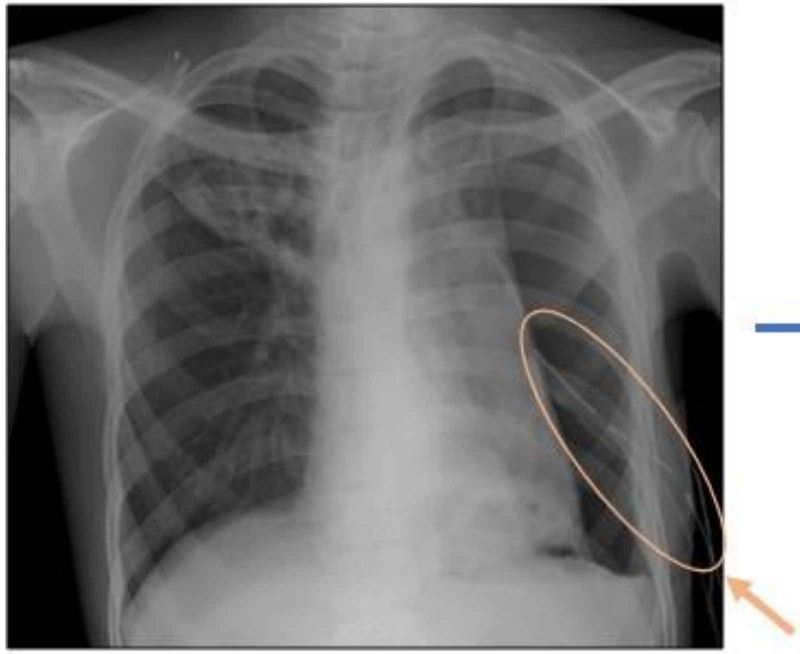
generality

specificity

[Oakden-Rayner, Dunnmon, Carneiro, R´e (2019)]

Robustness: spurious correlations

input x



Problem

Model uses treatment (chest drain) to diagnose the condition

Model is not accurate on patients who haven't been treated yet (!)

ML model

prediction — collapsed lung

true label $\mathbf{y}$ — collapsed lung

spurious correlation — chest drain = common treatment for collapsed lungs

Security

[Evtimov+ 2017]

[Sharif+ 2016]

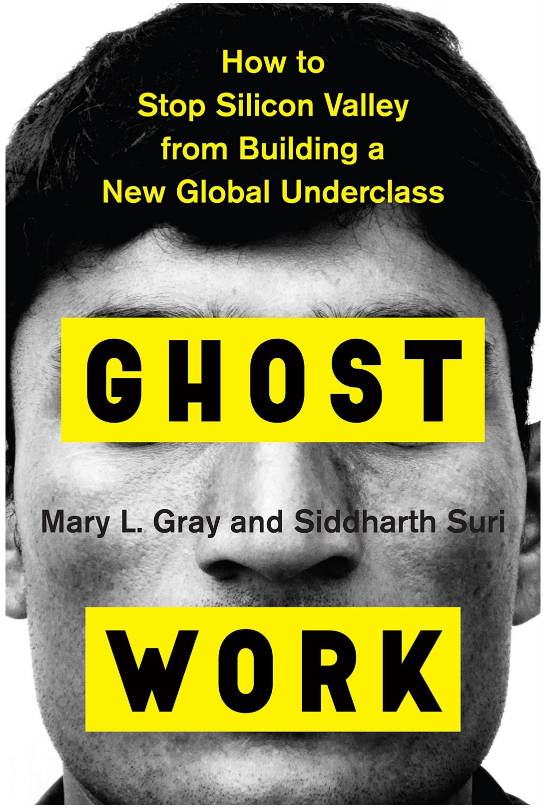
Adversaries at test time!

Data

- Web-scraped data can contain offensive content, historical biases

- Consent: Should a datum (e.g. a picture of my dog) whose owner or creator intended it for one use be allowed to be used in another application (e.g. scene classification) without permission?

Data



Data is produced by human labor

Transparency

Model Cards for Model Reporting

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru
{mmitchellai,simonewu,andrewzaldivar,parkerbarnes,lucyvasserman,benhutch,espitzer,tgebru}@google.com
deborah.raji@mail.utoronto.ca

Datasheets for Datasets

TIMNIT GEBRU, Black in AI
JAMIE MORGENSTERN, University of Washington
BRIANA VECCHIONE, Cornell University
JENNIFER WORTMAN VAUGHAN, Microsoft Research
HANNA WALLACH, Microsoft Research
HAL DAUMÉ III, Microsoft Research; University of Maryland
KATE CRAWFORD, Microsoft Research

Document potential issues

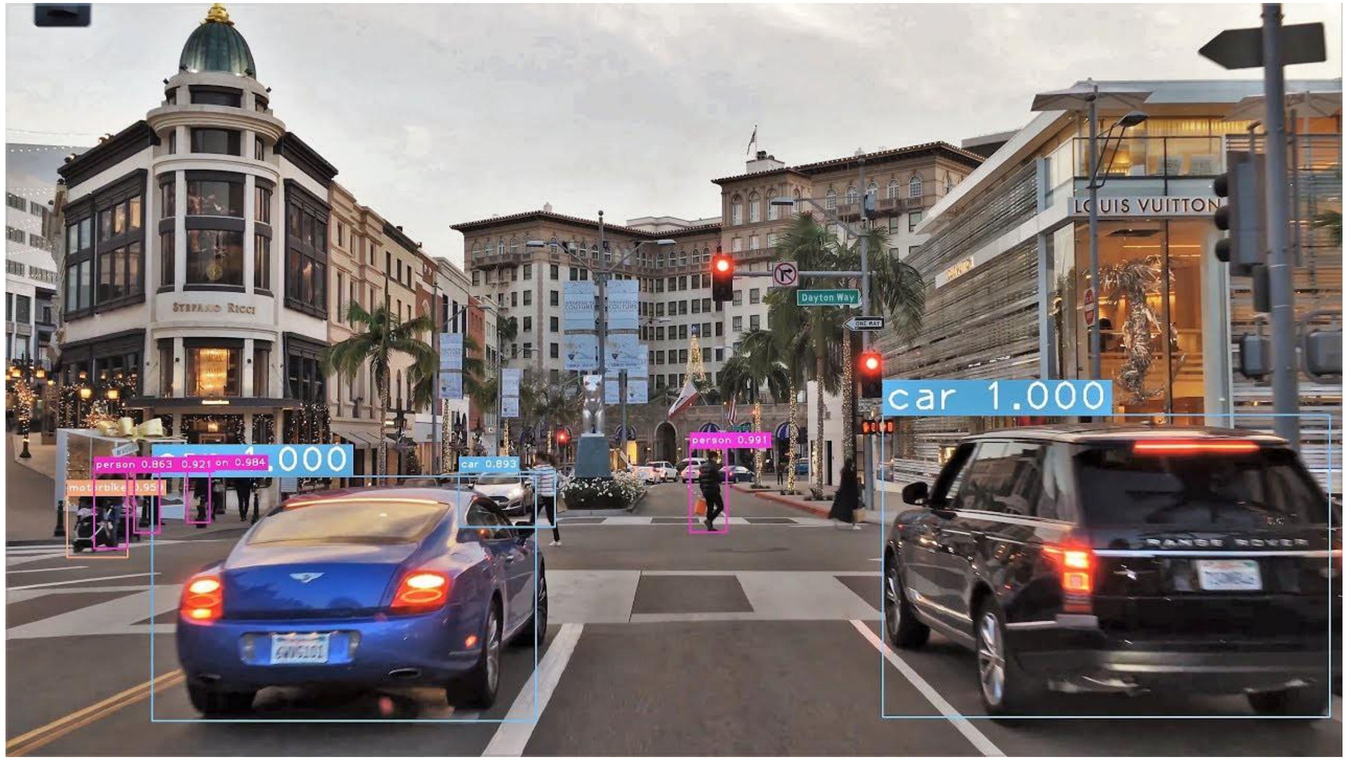
Roadmap

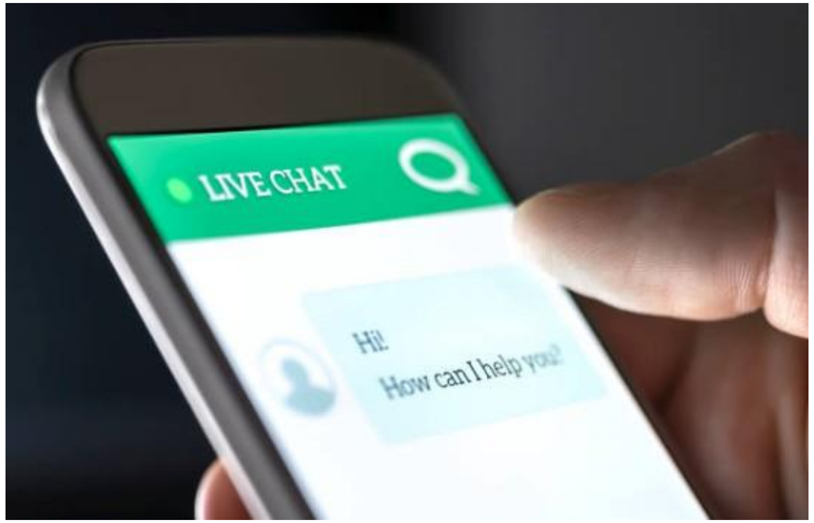
AI history

Ethics and responsibility

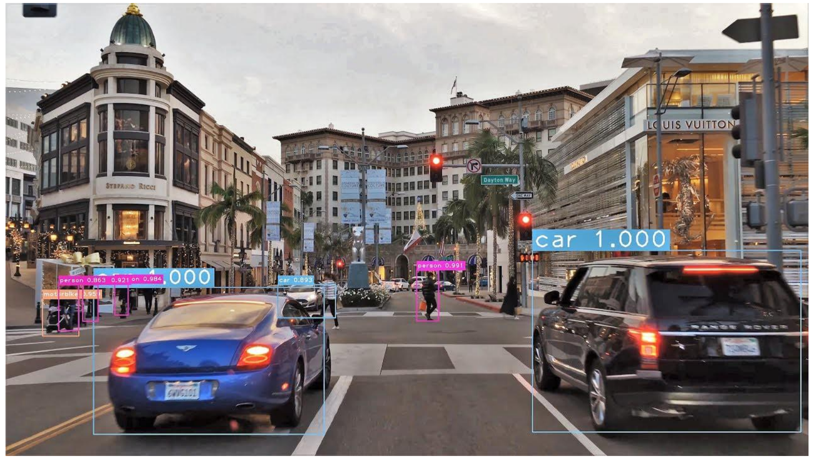
Course content

Complex real-world problems





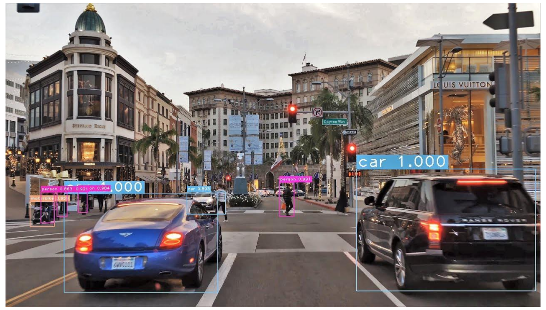
Bridging the gap



?

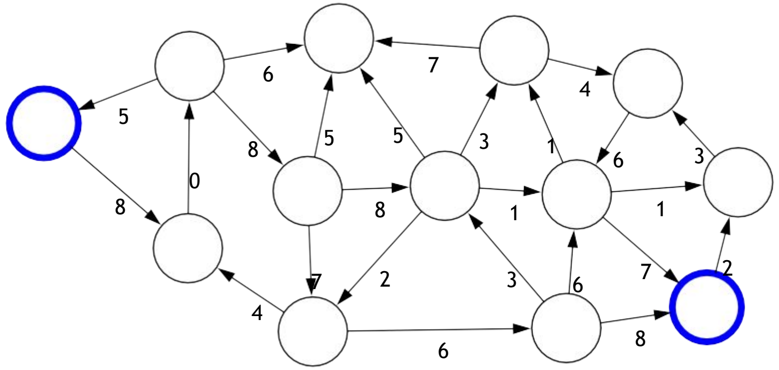
Paradigm: modeling

Real world

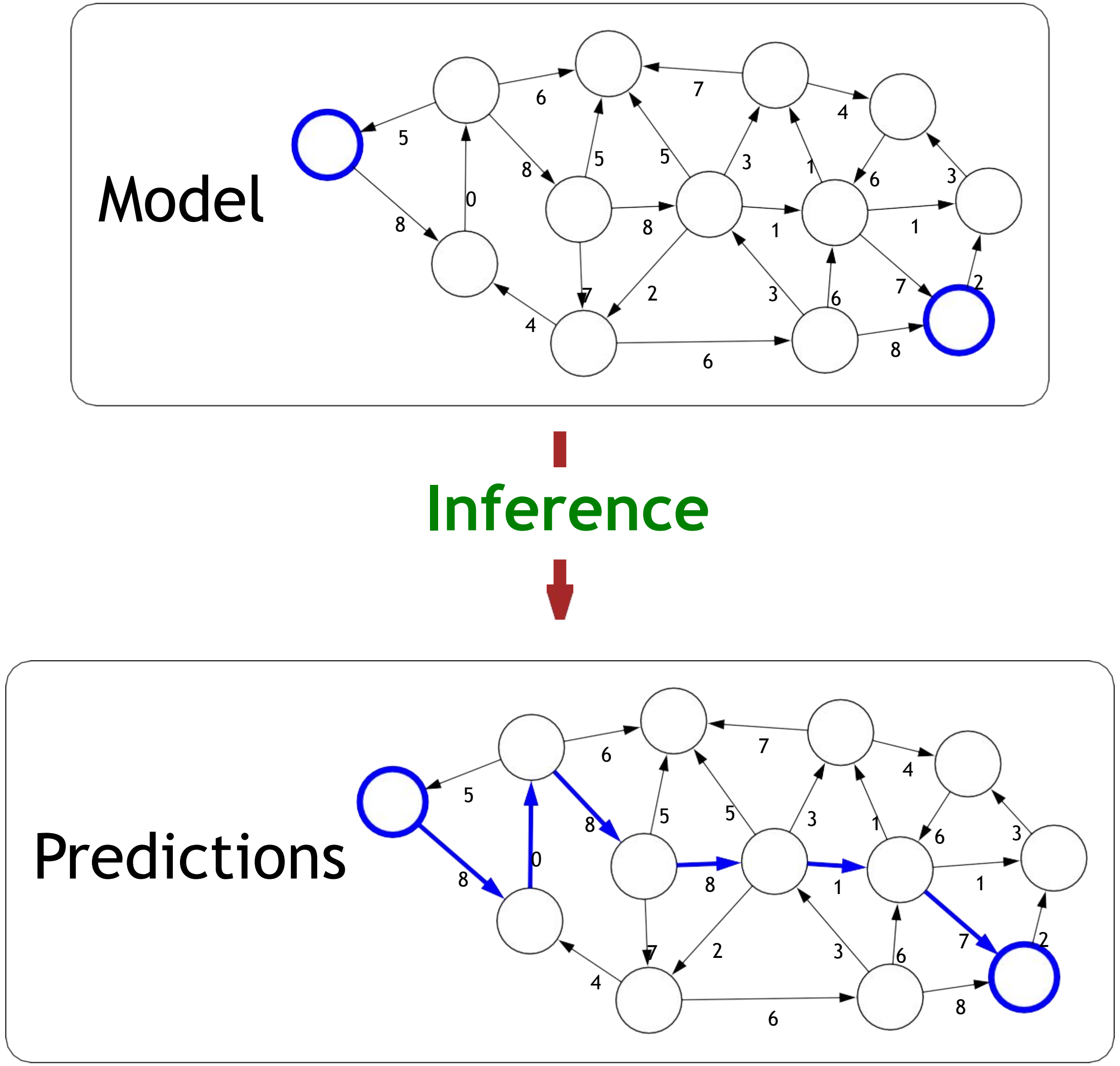


Modeling

Model

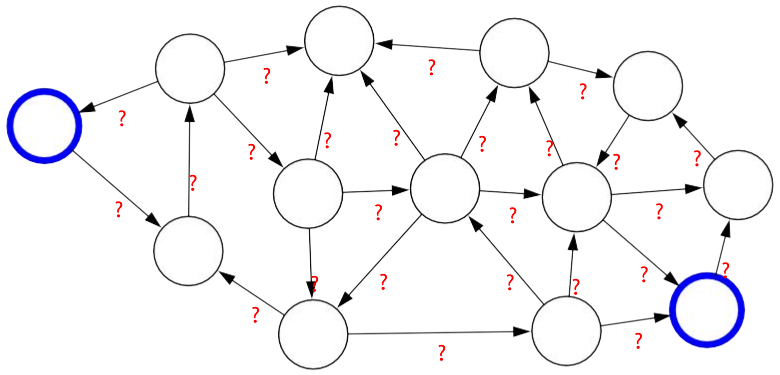


Paradigm: inference



Paradigm: learning

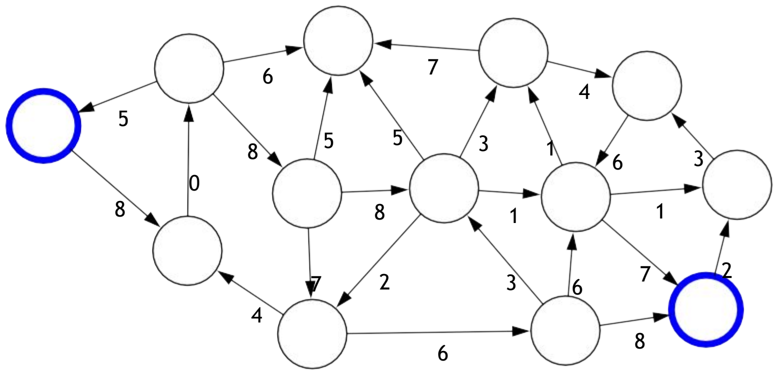
Model without parameters



+data

Learning

Model with parameters



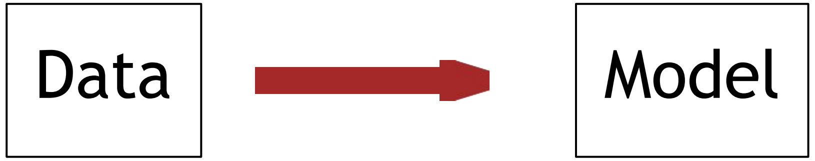
Paradigm

Modeling

Inference

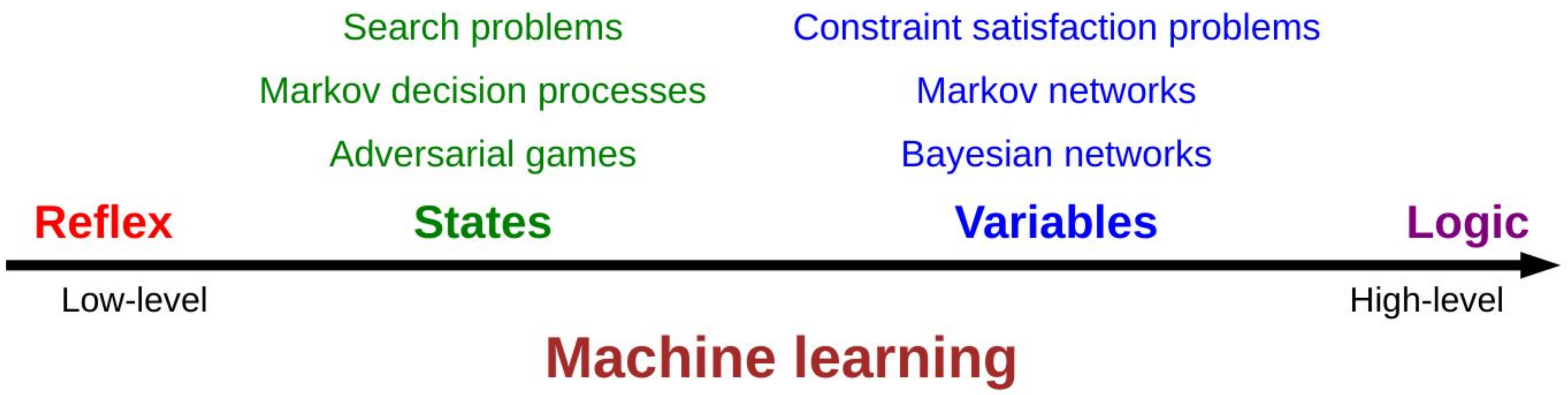
Learning

Machine learning



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

Course plan


Search problems
Markov decision processes
Adversarial games
Constraint satisfaction problems
Markov networks
Bayesian networks
Reflex
States
Variables
Logic
Low-level
High-level
Machine learning