



CPSC 100

Computational Thinking

Artificial Intelligence

Instructor: Firas Moosvi
Department of Computer Science
University of British Columbia

Agenda

- Artificial Intelligence
- Machine Learning
- Algorithmic bias
- Bias and Fairness
- Preview of Data Mining



Learning Goals

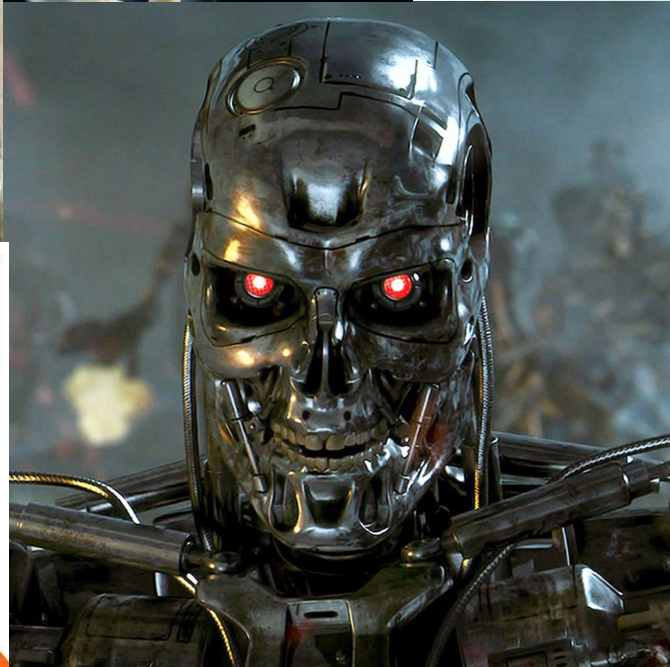
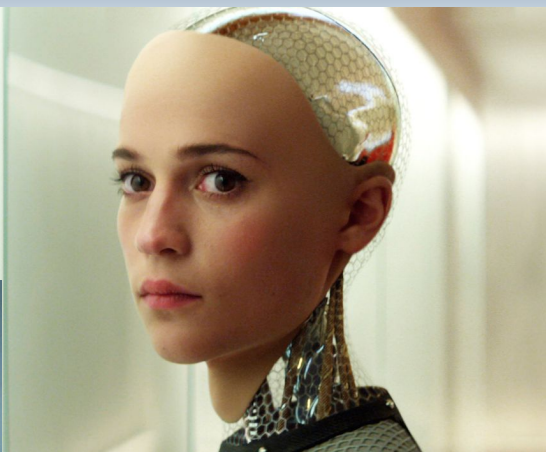
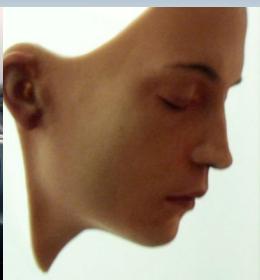
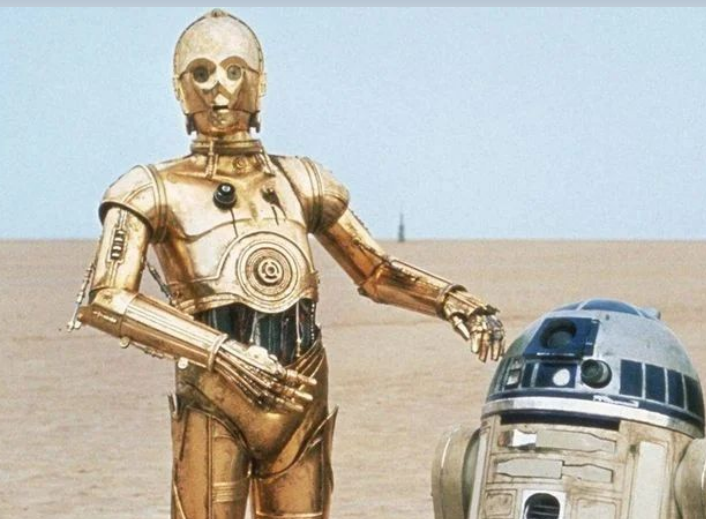
After this week's lecture, you should be able to:

- Explain the term “Artificial Intelligence” and be able to determine which methods are fall within its umbrella
- Explain what “Machine Learning” is conceptually
- Determine whether particular tasks are well-suited for machine learning or not
- Describe the difference between Supervised and Unsupervised Machine Learning
- List and describe common machine learning types
- Assess whether analyses have algorithmic bias

Artificial Intelligence

Slides courtesy of Dr. Kevin Leyton-Brown, Director of Center for AI
Decision-making and Action (CAIDA) and CIFAR Chair

What AI Means to Most People



But most modern AI systems look like...



What is Artificial Intelligence?

AI is the study, design, and development of computational processes that solve problems that **previously required human intelligence**

Not all of computer science is focused on AI.

But, it's hard to think of a field in CS that isn't **contributing fundamentally** to it

The “**AI Paradox**”: once we become familiar with a technology, we stop considering it AI

IBM computer Watson wins
clash

DeepFace: Challenge the
Yaniv Ta

Microsoft, Google Image Recognition

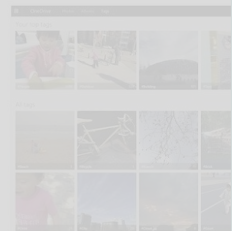
Deep learning algorithms compete

By R. Collins Johnson, CS&BS

Share Post

PORTLAND, Ore. — First computers
hurdle is image recognition — surely
too. Now Microsoft has programmed

The competition is fierce, with the first
the 2015 championship on December
they have one-upped humans too. For
benchmark of 5.1% errors with a 4.94
Microsoft by 0.04%.



The top row is a representation of the only
algorithm based on the features and the re
ception that the
(Source: Microsoft)

ImageNet, with hundreds of object ex
competition since 2010 with about 30

take the crown from the best human score. All the contest
algorithms, which are all derived from various versions of a
human brain works to varying degrees. Most of the contest

algorithm to most detail... in the context of some common

IN A
GOO
PLA

AI st

Letter

Andre E
Swetter

Nature

Abst

Skin c

the mo

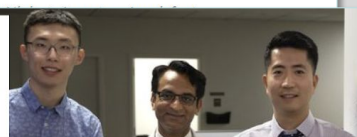
experts out

OpenAI's Dota 2 AI steamrolls world champion e-sports team with back-to-back victories

The Int
By Nick St

HEALTH + BEHAVIOR

Artificial intelligence
performs as well as
experienced



New AI Model Exceeds Human Performance at Question Answering

Microsoft's new AI translates Chinese-to-English as
well as a human translator

By Michael Hicks March 14, 2018 World Of Tech

It's all in how you teach the AI



Denise Hea

FINDINGS

Photo by N

OpenAI,

trained t

team. Th

ending t

note.

BACKGR

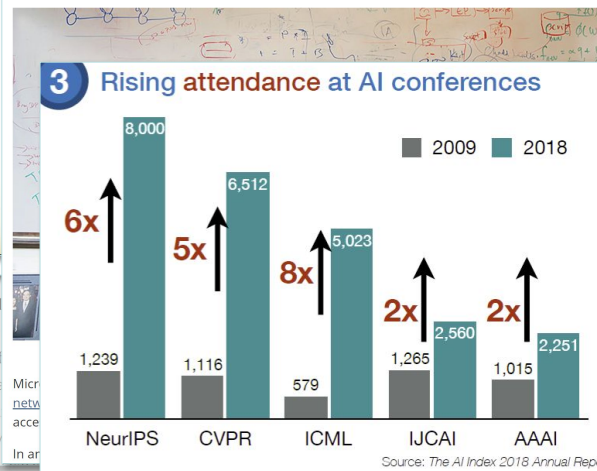
Radiologists

practicing on

accurately es

highly specia

METHOD



On May 11, 1997, IBM's Deep Blue® beat the world chess champion, Garry Kasparov, in a historic match. The match was a three-game series, with three draws, two wins for Deep Blue, and one win for Kasparov. The match was the culmination of a long history of computer chess, with the first match between a computer and a human in 1950. The match was a landmark event in the history of artificial intelligence, showing that computers could think and play like humans.

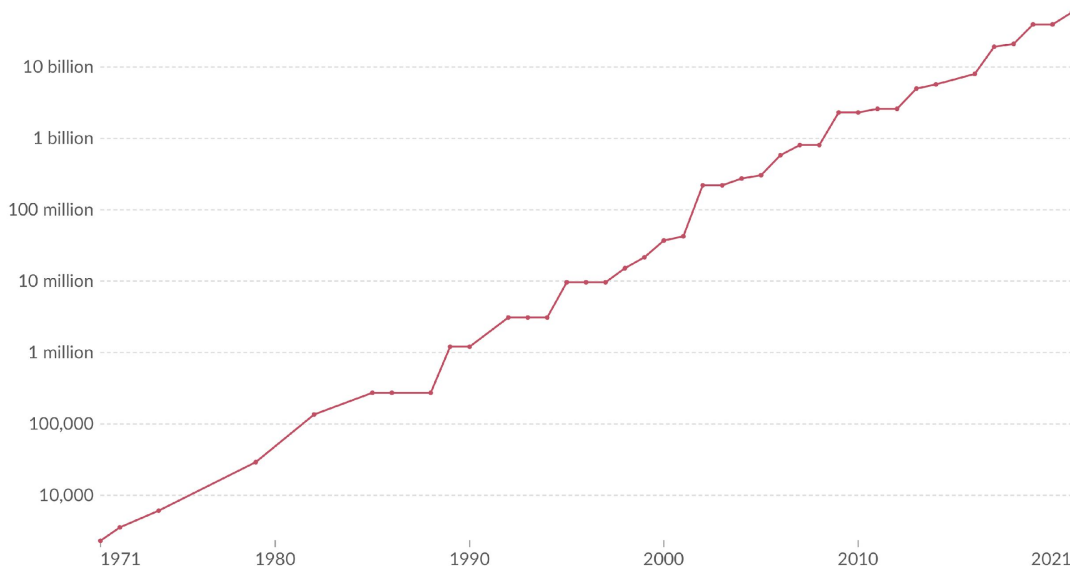
Why is it happening?

1. Scientific/mathematical **breakthroughs**, especially in machine learning
2. Growth in raw **computing power**

Moore's law: The number of transistors per microprocessor

Our World
in Data

Moore's law is the observation that the number of transistors in an integrated circuit doubles about every two years, thanks to improvements in production. It was first described by Gordon E. Moore, the co-founder of Intel, in 1965.



Data source: Karl Rupp, Microprocessor Trend Data (2022)

OurWorldInData.org/technological-change | CC BY

Current State of the Art

- **Predictive modeling**
limited mostly by quality of training data
- **Image recognition, face recognition**
becoming extremely reliable
- **Image generation** almost there
- **Language understanding & generation**
maturing extremely quickly. Watch this space
- **Fueling scientific breakthroughs**
from protein folding to astronomy
- **Robotics** remains a hard problem, but it's working well in some niche applications



Prompt: Here's a short rhyming poem by Dr. Seuss about Elon Musk's company, Neuralink which is building brain implants to achieve symbiosis with AI. Neuralink's mission statement is "If you Can't Beat'em, join'em". The poem has a structure typical of Dr. Seuss stories with simple alternate line rhymes. In the poem Elon explains why we need to merge with the AI to save humanity.

Neuralink, by Dr. Seuss

Response:

I have a friend, Elon Musk,
Who wants to join our brains to AI.
He thinks that humans can't compete
So he's gonna use AI to cheat.
They want to build a neural-link.
To link us to the AIs
And save humanity from its demise.
They want to put an AI in your brain.
They want to use your neurons as a frame.
They want to save humanity,
If you can't beat'em, join'em

<https://www.gpt-3.com> gpt-3

Machine Learning

- Narrow: building **predictive models** from data
- Broad: a new approach to computer programming, where computers **learn from experience** rather than being told what to do
- Where's it **headed**?
 - More and more data collection
 - sensors
 - behavior tracking
 - Image, text, speech understanding
 - Data mining/personalization/profiling
 - Self programming/optimization



Activity: Spot the Fakes

- Do the quiz once individually (3 minutes)
- How'd you do?
- Do the quiz together in a group of 3 or 5



Real or Not ?

The growing quality in AI images makes them harder to spot.
Can you tell if this image is real or AI generated?

Start Game

Activity: Spot the Fakes

Strategies for spotting the Fakes

- Continuity e.g Clothing: follow seams, holding guns wrong, knives, body structure
- Hands
- Blurriness
- Certain objects are hard to render (cameras, brands, logos)
- Lighting & Reflections
- Blatant errors (smiling, extra teeth)
- Physics doesn't add up (things that are wet)

Machine Learning



Machine Learning

A different way to solve problems

Machine learning uses computer programs to model data. It can be used to extract hidden patterns, make predictions in new situation, or generate novel content.

A field of study that gives computers the ability to learn without being explicitly programmed.

– Arthur Samuel (1959)



Machine Learning

Traditional programming vs. ML

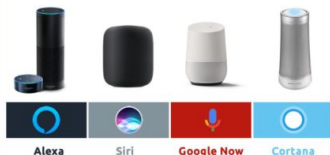
- Imagine writing a Python program for spam identification, i.e., whether a text message or an email is spam or non-spam.
- Traditional programming
 - Come up with rules using human understanding of spam messages.
 - Time consuming and hard to come up with robust set of rules.
- Machine learning
 - Collect large amount of data of spam and non-spam emails and let the machine learning algorithm figure out rules.

Machine Learning

Prevalence of ML

Let's look at some examples.

Voice assistants



Google news

- Armed man who broke into Trudeau residence charged with threatening to kill or injure PM
The Guardian - 1 hour ago
- Corey Hurren, alleged Rideau Hall intruder, threatened Trudeau: RCMP officer
Global News - 4 hours ago
 - Corey Hurren had multiple firearms, uttered threat against Trudeau, court documents allege
CBC.ca - 2 hours ago
 - Man arrested near Rideau Hall had several weapons, threatened PM Trudeau: RCMP
CTV News - 22 minutes ago

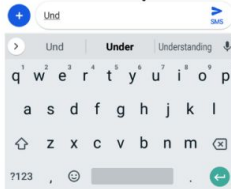
Recommendation systems



Face recognition



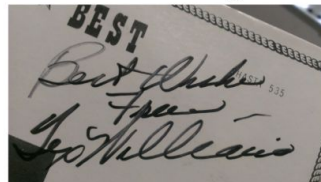
Auto-completion



Stock market prediction



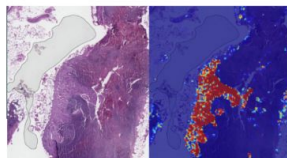
Character recognition



Self-driving car



Cancer diagnosis



Drug discovery



AlphaGo



Machine Learning

Activity: For what type of problems ML is appropriate? (~5 mins)

Discuss with your neighbour for which of the following problems you would use machine learning

- Finding a list of prime numbers up to a limit
- Given an image, automatically identifying and labeling objects in the image
- Finding the distance between two nodes in a graph

Machine Learning

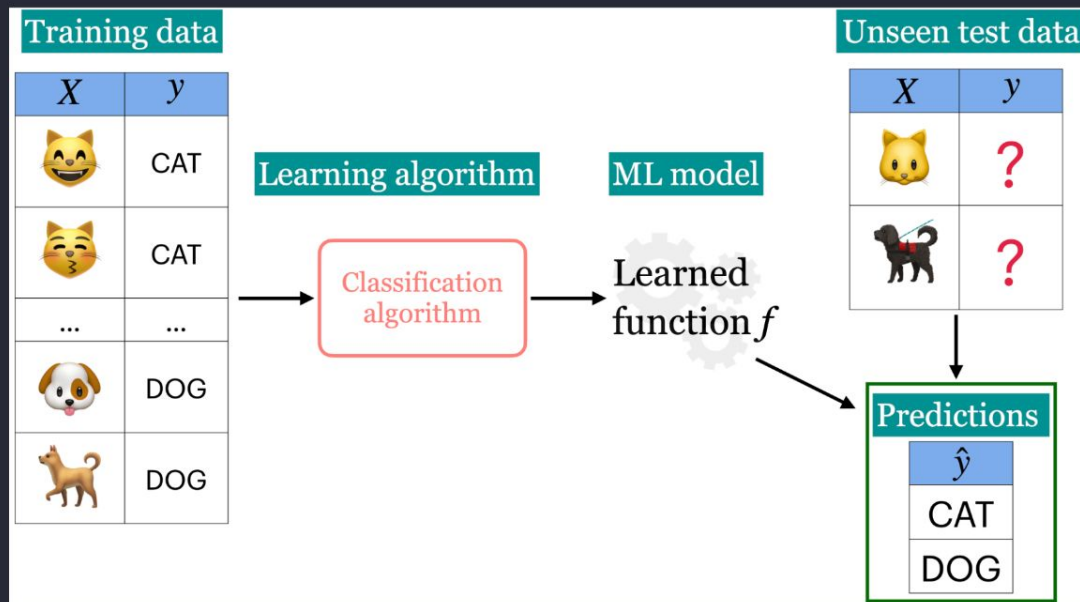
Types of machine learning

Here are some typical learning problems.

- **Supervised learning** (Gmail spam filtering)
 - Training a model from input data and its corresponding targets to predict targets for new examples.
- Unsupervised learning (Google News)
 - Training a model to find patterns in a dataset, typically an unlabeled dataset.
- Reinforcement learning (AlphaGo)
 - A family of algorithms for finding suitable actions to take in a given situation in order to maximize a reward.
- Recommendation systems (Amazon item recommendation system)
 - Predict the “rating” or “preference” a user would give to an item.

What is supervised learning?

- Training data comprises a set of observations (X) and their corresponding targets (y).
- We wish to find a model function f that relates X to y .
- We use the model function to predict targets of new examples.



Unsupervised learning

- Training data consists of observations (X) without any corresponding targets.
- Unsupervised learning could be used to group similar things together in X or to find underlying structure in the data.

Training data



Learning algorithm

Clustering
algorithm

ML model



CATS



DOGS



Bias and Fairness

Human bias

Bias in people refers to our tendency to take quick decisions based on little information

Published online 11 April 2011 | Nature | doi:10.1038/news.2011.227

News

Hungry judges dispense rough justice

When they need a break, decision-makers gravitate towards the easy option.

Zoë Corbyn

Journal of Economic Perspectives—Volume 12, Number 2—Spring 1998—Pages 41–62

Evidence on Discrimination in Mortgage Lending

Helen F. Ladd

Science faculty's subtle gender biases favor male students

Corinne A. Moss-Racusin^{1,2}, John F. Dovidio³, Victoria L. Brescoll⁴, Mark A. Graham^{1,2}, and Jo Handelsman^{1,2}

¹Department of Molecular, Cellular, and Developmental Biology, ²Department of Psychology, ³School of Management, and ⁴Department of Psychiatry, Yale University, New Haven, CT 06520

Edited by Shirley Tilghman, Princeton University, Princeton, NJ, and approved August 21, 2012 (received for review July 2, 2012)

The Observer
Stop and search


Racial bias in police stop and search getting worse, report reveals

Despite reforms, black people are nine times more likely than white people to be checked for drugs

Mark Townsend
Home Affairs
Editor

4 comments
Sat 10 Jul 2010 20:13 BST

1716
The article is one of 7 comments

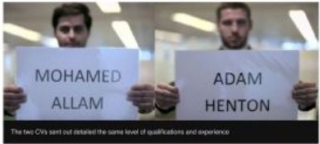


Is it easier to get a job if you're Adam or Mohamed?

By Zack Abramson and Clara Mancuso
BBC Inside Out

6 February 2017

f t e g+ Share



The two CVs sent out detailed the same level of qualifications and experience

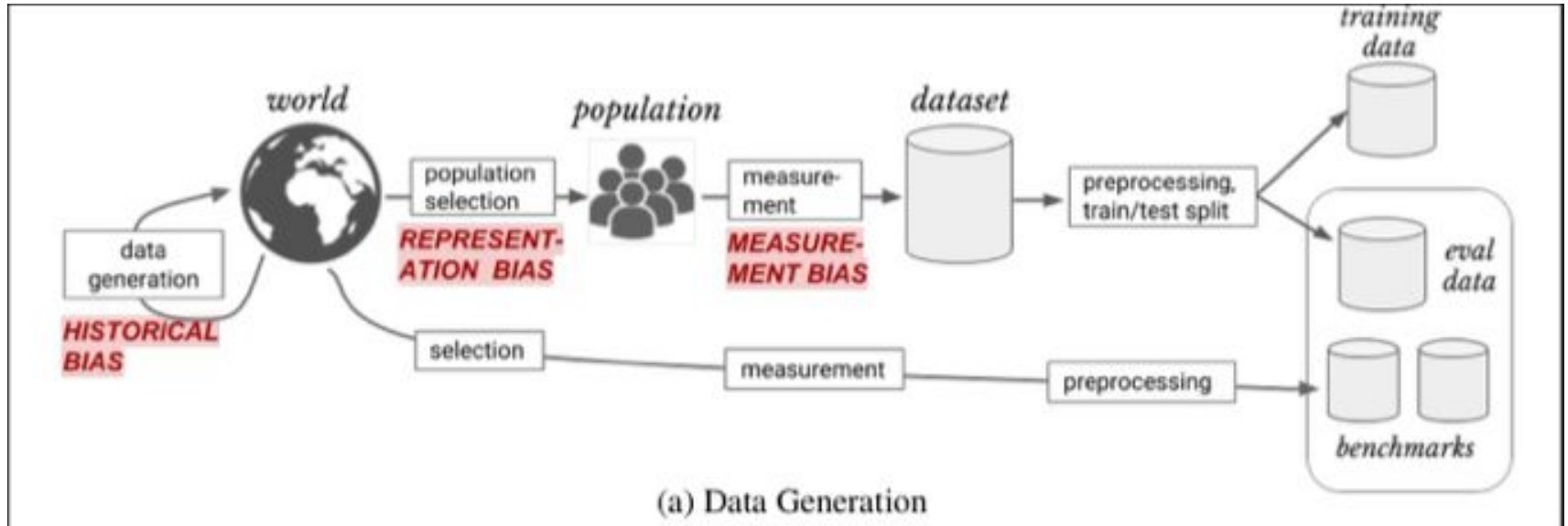
A job seeker with an English-sounding name was offered three times the number of interviews than an applicant with a Muslim-sounding name, BBC test found.

Why worry about bias in algorithms

Decisions made by a ML algorithm are:

- Cheap
- Scalable
- Automated
- Self-reinforcing
- Seemingly objective
- Often lacking appeals processes
- Not just predicting but also causing the future

Sources of bias in ML algorithms



Representation bias

Representation bias arises while defining and sampling a development population. It occurs when the development population under-represents, and subsequently fails to generalize well, for some part of the use population. 1

1. **The sampling methods only reach a portion of the population.** For example, datasets collected through smartphone apps can under-represent lower-income or older groups, who are less likely to own smartphones. Similarly, medical data for a particular condition may be available only for the population of patients who were considered serious enough to bring in for further screening.
2. **The population of interest has changed or is distinct from the population used during model training.** Data that is representative of Boston, for example, may not be representative if used to analyze the population of Indianapolis. Similarly, data representative of Boston 30 years ago will likely not reflect today's population.

Measurement bias

Measurement Bias arises when choosing and measuring features and labels to use; these are often proxies for the desired quantities. The chosen set of features and labels may leave out important factors or introduce group- or input-dependent noise that leads to differential performance.

3. **The defined classification task is an oversimplification.**

In order to build a supervised ML model, some label to predict must be chosen. Reducing a decision to a single attribute can create a biased proxy label because it only captures a particular aspect of what we really want to measure. Consider the prediction problem of deciding whether a student will be successful (e.g., in a college admissions context). Fully capturing the outcome of 'successful student' in terms of a single measurable attribute is impossible because of its complexity. In cases such as these, algorithm designers resort to some available label such as 'GPA' (Kleinberg et al. 2018), which ignores different indicators of success achieved by parts of the population.

1. **The measurement process varies across groups.** For example, if a group of factory workers is more stringently or frequently monitored, more errors will be observed in that group. This can also lead to a feedback loop wherein the group is subject to further monitoring because of the apparent higher rate of mistakes (Barocas and Selbst 2016).

2. **The quality of data varies across groups.** Structural discrimination can lead to systematically higher error rates in a certain group. For example, women are more likely to be misdiagnosed or not diagnosed for conditions where self-reported pain is a symptom (Calderone 1990). In this case, "*diagnosed* with condition X" is a biased proxy for "has condition X."

Historical bias

Historical bias arises when there is a misalignment between world as it is and the values or objectives to be encoded and propagated in a model. It is a normative concern with the state of the world, and exists even given perfect sampling and feature selection.

1

Example: image search In 2018, 5% of Fortune 500 CEOs were women (Zarya, 2018). Should image search results for “CEO” reflect that number? Ultimately, a variety of stakeholders, including affected members of society, should evaluate the particular harms that this result could cause and make a judgment. This decision may be at odds with the available data even if that data is a perfect reflection of the world. Indeed, Google has recently changed their Image Search results for “CEO” to display a higher proportion of women.

Fairness in algorithms

- Increasing attention on algorithms being **fair**, not just accurate
- Fairness can be measured as:
 - demographic (or statistical) parity: population percentage should be reflected in the output classes
 - Equality of false negatives or equalized odds: constant false-negative (or both false-negative and true-negative) rates across groups.
 - Equal opportunity: equal True Positive Rate for all groups
 - Other metrics...
- Accuracy and fairness tend to be at odds with each other.
- Algorithms can be audited to test their fairness.
- *Are we ethically required to sacrifice accuracy for fairness? To what extent?*

When the metric becomes the target (Goodhart's Law)

"When a measure becomes a target it ceases to be a good measure"

- Metrics introduced in the [British public healthcare system](#) (e.g. waiting time in ER) caused people to game it:
 - Cancelled scheduled operations to draft extra staff to ER
 - Required patients to wait outside the ER, e.g. in ambulances
 - Put stretchers in hallways and classified them as "beds"
 - Hospital and patients reported different wait times
- Big Data is significantly changing college applications
 - Universities are given higher ranking for things such as receiving more applications, being more selective, and having more students accept their offers (while tuition is not considered)
 - This even pushed some mid-tier universities to reduce the number of offer letter sent out, especially to good students who they think will not accept. Can affect applications to "safety schools"
- *Can you think of more examples?*
- *Can you think of ways to avoid this trap?*

Ethics of pricing algorithms

- Algorithms are currently used to adjust prices based on:
 - Willingness of buyer
 - Availability
- Uber surge pricing:
 - In 2014, terrorists attacked a café in Sidney, holding 10 customers and 8 employees hostage for 16 hours
 - During this time, people from the surrounding areas were evacuated. Transportation was disrupted.
 - Uber prices adapted by increasing the rate to a minimum of 100\$
 - In general, underserved (poorer) areas get worse rates under current pricing policy
 - Drivers are also subjected to different pricing/waiting times
 - https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4331080
- *Is Uber morally obliged to avoid such pricing disparities?*

From trading to gambling addiction?

- Case study: Robinhood Trading app
- Designed to make trading more accessible and equitable. Robinhood's mission statement is "To democratize finance for all"
- Concerns around the most “dopamine-inducing” features ([source](#)):
 - Green confetti to celebrate transactions.
 - A constant update of stock related articles.
 - A colorful, eye-catching interface.
 - Emoji phone notifications.
 - One-click trading for instant gratification.
 - Free stocks in the shape of lottery tickets.



Technology that will be in widespread use

- Tailored solutions for **specific tasks**, not general intelligence
- Prototypes **that work today** in labs & narrow deployments
- Some examples:
 - **Non-text input modalities** (vision; speech)
 - **Consumer modeling** (recommendation; marketing)
 - **Cloud services** (translation; question answering; AI-mediated outsourcing)
 - **Transportation** (automated trucking; some self-driving cars)
 - **Industrial robotics** (factories; some drone applications)
 - **AI knowledge work** (logistics planning; radiology; legal research; call centers)
 - **Policing & security** (electronic fraud; cameras; predictive policing)



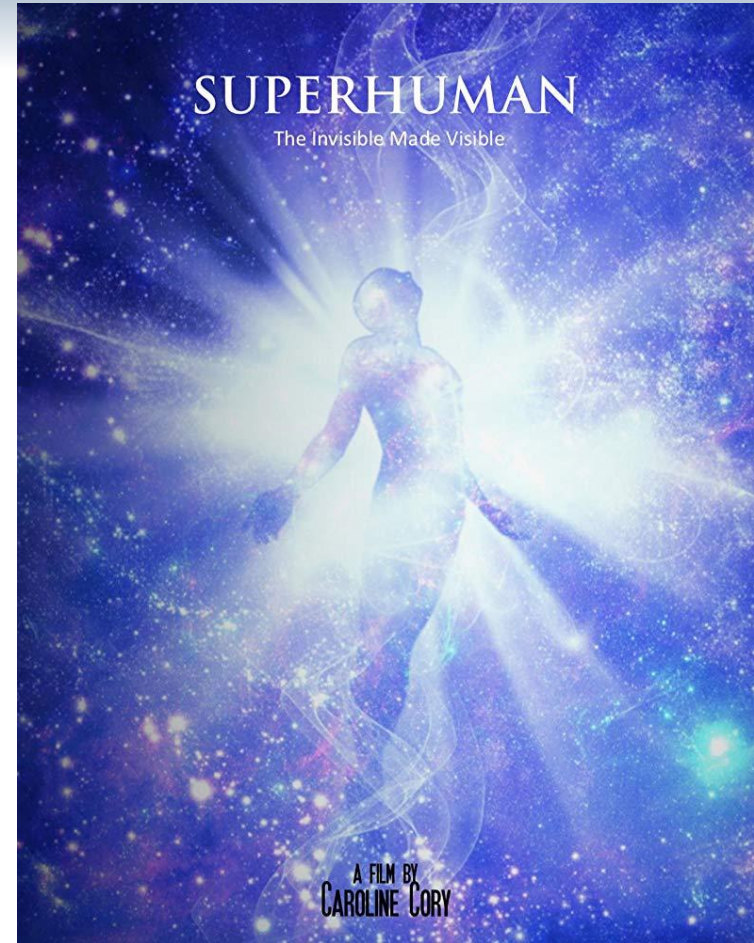
Technologies that won't take off as quickly

- Overall, areas in which
 - major entrenched **regulatory regimes** need to be navigated
 - there exist **social/cultural barriers** to adoption
 - the **human touch** is crucial
 - substantial **new hardware** would need to be developed
- Some **examples**:
 - childcare, healthcare, eldercare
 - education
 - coaching, counselling
 - consumer robots beyond niche applications
 - semantically rich language understanding



Superhuman Intelligence

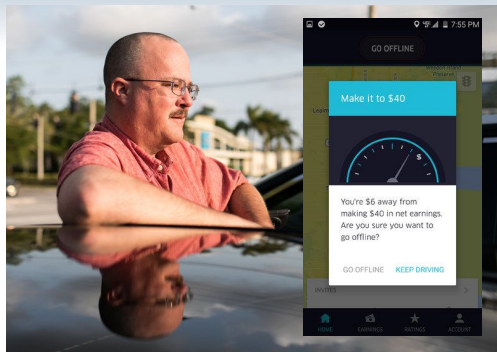
- AI systems will increasingly be capable of reaching **human-level performance**
- **Superhuman intelligence** isn't such a foreign, scary thing
 - governments, corporations, NGOs exhibit behavior much more sophisticated and complex than that of any individual
- Many important problems need superhuman intelligence; AI can help
 - improved **collective decision making**
 - more efficient use of **scarce resources**
 - addressing **underserved communities**
 - **climate change**; other societal challenges



Ethics of AI

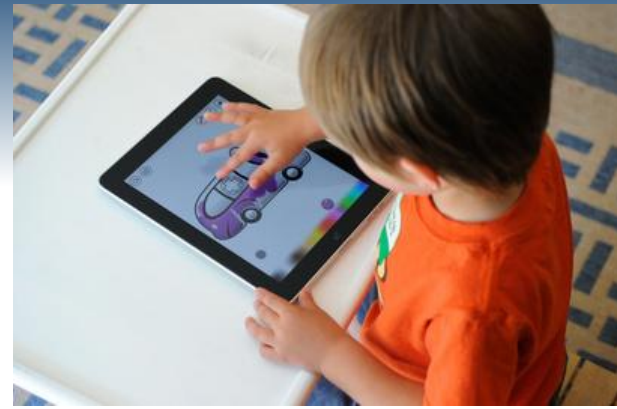
Will a new technology:

- disempower **individuals vs corporations?**
 - ⇒ user modeling; **data mining**; fostering addictive behaviors; developmental effects on children
- disempower **individuals vs governments?**
 - ⇒ facilitate disinformation (deep fakes; bots masquerading as people; filter bubbles); enable qualitatively new military or security tactics
- take **autonomous actions** in a way that obscures responsibility
 - ⇒ autonomous weapons; self-driving cars; loan approval systems
- disproportionately affect **vulnerable/marginalized groups**
 - ⇒ automated decision making tools trained in ways that may encode existing biases



Social Impact

- How will AI technologies **transform society**?
- Will there be a **social backlash** against AI?
 - If so, what will be considered AI?
- This **generation of children** will grow up taking for granted many technologies that strike us as magical
- How will **human relationships** change in the presence of always-available social agents?
- As we are increasingly **augmented by AI**, what are our inherent cognitive/emotional/motivational limitations, beyond which augmentation won't help?



Ethical consideration of advances in AI

- *Is it wrong to create machines capable of making human labor obsolete? Will humans become demoralized by the presence of vastly more intelligent robots?*
- *How can we ensure that intelligent robots will not be put to an evil purpose by a malevolent human? How can we ensure they do not adopt malevolent purposes themselves?*
- *Is it morally acceptable to create “personal” (self-conscious) AI?*

