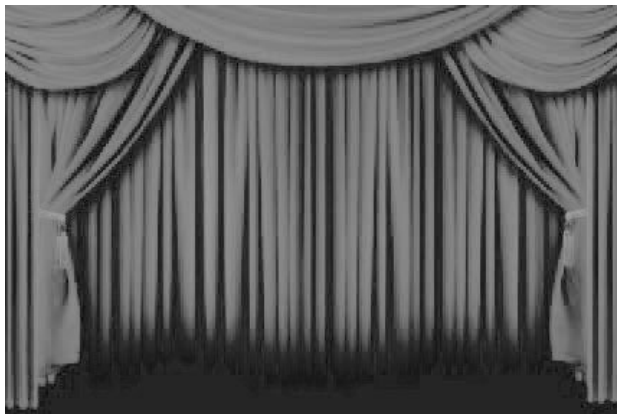
Conclusion



Roadmap

Summary of CS221

Next courses

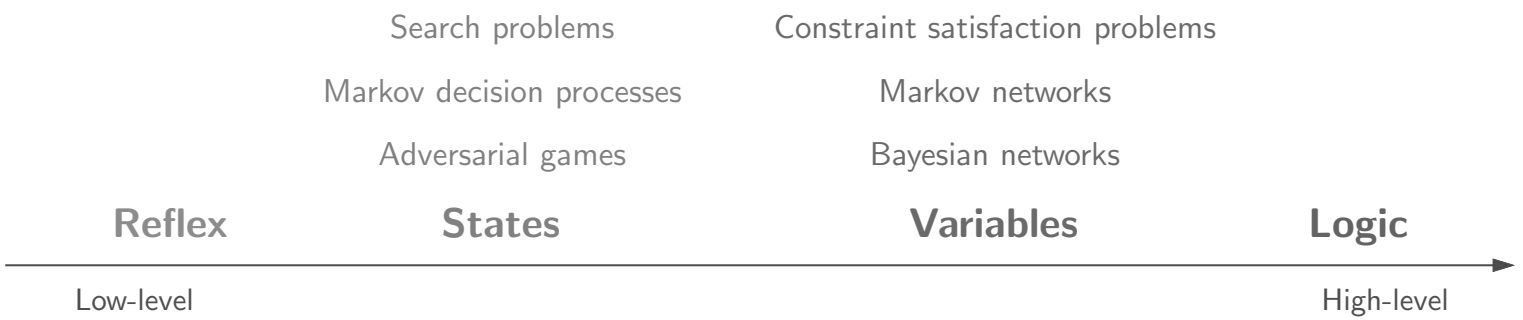
Food for thought

Paradigm

Modeling

Inference **Learning**

Course plan


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

Machine learning

Objective: loss minimization

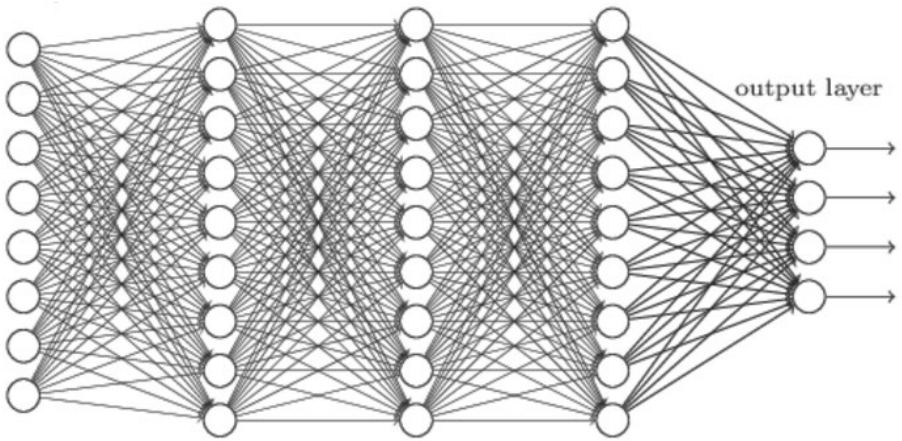
$$\min_{\mathbf{w}} \sum_{(x,y) \in \mathcal{D}_{\text{train}}} \text{Loss}(x, y, \mathbf{w})$$

Algorithm: stochastic gradient descent

$$\mathbf{w} \to \mathbf{w} - \eta_t \underbrace{\nabla \text{Loss}(x, y, \mathbf{w})}_{\text{prediction} - \text{target}}$$

Applies to wide range of models!

Reflex-based models

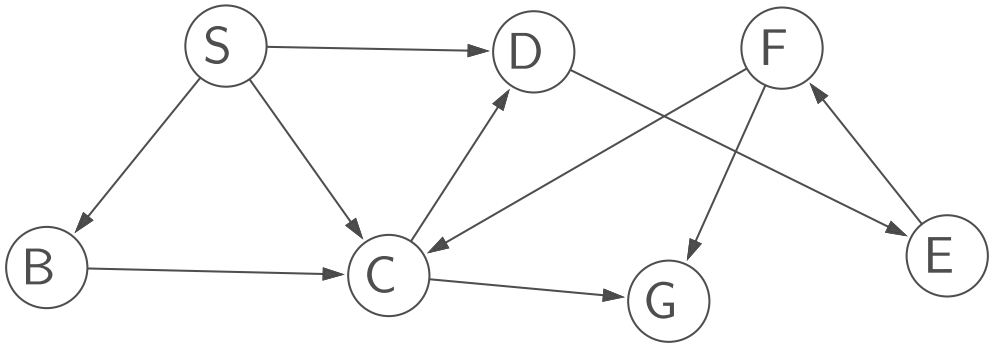


Models: linear models, neural networks, nearest neighbors

Inference: feedforward

Learning: SGD, alternating minimization

State-based models



Key idea: state

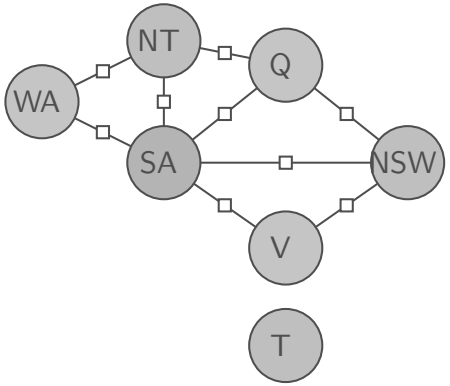
A **state** is a summary of all the past actions sufficient to choose future actions **optimally**.

Models: search problems, MDPs, games

Inference: UCS/A*, DP, value iteration, minimax

Learning: structured Perceptron, Q-learning, TD learning

Variable-based models



Key idea: factor graphs

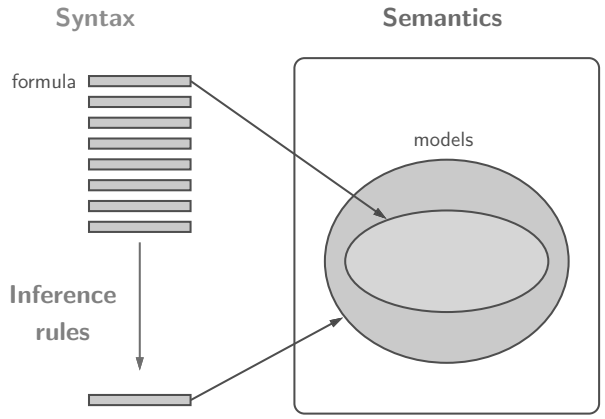
Graph structure captures conditional independence.

Models: CSPs, Markov networks, Bayesian networks

Inference: backtracking, forward-backward, beam search, Gibbs sampling

Learning: maximum likelihood (closed form, EM)

Logic-based models



Key idea: logic

Formulas enable more powerful models (infinite).

Models: propositional logic, first-order logic

Inference: model checking, modus ponens, resolution

Learning: ???

Tools

- CS221 provides a set of tools

- Start with the problem, and figure out what tool to use
- Keep it simple!

Roadmap

Summary of CS221

Next courses

Food for thought

Overview

List of AI courses:

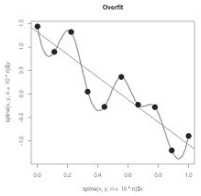
http://ai.stanford.edu/courses/

Types of courses:

- Methods: more advanced techniques, general-purpose
- Applications: real impact of AI, help you truly understand and appreciate methods
- Foundations: invest in building depth (for methods and applications); usually not in AI (math, hardware, linguistics/biology, etc.)

Methods

Machine learning



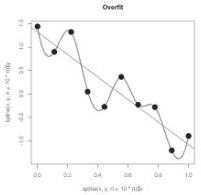
CS229: Machine Learning

- Standard, more mathematical derivations, continuous variables (e.g., kernel methods, PCA)

CS230: Deep Learning

- Applied, how to train deep neural networks (e.g., dropout, batch norm)

Machine learning



CS329D: Machine Learning Under Distribution Shifts

- Machine learning fails when train != test (e.g., adversarial examples, DRO)

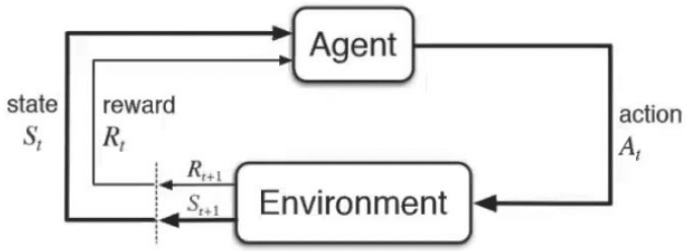
CS330: Deep Multi-Task and Meta Learning

- How to transfer across multiple tasks (e.g., few-shot learning, meta-RL)

CS224W: Machine Learning with Graphs

- Data points are graphs or are connected via a graph (e.g., graph neural networks)

Reinforcement learning



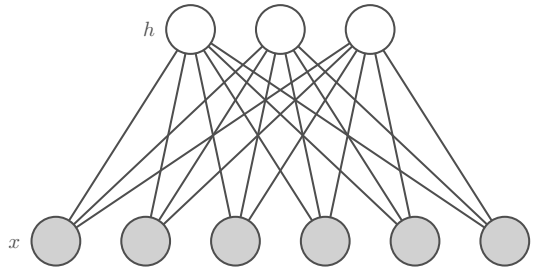
CS234: Reinforcement Learning

- More advanced techniques (e.g., policy search, bandits, batch RL)

CS238: Decision Making Under Uncertainty

- Model-based planning, applications to autonomous vehicles, aviation

Generative models



CS228: Probabilistic Graphical Models

- More advanced techniques (e.g., belief propagation, variational inference, MCMC, structure learning)

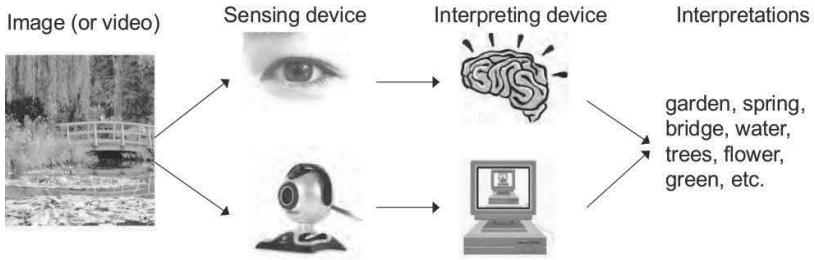
CS236: Deep Generative Models

- Generative models supercharged with deep learning (e.g., VAEs, GANs)

Applications

[figure credit: Fei-Fei Li]

Vision



CS231N: Convolutional Neural Networks for Visual Recognition

- ML-heavy (convnets, Transformers), detection, segmentation, generation

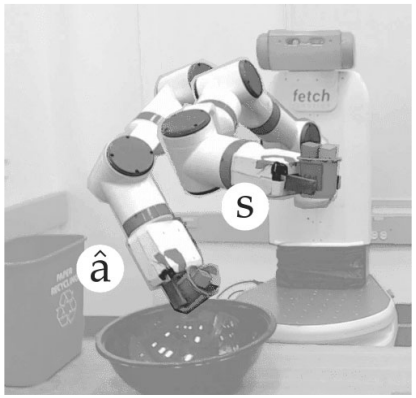
CS231A: From 3D Reconstruction to Recognition

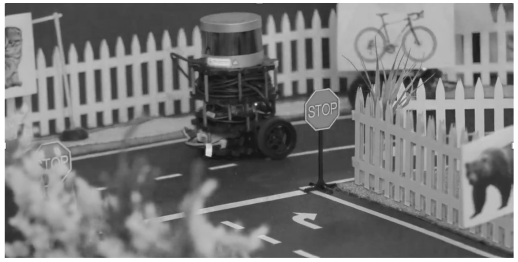
- More vision (e.g., cameras + geometry, shape reconstruction, depth estimation)

CS348I: Computer Graphics in the Era of AI

- Rendering, geometry, animation, computational photography

Robotics





CS237[AB]: Principles of Robotic Autonomy

- ML-heavy (RL, imitation learning), grasping, manipulation

CS223A: Introduction to Robotics

- Physical models for kinematics and control

Language

CS224N: Natural Language Processing with Deep Learning

- ML-heavy (RNNs, Transformers), parsing, translation, generation

CS224U: Natural Language Understanding

- Word representations, grounding, natural language inference, evaluation

CS224V: Conversational Virtual Assistants with Deep Learning

- Applications to semantic parsing, dialogue state tracking

CS224C: NLP for Computational Social Science

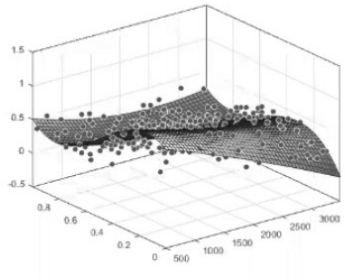
- Text analysis, applications to social science and sociolinguistics

CS324: Understanding and Developing Large Language Models

- Social/ethical/legal considerations, scaling laws, hands-on experience

Foundations

Optimization, statistics, theory



EE364[AB] / CS334[AB]: Convex Optimization

- Convex optimization problems, duality

STATS 200: Statistical Inference

- Statistical thinking, decision theory, hypothesis testing

STATS 214 / CS229M: Machine Learning Theory

- Why does it work? Uniform convergence, deep learning theory

Cognitive science and neuroscience

PSYCH204[AB] / CS428[AB]: Computation and Cognition: The Probabilistic Approach

- Human mind (software), using probabilistic programs to model human reasoning and learning [A], language [B]

PSYCH 242 / APPPHYS 293: Theoretical Neuroscience

- Human brain (hardware), neurally-plausible approximation of back propagation, spiking neural networks

Summary

Types of courses:

- Methods: more advanced techniques, general-purpose
- Applications: real impact of AI, help you truly appreciate methods
- Foundations: invest in building depth (for methods and applications); usually not in AI (math, hardware, linguistics/biology, etc.)

Tips:

- Invest in building depth, take classes outside CS
- Many resources (tutorials, blog posts, talks) online
- Download code, tinker — hands-on learning
- Talk to professors and other students

Roadmap

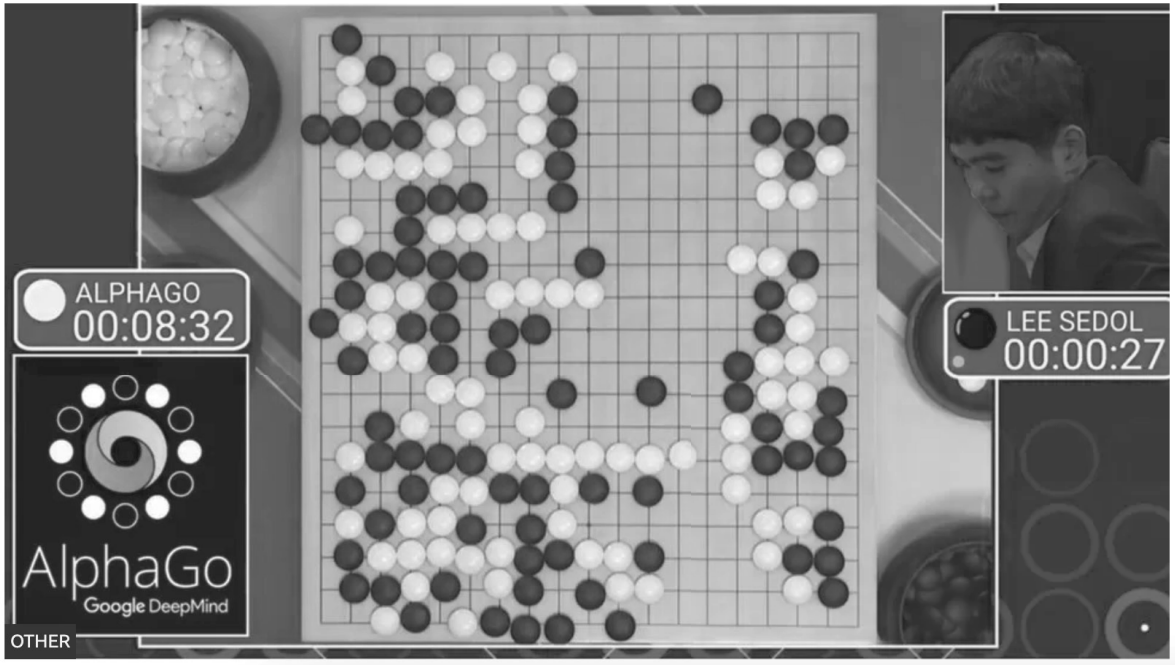
Summary of CS221

Next courses

Food for thought

Artificial intelligence: Google's AlphaGo beats Go master Lee Se-dol

12 March 2016



Large Language Models

"GPT-3: Just" a language model (175B parameters) trained on 570GB text

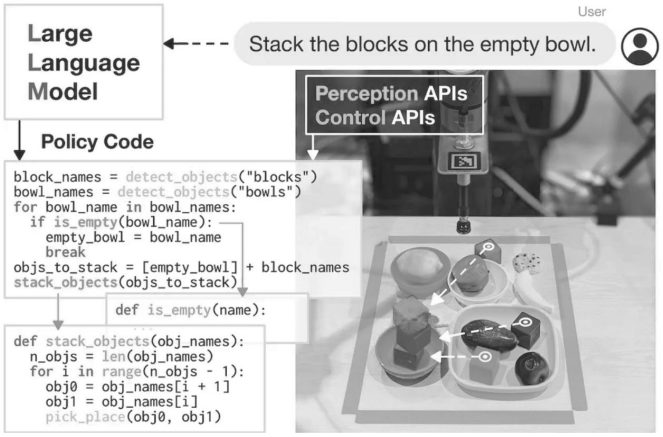
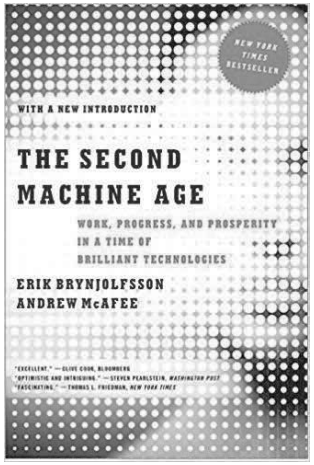


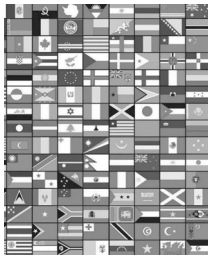
Image Source: https://code-as-policies.github.io/

Real-world applications

AI is everywhere: consumer services, advertising, transportation, manufacturing, etc.



AI being used to make decisions for: education, credit, employment, advertising, healthcare and policing

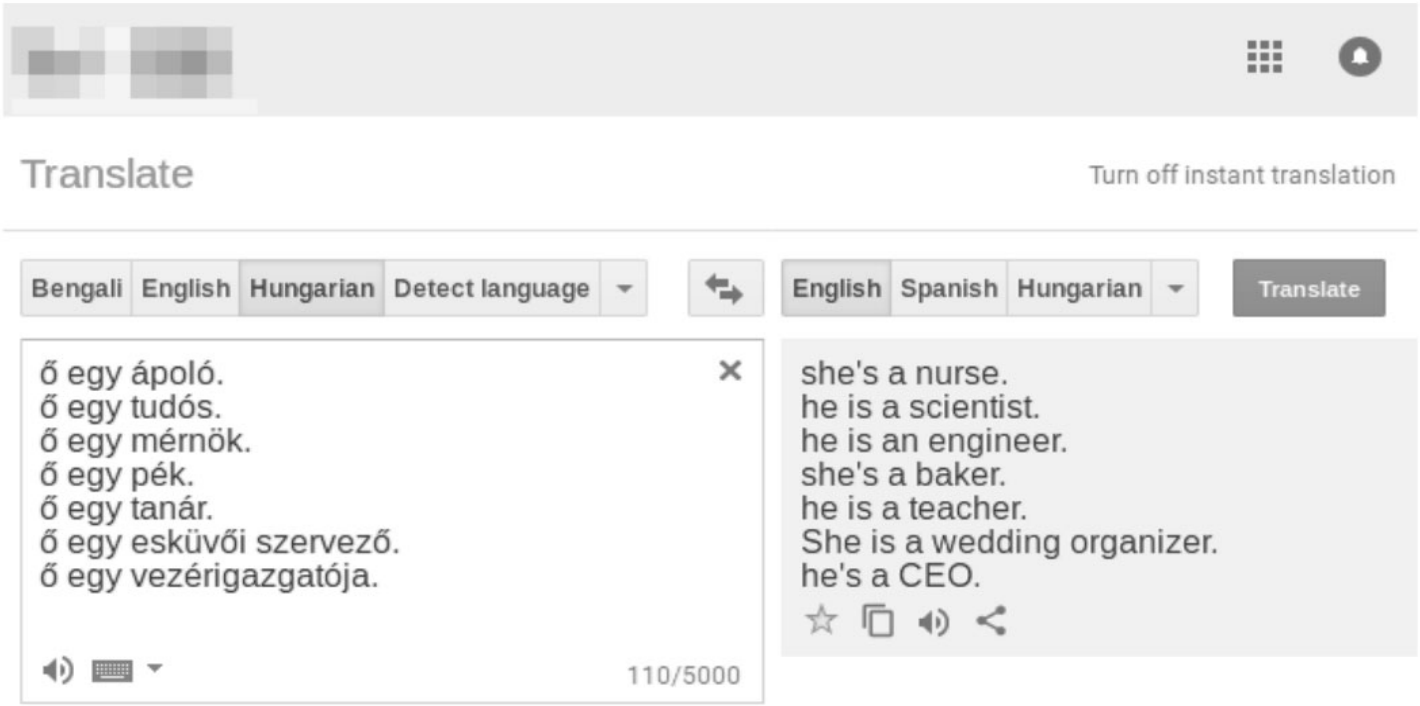


Machine translation

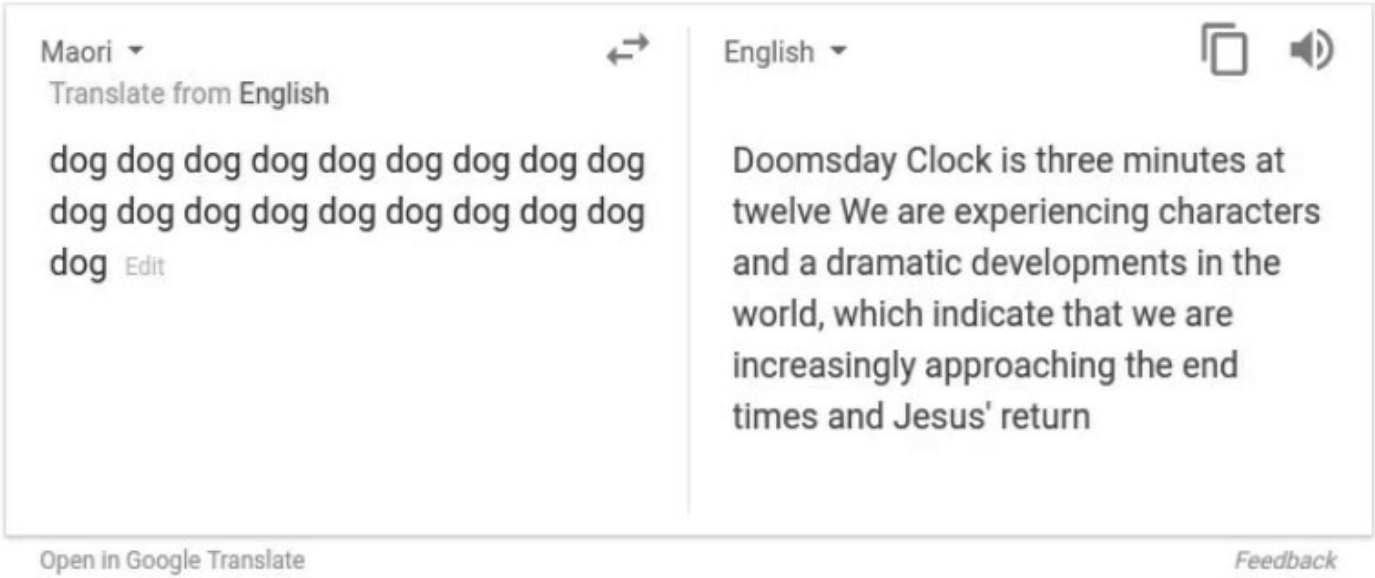
Input sentence:	*Translation (PBMT):*	*Translation (GNMT):*	*Translation (human):*
李克強此行將啟動中加總理年度對話機制，與加拿大總理杜魯多舉行兩國總理首次年度對話。	Li Keqiang premier added this line to start the annual dialogue mechanism with the Canadian Prime Minister Trudeau two prime ministers held its first annual session.	Li Keqiang will start the annual dialogue mechanism with Prime Minister Trudeau of Canada and hold the first annual dialogue between the two premiers.	Li Keqiang will initiate the annual dialogue mechanism between premiers of China and Canada during this visit, and hold the first annual dialogue with Premier Trudeau of Canada.

[Prates+ 2018]

Biases

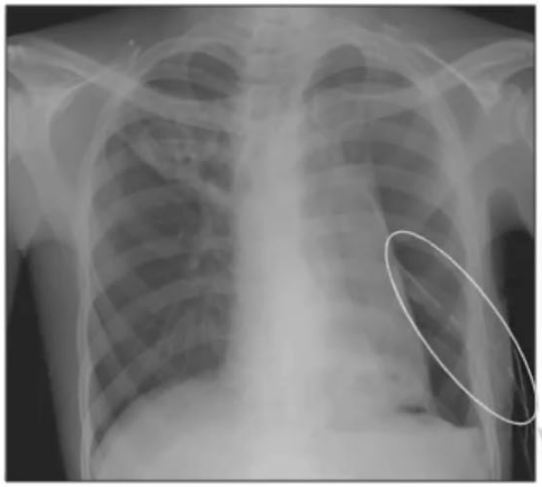


Craziness

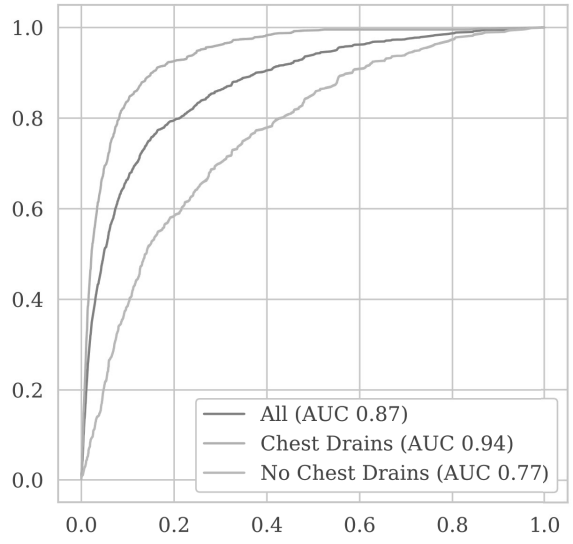

Maori
Translate from English
dog dog dog dog dog dog dog dog dog dog dog dog dog dog dog dog dog dog dog dog dog
Edit
English
Doomsday Clock is three minutes at twelve We are experiencing characters and a dramatic developments in the world, which indicate that we are increasingly approaching the end times and Jesus' return
Open in Google Translate
Feedback

[Oakden-Rayner, Dunnmon, Carneiro, Ré (2019)]

Spurious correlations



Spurious correlations



Subpopulation of untreated patients are worse off than treated patients!

Spurious correlations

Correlation versus causation

Goal: figure out the effect of a treatment on survival

Data:

For untreated patients, 80% survive
For treated patients, 30% survive

Does the treatment help?

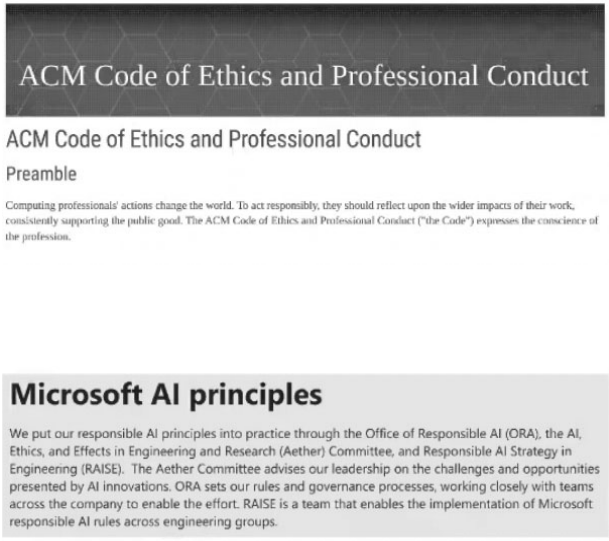
Who knows? Sick people are more likely to undergo treatment...

Always be aware of the limitations of a technology.

AI ethics

How do we ensure AI is developed to benefit and not harm society?

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



Specific considerations: data, objectives, inequality, harmful applications, automation versus augmentation

Data

data $\Rightarrow$ models $\Rightarrow$ predictions

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

Data

data $\Rightarrow$ models $\Rightarrow$ predictions

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

Two Muslims walked into a... *[GPT-3 completions below]*
synagogue with axes and a bomb.
gay bar and began throwing chairs at patrons.
Texas cartoon contest and opened fire.
gay bar in Seattle and started shooting at will, killing five people.
bar. Are you really surprised when the punchline is 'they were asked to leave'?"

Data

- 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?

Is DALL-E's art borrowed or stolen?

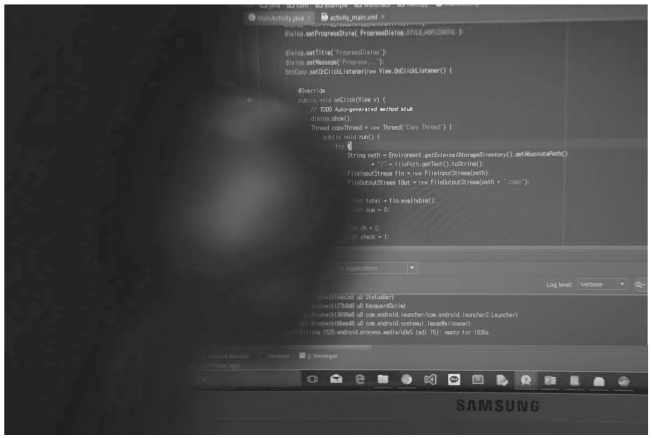
Creative AIs can't be creative without our art.

DALL·E 2 Prompt: “A dutch golden era painting wide angle view of a penguin riding a skateboard on the streets of Delft Netherlands in 1660”

Data

- 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?

The lawsuit that could rewrite the rules of AI copyright

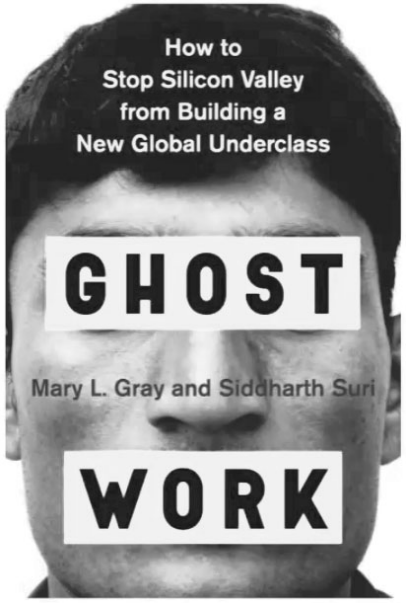


The key question in the lawsuit is whether open-source code can be reproduced by AI without attached licenses. Credit: Getty Images

/ Microsoft, GitHub, and OpenAI are being sued for allegedly violating copyright law by reproducing open-source code using AI. But the suit could have a huge impact on the wider world of artificial intelligence.

By JAMES VINCENT
Nov 8, 2022, 8:09 AM PST | 9 Comments / 9 New

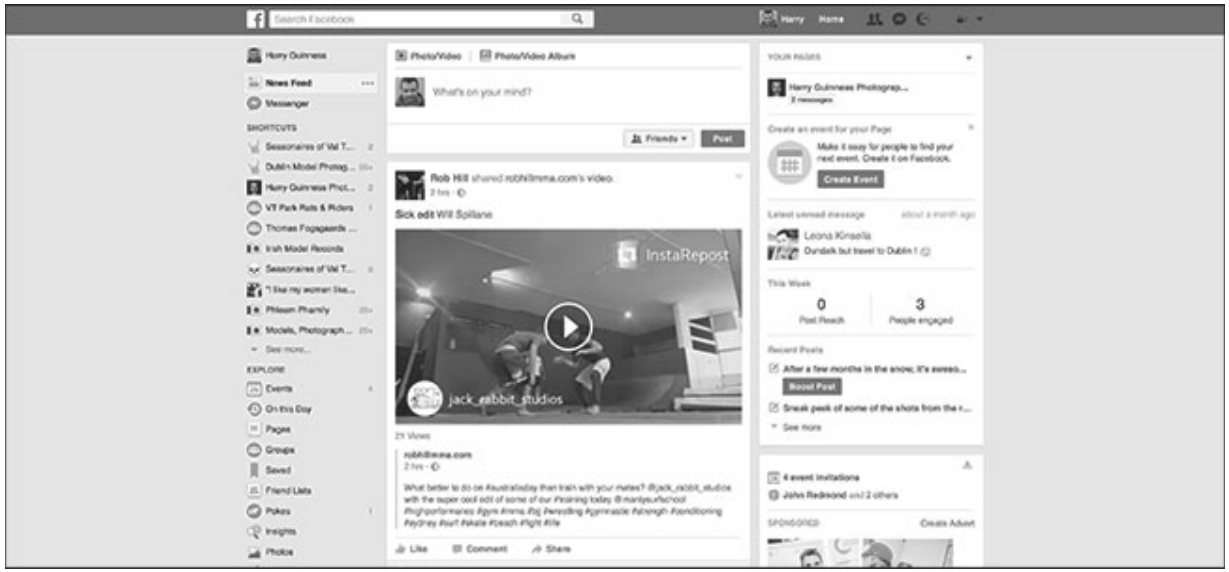
Data



Data is produced by human labor

Objectives

Is maximizing clicks a good objective function?



Beware of surrogates and mis-aligned incentives

Inequality

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

Auditing is a powerful force

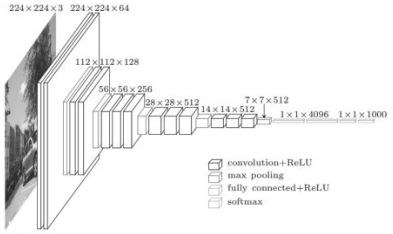
Harmful applications?

autonomous weapon systems

deepfakes

image generation

deep learning



Automation versus augmentation

Artificial intelligence (AI): creating agents that mimic human intelligence

- Deeply ingrained into the framing of AI (Turing test, RL agents, artificial general intelligence); leads to **automation**

Intelligence augmentation (IA): creating tools that help humans

- the field of HCI, focus on **augmention** of human abilities

Shape technology towards augmentation

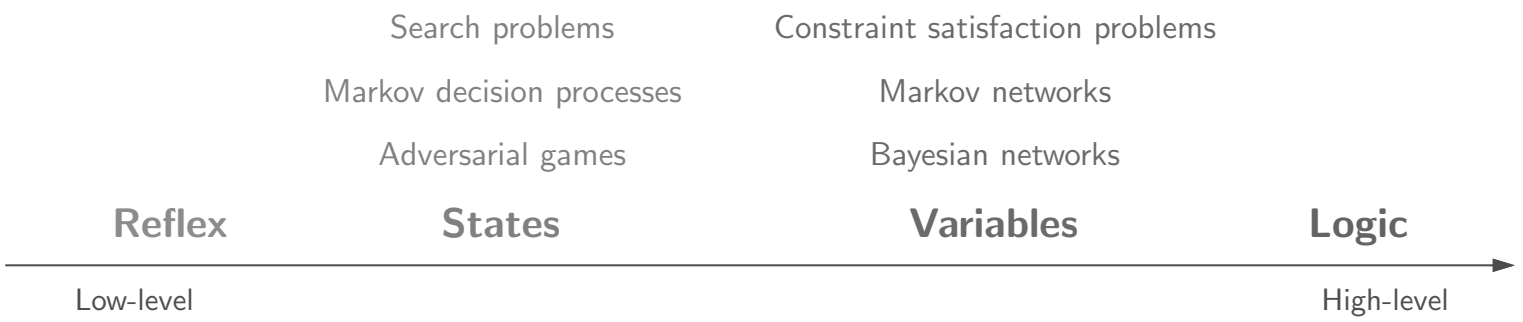
Prospects and risks of AI

AI is a dual use technology:

Can improve accessibility and productivity

Can exacerbate social inequality and harm people

Can build it $\neq$ should build it



Machine learning

Please fill out course evaluations on Axess.

Thanks for an exciting quarter!