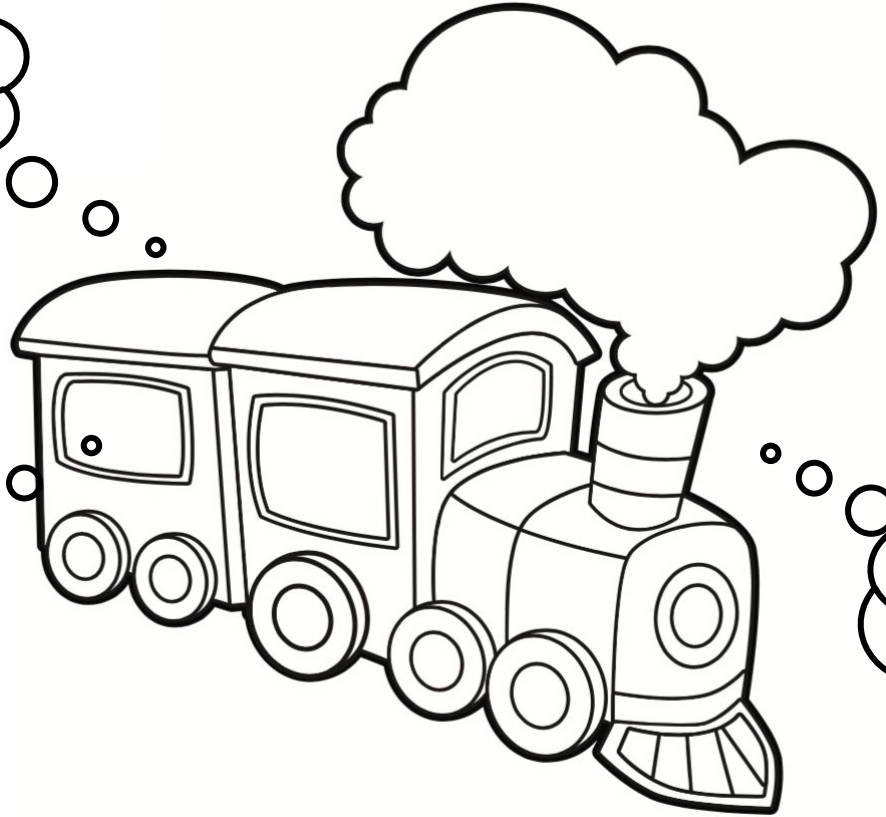
Machine learning can't do the thinking

Inga Strümke, December 2019

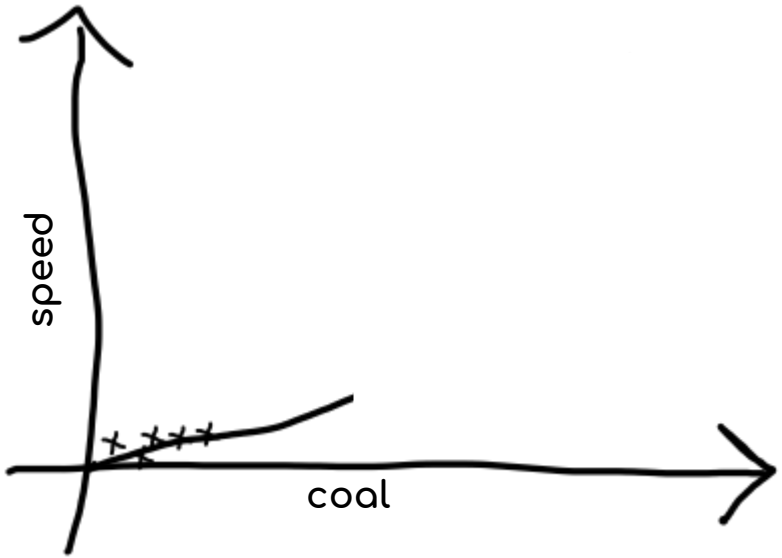
pwc

omg so fast

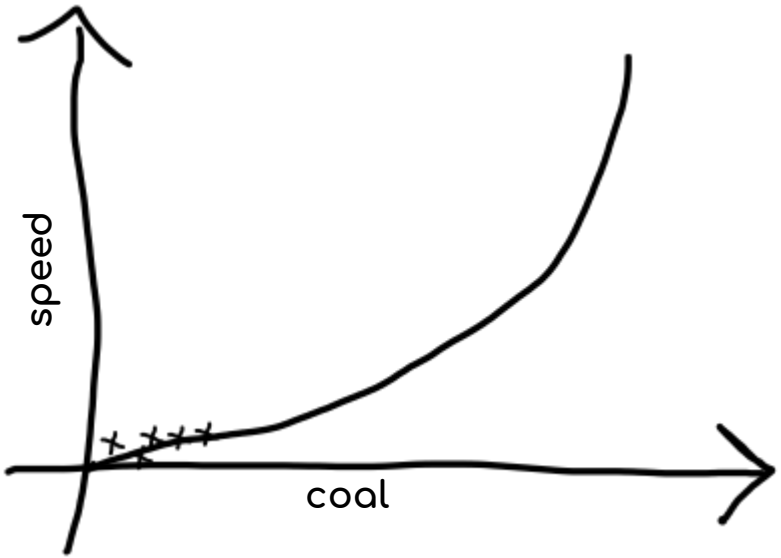


... fastest ever?

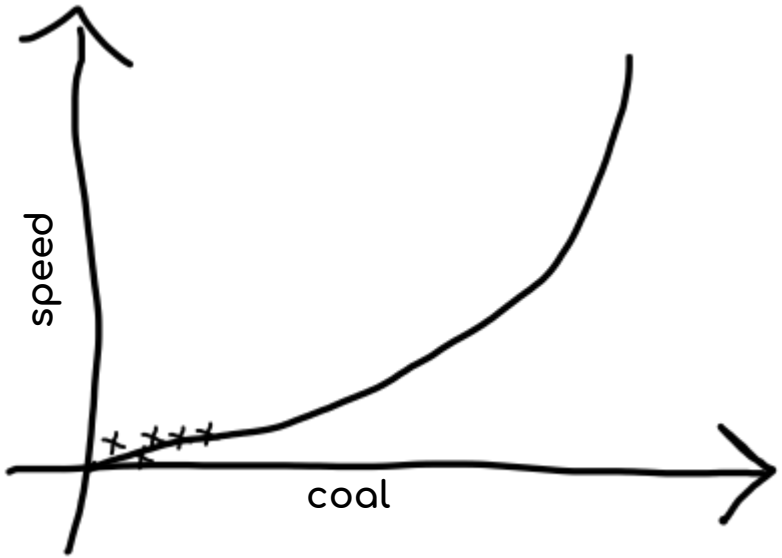
more coal, more fast!


speed
coal

Q: will $v_{train} \to \infty$ if the engine gets big enough?



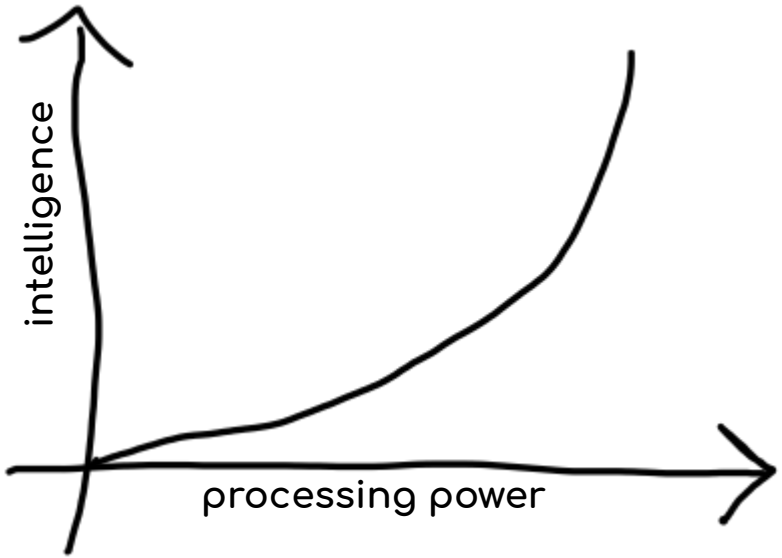
Q: will $v_{train} \rightarrow \infty$ if the engine gets big enough?



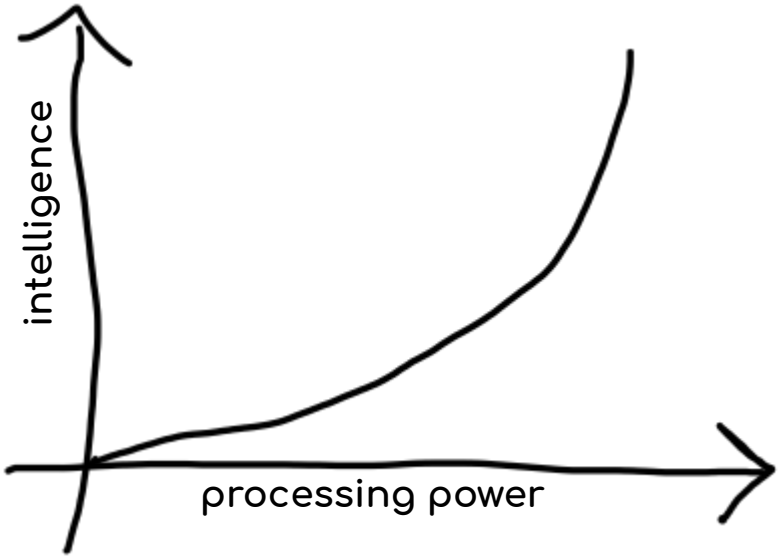
A: No, because

1. Architecture: Rails, friction, air resistance, locomotive
2. Physical limitations: Sound barrier, light speed,

Q: will AGI arise if we increase processing power?

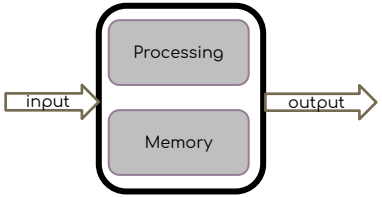


Q: will AGI arise if we increase processing power?



A: ... :)

1. Architecture



Moravec's paradox

2. Physical limitations?

No need to worry about AGI yet (unfortunately)

How then to achieve AI world domination?

(and how to make this a good world)

Image courtesy Shutterstock

The road to AI world domination:

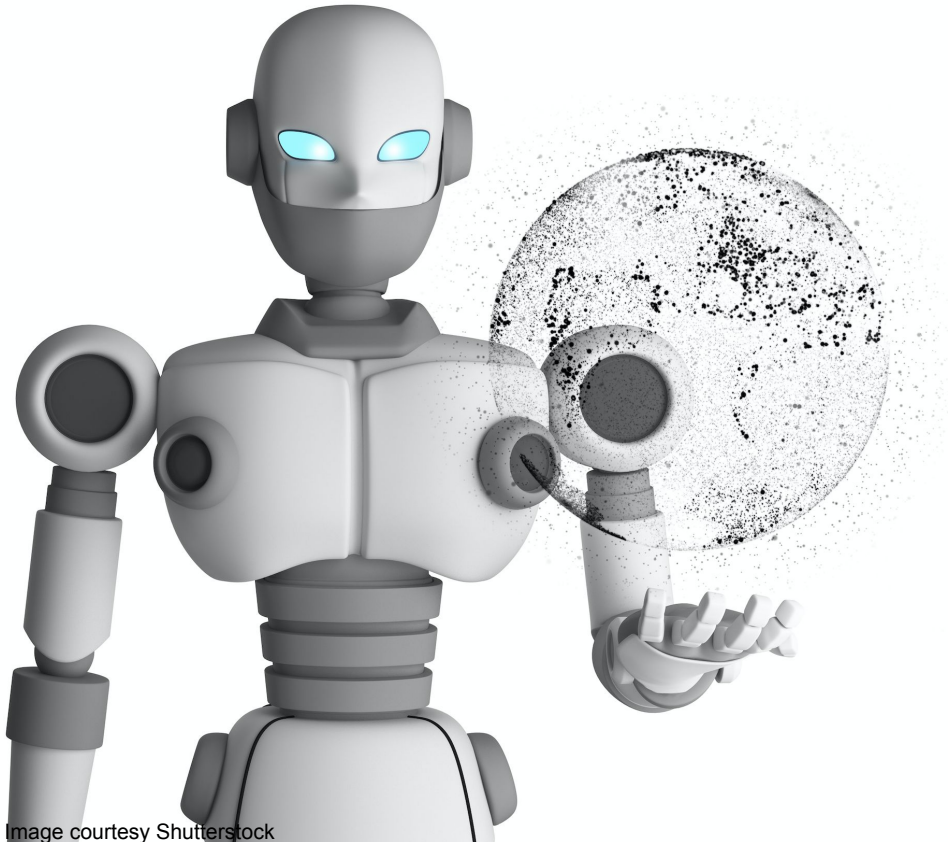


Image courtesy Shutterstock

- DATA

Luca Pacioli

The accountant who changed the world

Venice 14xx: Invented “double-entry bookkeeping”

E.g.: You sell 48 pepper for 3 gold coins ⇒

Gold coins	**Pepper**
10	200
13	152

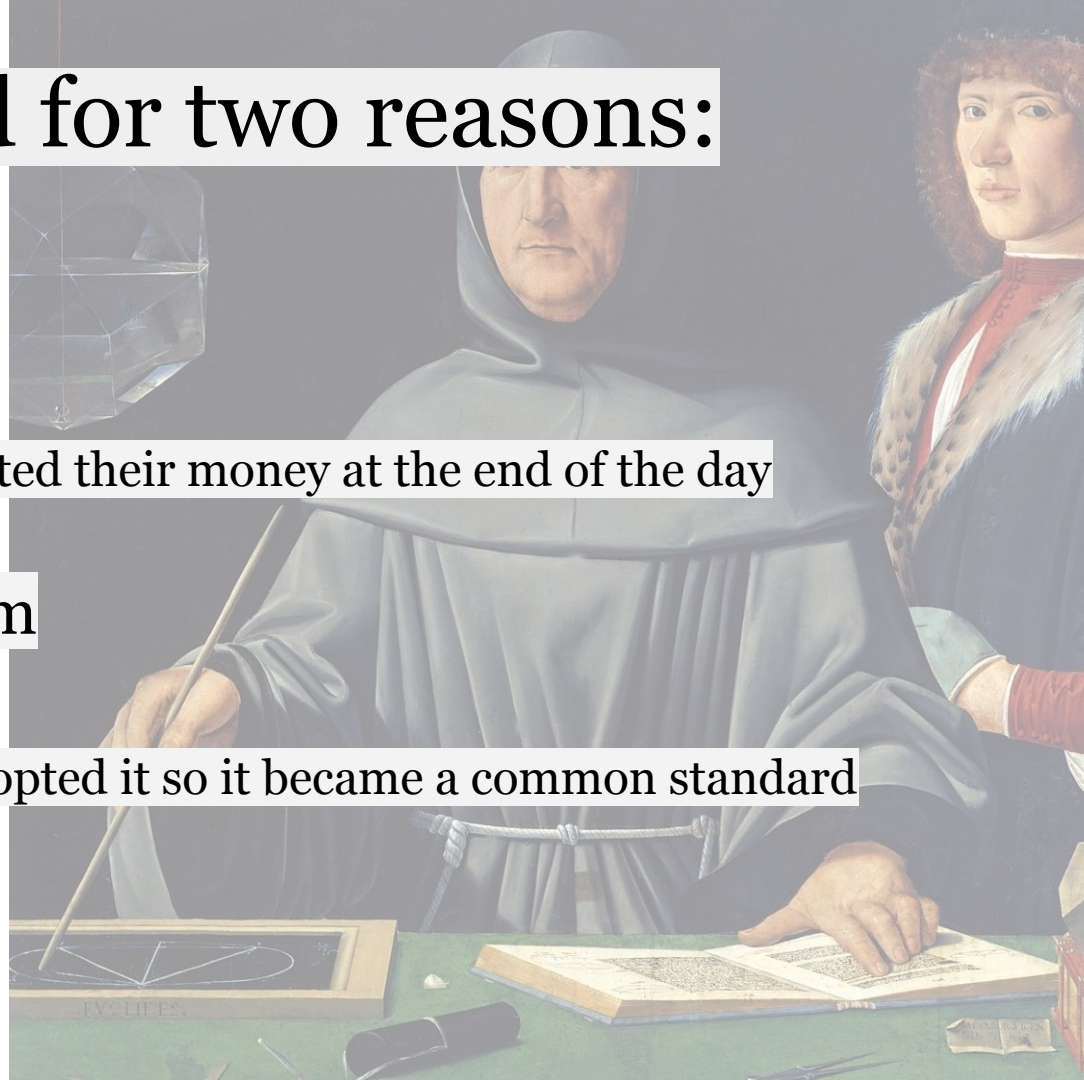
Changed the world for two reasons:

1. It works REALLY well

Before, people kept diaries and counted their money at the end of the day

2. Part of a powerful system

The large and influential systems adopted it so it became a common standard



Use standardised format

⇒

Containers

Imagine a future where you say

“it used to take me a week to get all the data for my project. These days, it takes two hours”

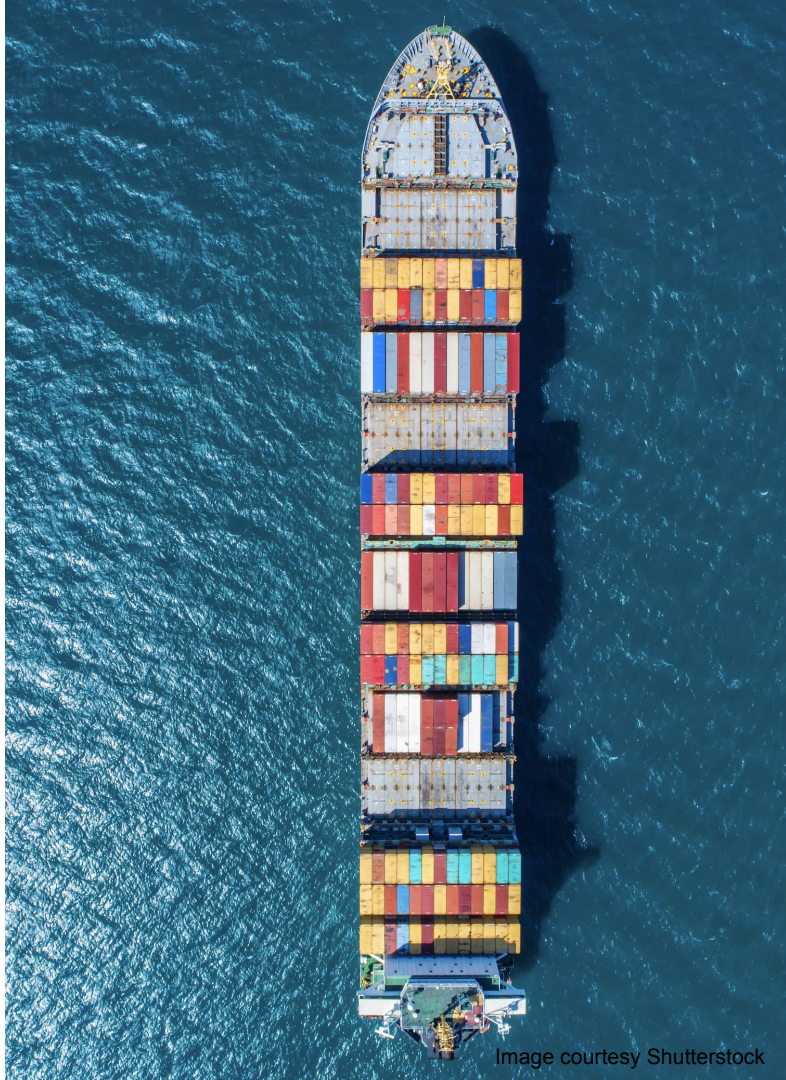


Image courtesy Shutterstock

About two years ago, we, at our newly formed Machine Learning Infrastructure team started asking our data scientists a question: “What is the hardest thing for you as a data scientist at Netflix?” We were expecting to hear answers related to large-scale data and models, and maybe issues related to modern GPUs. Instead, we heard stories about projects where getting the first version to production took surprisingly long — mainly because of mundane reasons related to software engineering. We heard many stories about difficulties related to data access and basic data processing. We sat in meetings where data scientists discussed with their stakeholders how to best version different

Savin Goyal · 3rd

Machine Learning Infrastructure at Netflix

Once we've solved the data issues, what could ruin it for us?

Image courtesy Shutterstock

1. Not using AI where we should use it (missed use)

2. Using AI in a bad way (misuse 1: being unethical)

3. Using AI the wrong way (misuse 2: not understanding it)

Example from Norway
Targeted marketing

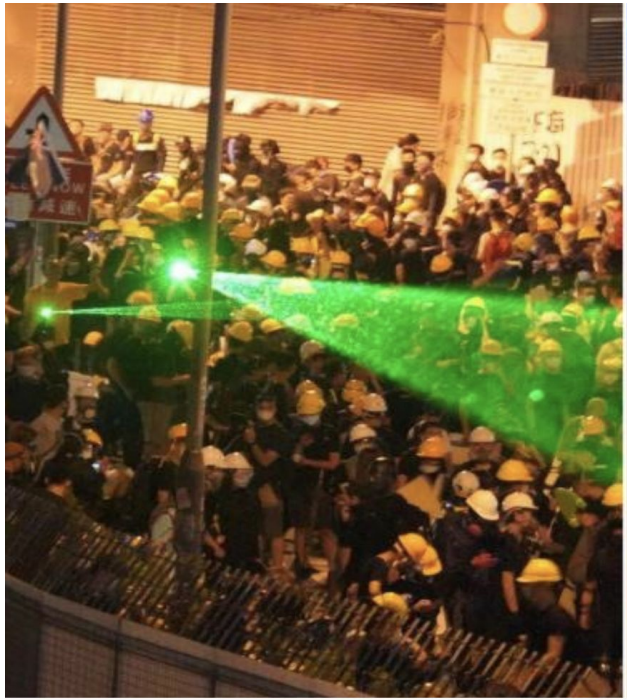
Peppes Pizza:

Largest pizza restaurant chain in Scandinavia

Example from Norway
Targeted marketing

Peppes Pizza:

Caused a small media storm and would have been a **reputational disaster** for a smaller/younger company



Hong Kong protesters use lasers to confuse police and to interfere with their cameras. (Photos posted o

HONG KONG / CCTV - *08/06/2019*

Hong Kong protesters use lasers to confuse police and damage cameras

This is awesome. We fight crime like never before, and society gains from it. Everybody's better off!

Get your hands off MY data

Different ethics, different metrics, but one thing is certain:

The "eastern model" for sharing data provides perfect conditions for machine learning.

The road to AI world domination:

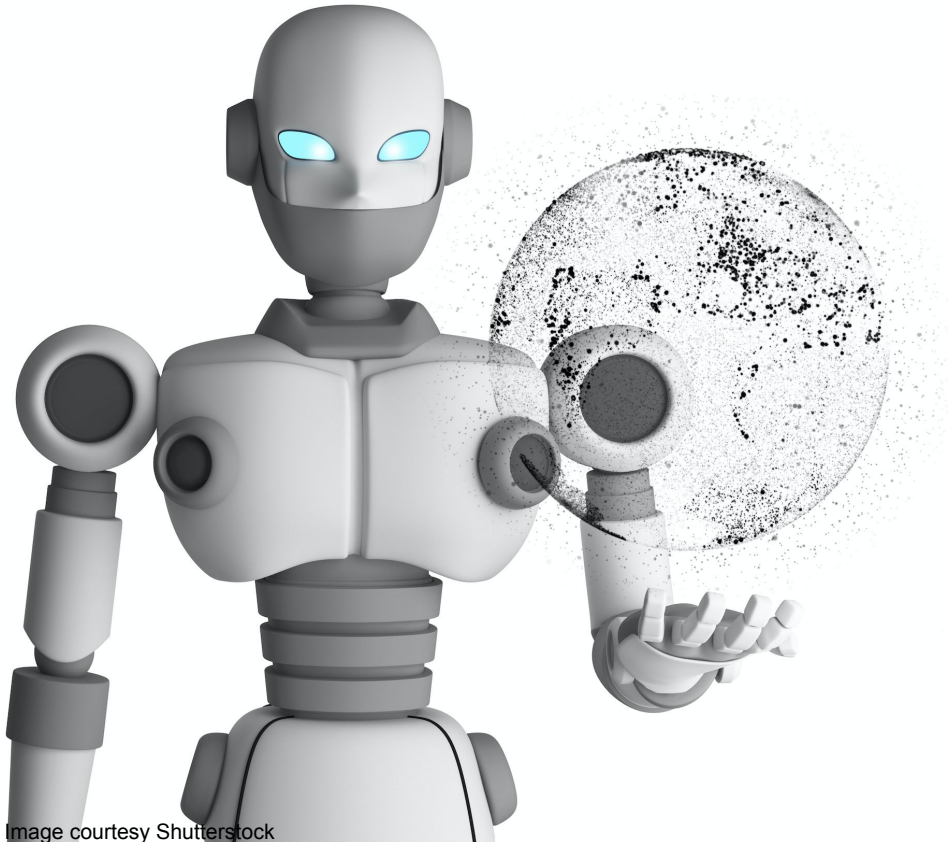


Image courtesy Shutterstock

- DATA ✓
- ETHICS

Exclusive: Google cancels AI ethics board in response to outcry

The controversial panel lasted just a little over a week.

By Kelsey Piper | Apr 4, 2019, 7:00pm EDT

SHARE

Our favourite bad guy these days: BIAS
(the part where an algorithm learns from the past)

Amazon reportedly scraps internal AI recruiting tool that was biased against women

Bias is not as simple as over- and underrepresented groups

Task: Determine risk score for selecting which patients to treat first

Problem: White patients receive higher risk scores, than black patients who are equally sick

Why?

- Medical histories are used to predict **how much people will cost the health-care system**
- Black patients have historically incurred **lower costs** than white patients with the same conditions

⇒ The problem is not unequal representation in the data, but the **loss metric** in the cultural and historical (i.e. data) context

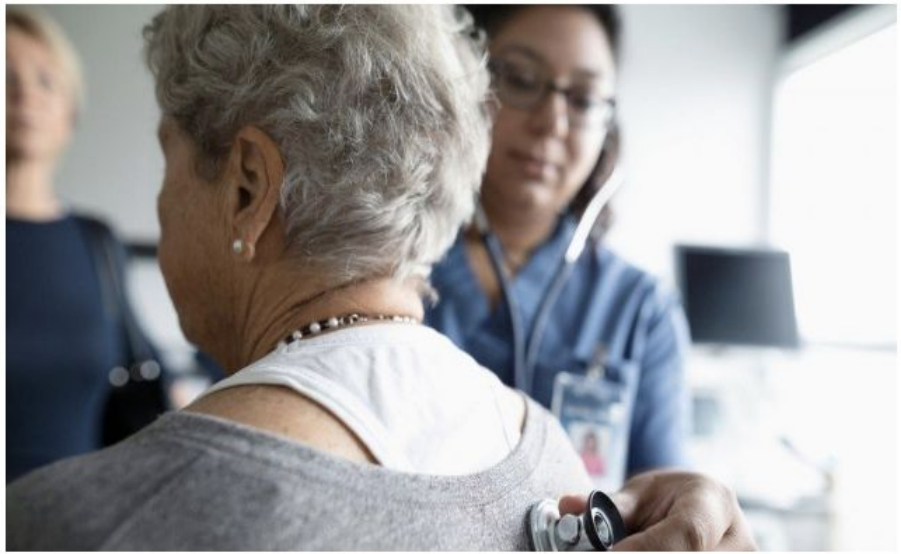
Easy fix: Disparity can be reduced >80% by predicting also the number of times a chronic condition will reoccur

Algorithmic bias can be corrected, **if you can catch it.**

MIT Technology Review

Artificial Intelligence Oct 25

A biased medical algorithm favored white people for health-care programs



technologyreview.com/f/614626/a-biased-medical-algorithm-favored-white-people-for-healthcare-programs

Bias is hard to get rid of.

“

… when Sussman was a novice, Minsky once came to him as he sat hacking at the PDP-6.

"What are you doing?", asked Minsky.
"I am training a randomly wired neural net to play Tic-tac-toe", Sussman replied.
"Why is the net wired randomly?", asked Minsky.
"I do not want it to have any preconceptions of how to play", Sussman said.

Minsky then shut his eyes.

"Why do you close your eyes?" Sussman asked his teacher.
"**So that the room will be empty.**"

At that moment, Sussman was enlightened.

What else is new?

Marvin Minsky, “Father of Artificial Intelligence”

The New York Times

ECONOMIC VIEW

Biased Algorithms Are Easier to Fix Than Biased People

Racial discrimination by algorithms or by people is harmful — but that's where the similarities end.

AI - in the sense of advanced data analysis - is actually good news

They highlight existing problems

The algorithms don't decide what is ethical

Bias vs fairness

While **bias** is a well-defined concept and can be controlled, **fairness** is a societal construct where different interests are put up against each other

COMPAS

- **Task: Calculate risk score** → should defendant be kept in jail while awaiting trial?
- Aim: Make the criminal legal system fairer, by replacing the judges' intuition with a testable tool
- *Algorithm does not have access to protected attributes (ethnicity, age, gender, disability, etc)*

No ethnicity info in the training data

Can the algorithm still have racial bias?

Why?

MIT Technology Review: courtroom algorithm game

COMPAS

- **Task: Calculate risk score** → should defendant be kept in jail while awaiting trial?
- Aim: Make the criminal legal system fairer, by replacing the judges' intuition with a testable tool
- *Algorithm does not have access to protected attributes (ethnicity, age, gender, disability, etc)*

The algorithm has racial bias, as demonstrated in 2016 by ProPublica.

Why?

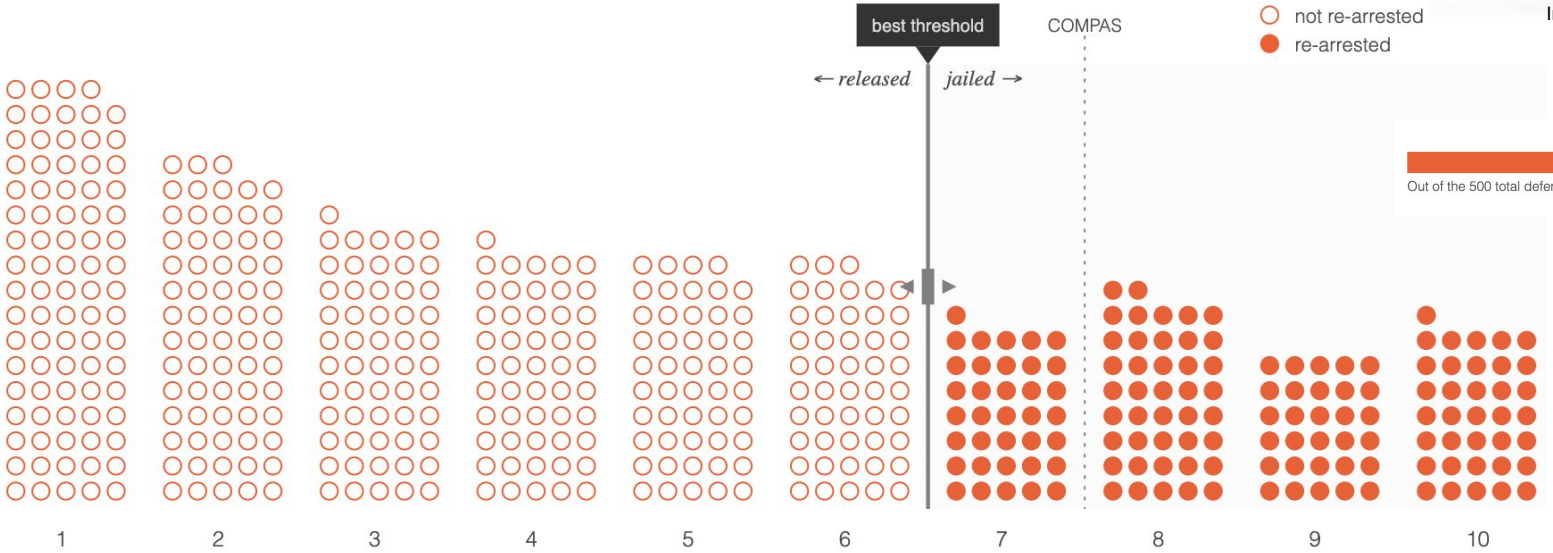
Because there are predictors for ethnicity in the data set - i.e. the data could be used to build an ethnicity-predictor.

MIT Technology Review: courtroom algorithm game

COMPAS - can we make it fair?

Definition
ERROR RATE: How many people were not jailed but still reoffended

Image courtesy Shutterstock



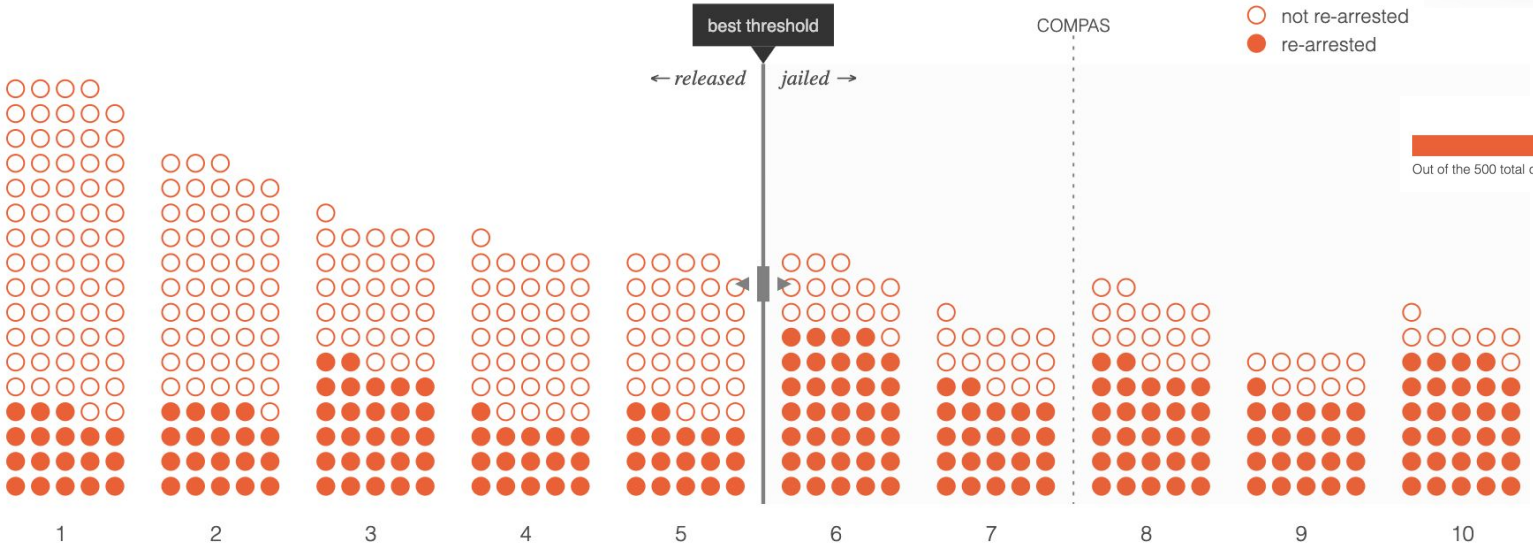
[MIT Technology Review: courtroom algorithm game]

COMPAS - can we make it fair?

Definition
ERROR RATE: How many people were not jailed but still reoffended

Can’t predict with 100% certainty how a person will act

Image courtesy Shutterstock



[MIT Technology Review: courtroom algorithm game]

COMPAS - can we make it fair?

Definition
ERROR RATE: How many people were not jailed but still reoffended

Complication: White and black people are jailed with different frequencies. Reason? Historical discrimination?

Consequence: Two groups arise in the model. We wish this wasn't the case - put now that it is, we have to deal with it.

WHAT IS FAIR?

1. Keep the error rates comparable between groups (wrongfully arrest the same number of black and white people)
2. Treat people with the same risk scores in the same way

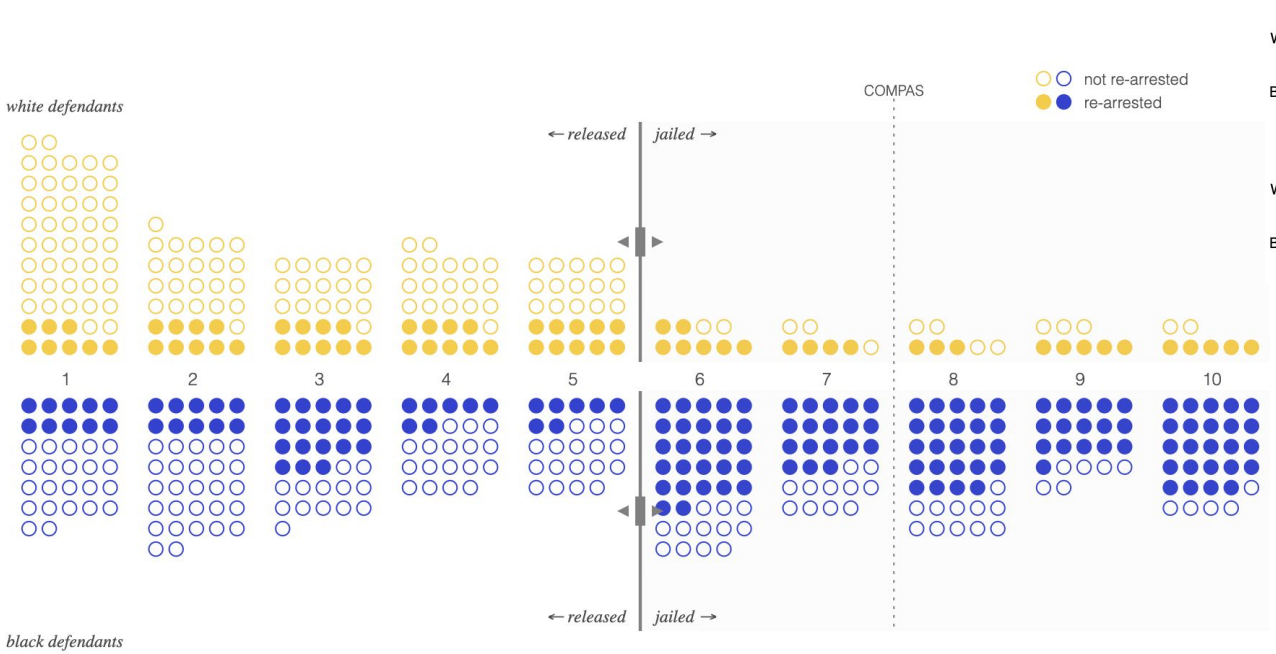
MIT Technology Review: courtroom algorithm game

COMPAS - can we make it fair?

FAIRNESS:

1. Keep the error rates comparable between groups (wrongfully arrest the same number of black and white people)
2. Treat people with the same risk scores in the same way

PROBLEM: The error rates are different - more black than white people are needlessly arrested!



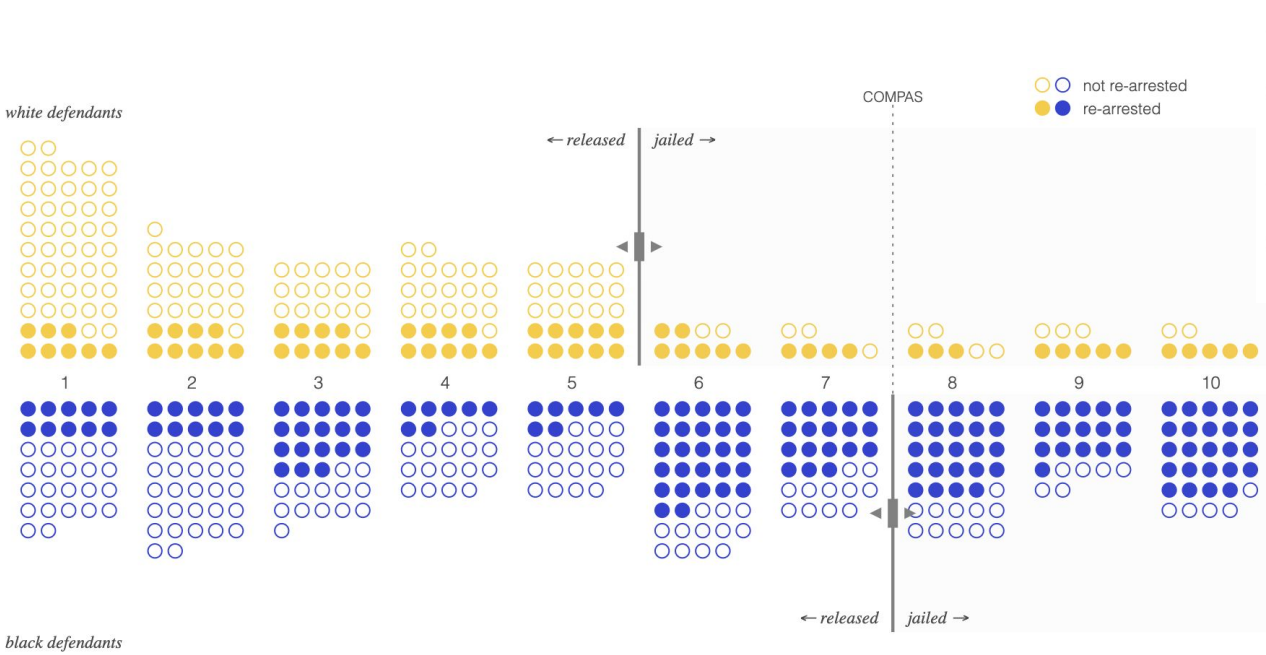
MIT Technology Review: courtroom algorithm game

COMPAS - can we make it fair?

FAIRNESS:

1. Keep the error rates comparable between groups (wrongfully arrest the same number of black and white people)
2. Treat people with the same risk scores in the same way

PROBLEM: Two people with the same risk score are treated differently based on skin colour



MIT Technology Review: courtroom algorithm game

There are at least 20 definitions of fairness, and when we choose one, we violate some aspect of the others.

Some fairness metrics examples:

Statistical Parity
Difference of the rate of favorable outcomes for the unprivileged group to the privileged group

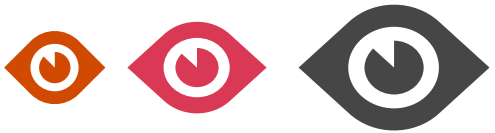
Disparate Impact
Ratio of the rate of favorable outcome for the unprivileged group to the privileged group.

Equal Opportunity
Difference of true positive rates between the unprivileged and the privileged groups

Average Odds
Average difference of false positive rate and true positive rate between unprivileged and privileged groups.

Group fairness

Perspective: All groups have the same underlying abilities



Individual fairness

Perspective: The observations reflect ability

fairware.cs.umass.edu/papers/Verma

Fairness is a societal construct

It cannot be left to developers and programmers to decide which fairness definition to use or which group to protect

Answering such questions requires a democratic process

Image courtesy Shutterstock

Finally: Not understanding

Finally: Not understanding

Let's take a step back and think about what machine learning *does*

Data

↓

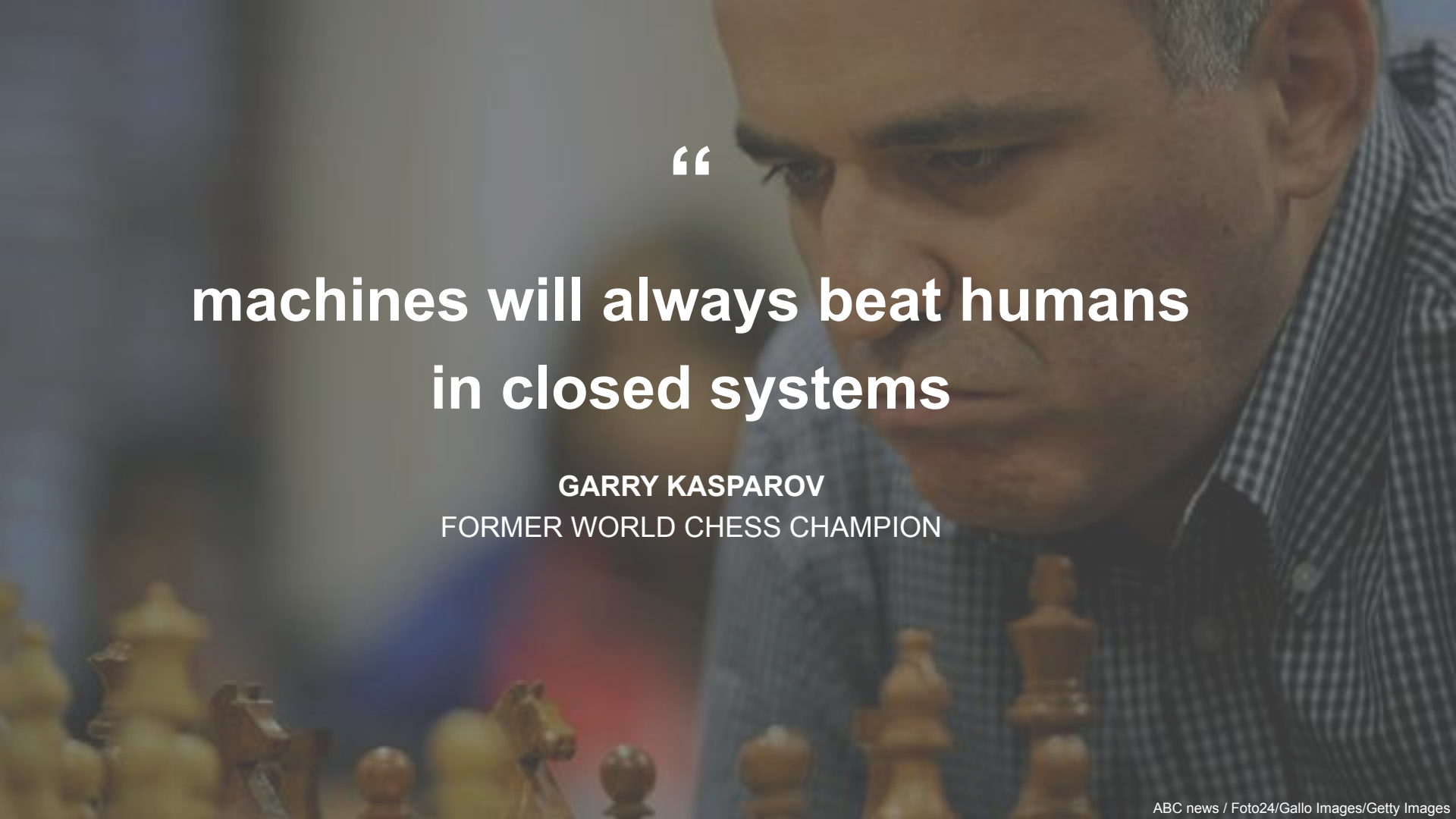
Model

↓

Prediction

Machine learning is about predicting the future **based on the past**

… so it's strictly speaking association more than prediction



“

machines will always beat humans in closed systems

GARRY KASPAROV

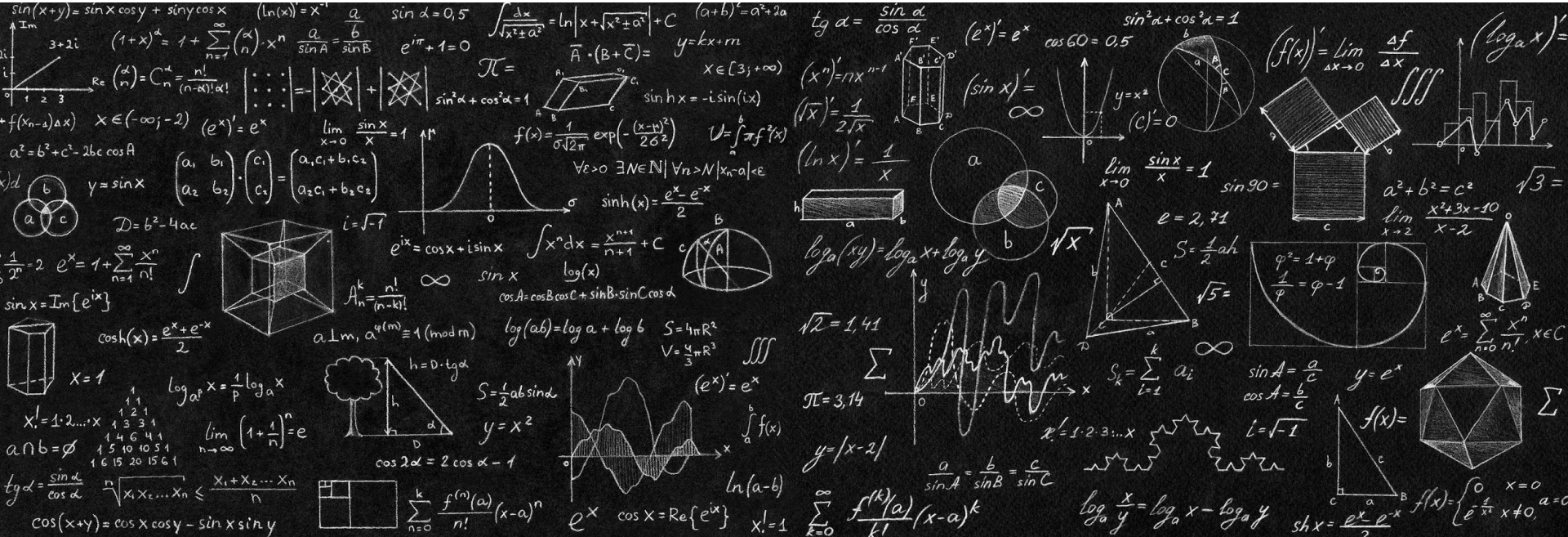
FORMER WORLD CHESS CHAMPION

ABC news / Foto24/Gallo Images/Getty Images

Observations:

- Reality can be described
- Machine learning can do this

Coincidence??



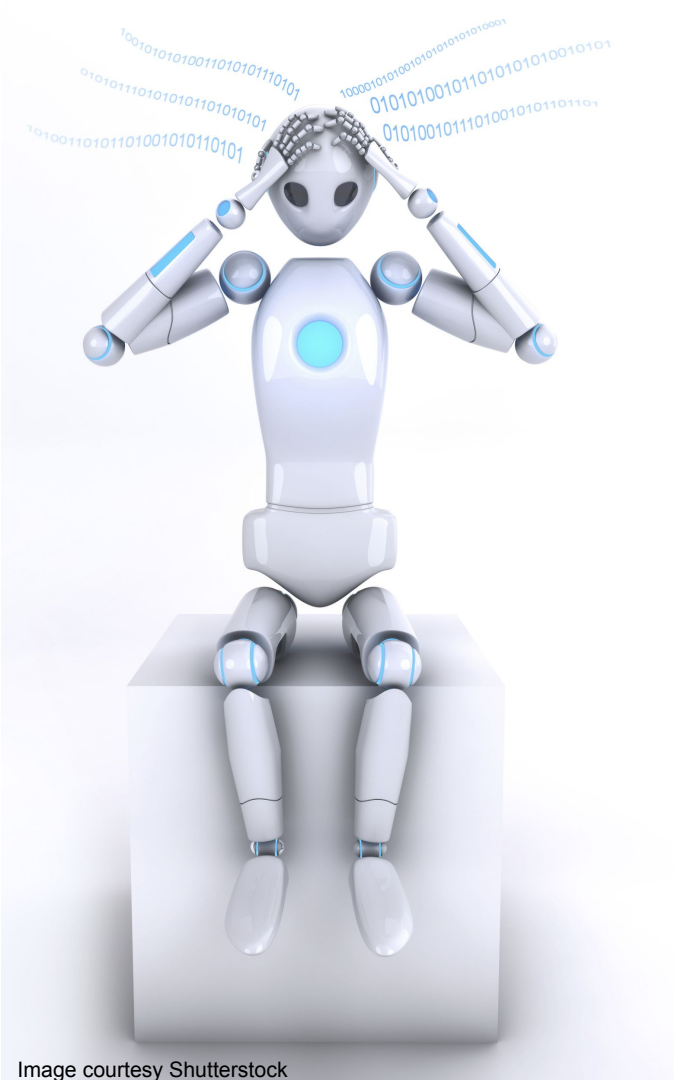


Image courtesy Shutterstock

Machine learning is correlation detection.

We still have to do the thinking

… and there is a huge effort to be done also in XAI

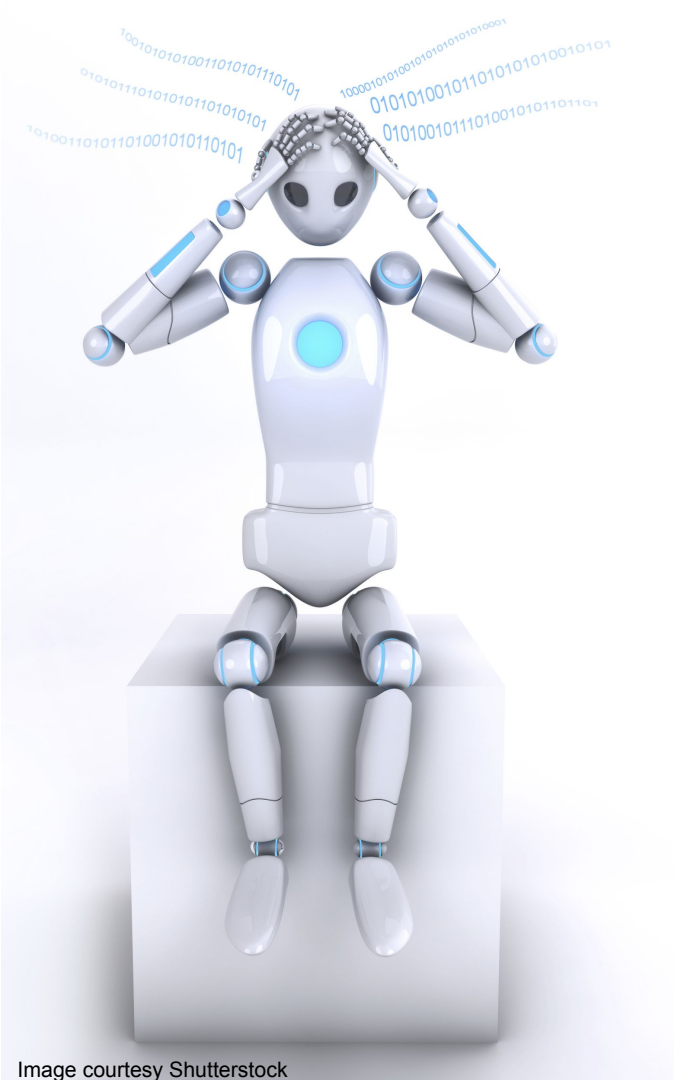


Image courtesy Shutterstock

Thank you!

inga.struemke@pwc.com