

Data Analysis and Machine Learning 4

Week 4: Machine Learning and Ethics

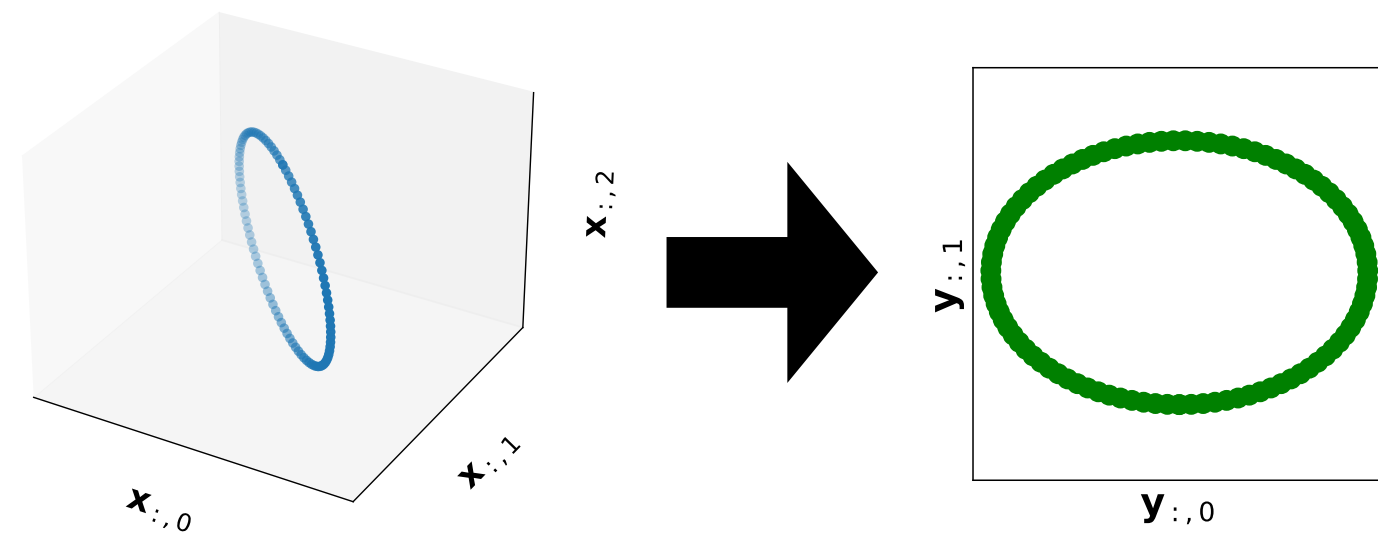
Elliot J. Crowley, 6th February 2023



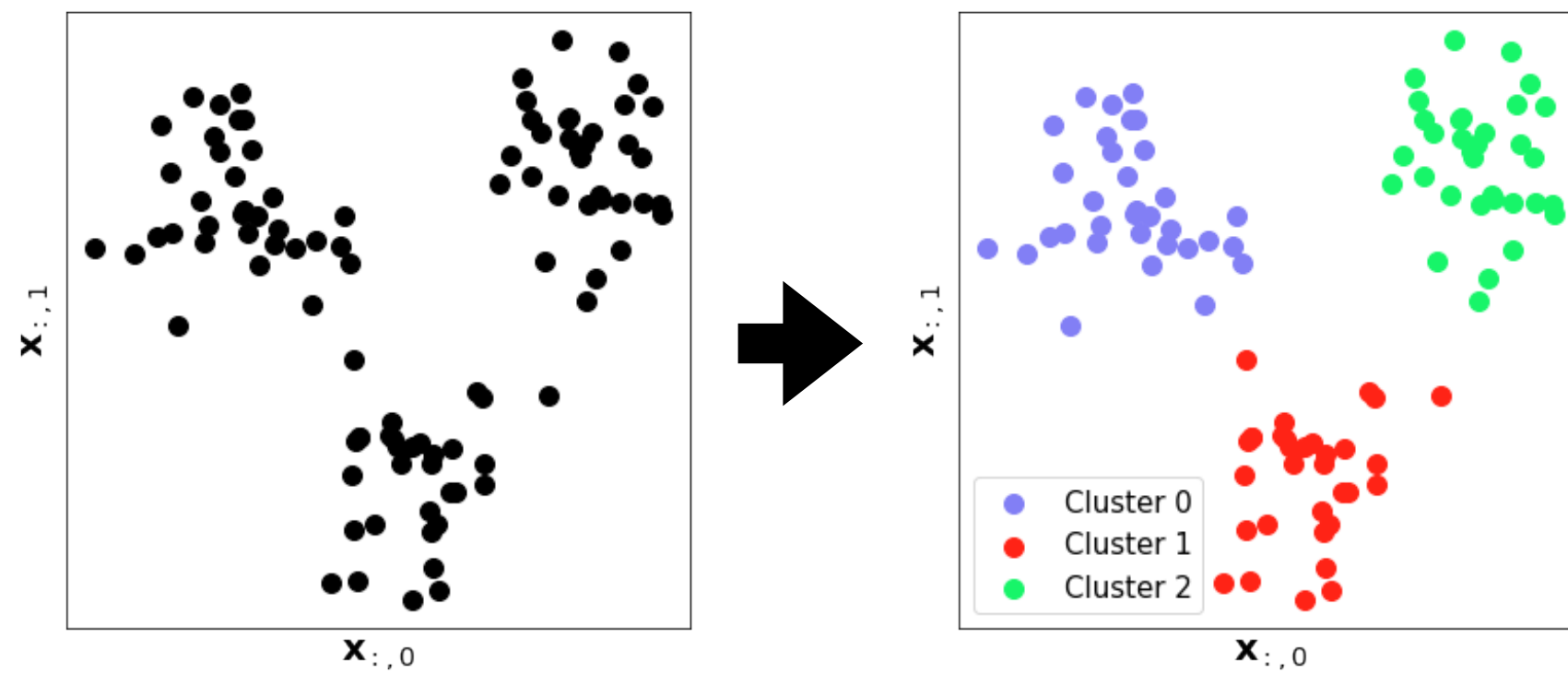
THE UNIVERSITY
of EDINBURGH

Recap

- We learned about principal component analysis (PCA)



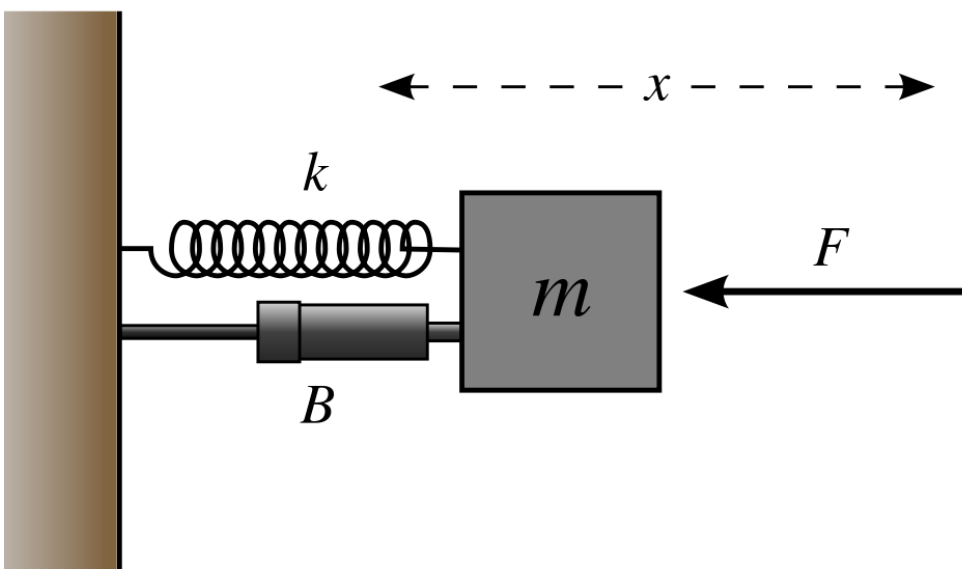
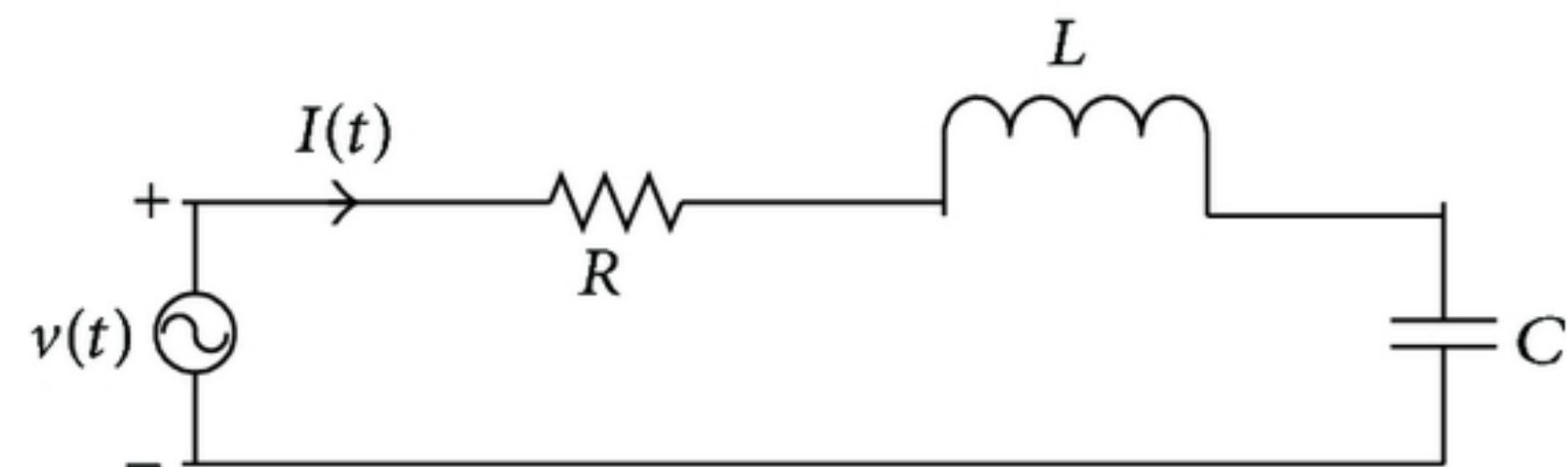
- We found out how to perform clustering using K-means



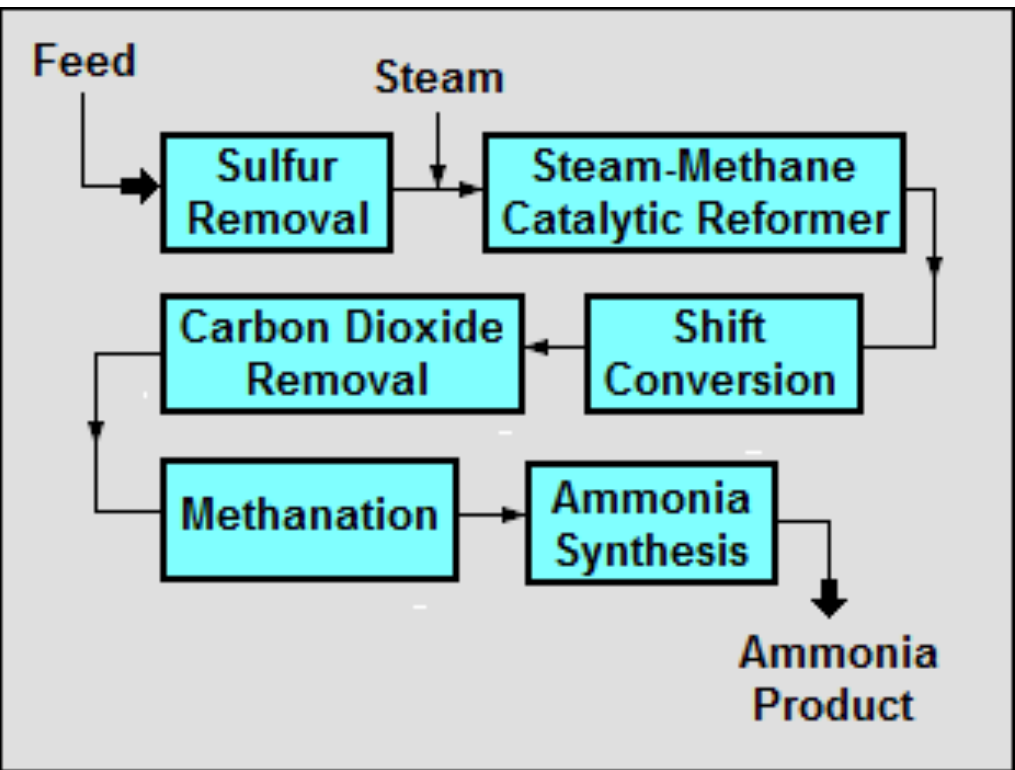
Machine Learning and Supervised Learning

Problem solving

Some problems can be solved analytically



Some problems can be solved by following instructions or rules



	No oxygen support (early COVID-19, but at high risk of progression)	Low-flow oxygen (COVID-19 pneumonia)	High-flow oxygen/CPAP/ mechanical ventilation (COVID-19 pneumonia)
Recommended	Neutralising monoclonal antibodies (See policy for more details) <ul style="list-style-type: none">♦ Aged 12 or over, and weight 40 kg or over, and♦ who are not in hospital	Corticosteroids (dexamethasone, or either hydrocortisone or prednisolone)	
		Tocilizumab (See policy for more details) If C-reactive protein is 75 mg/litre or more	Tocilizumab (See policy for more details) If within 48 hours of starting this level of support
		Baricitinib Adults	
		Low molecular weight heparin (standard prophylactic dose) Adults or young people, if within 14 hours of admission and no increased bleeding risk	

Problem solving

- Many real-world problems cannot be solved analytically or with rules
- However, we have access to more data than ever
- Can we leverage all this data to solve problems?

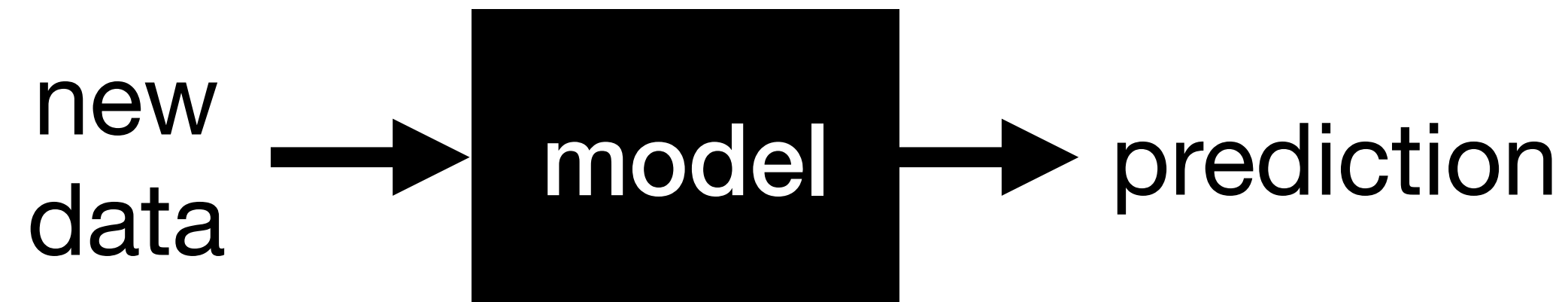
Machine Learning is...

“the study of algorithms that can learn from training data in order to make predictions on new data.”

Elliot J. Crowley

Machine Learning

- We want a model that takes in a new data point and outputs a prediction



- For the model to be accurate it must first learn from training data
- Often, models are parameterised functions and learning = finding the best parameters

Supervised Learning

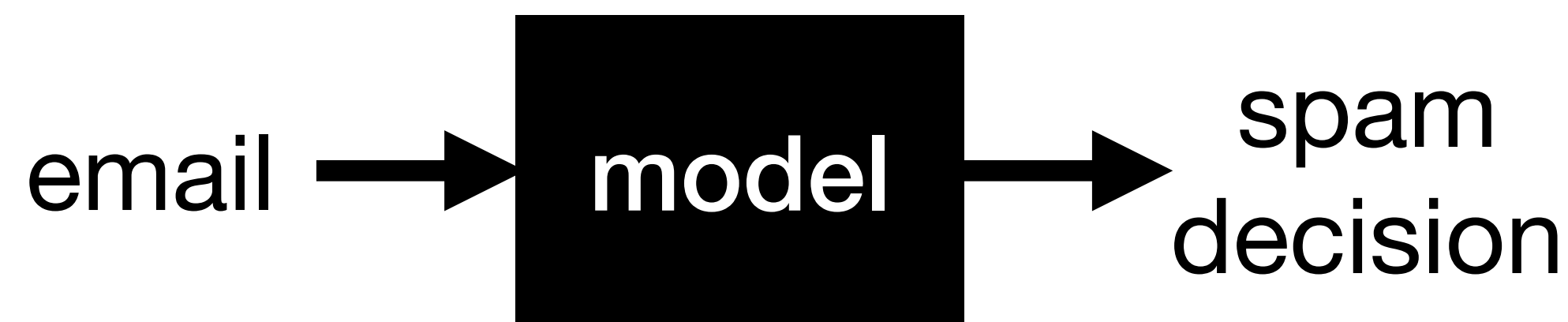
- In supervised learning, training data is labelled
- The label says what the **prediction** for that data point **should be**
- In unsupervised learning we do not have labels for our training data. Some people consider K-means and PCA to be unsupervised ML techniques
- Supervision can be seen as a spectrum e.g. we can have semi-supervised ML

We will only consider fully supervised machine learning for the remainder of the technical material on this course

Some use cases you see this lecture involve weaker supervision however :)

Example: Spam classification

- We want a model that takes in a new email \mathbf{x} and returns a prediction $y \in \mathbb{Z}_{<2}^+ = \{0,1\}$ where 0 is NOT SPAM and 1 is SPAM
- Our training data consists of emails that are labelled as spam or not spam



Dear Elliot J Crowley,

Paper#: 1356 Title: Prediction-Guided Distillation for Dense Object Detection

Congratulations on having your paper accepted to ECCV 2022. Follow the link below for instructions on formatting and submitting your final ECCV 2022 camera-ready files.

Camera-Ready Submission Instructions: <https://docs.google.com/document/d/1vRlxDLW7rNzgB5-fq-XiSBm5bmbjYuOeR3kjDf8zeg/>

Read through these instructions and follow them carefully to avoid any problems with your camera-ready paper submission.



THE UNIVERSITY
of EDINBURGH

We have just prevented a sign-in attempt on your **MyEd account** from another location, [CLICK HERE](#) to verify your profile.

Your prompt response regarding this matter is appreciated.

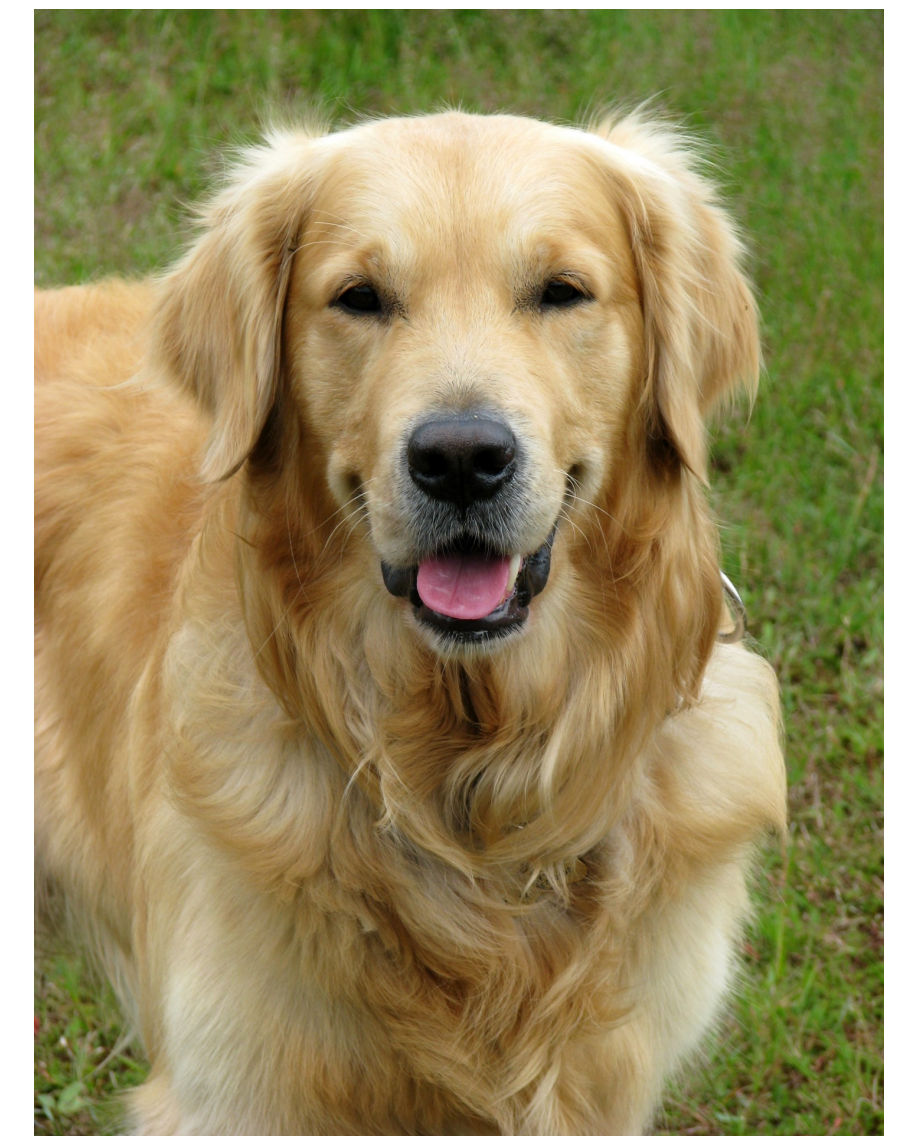
Sincerely

Management



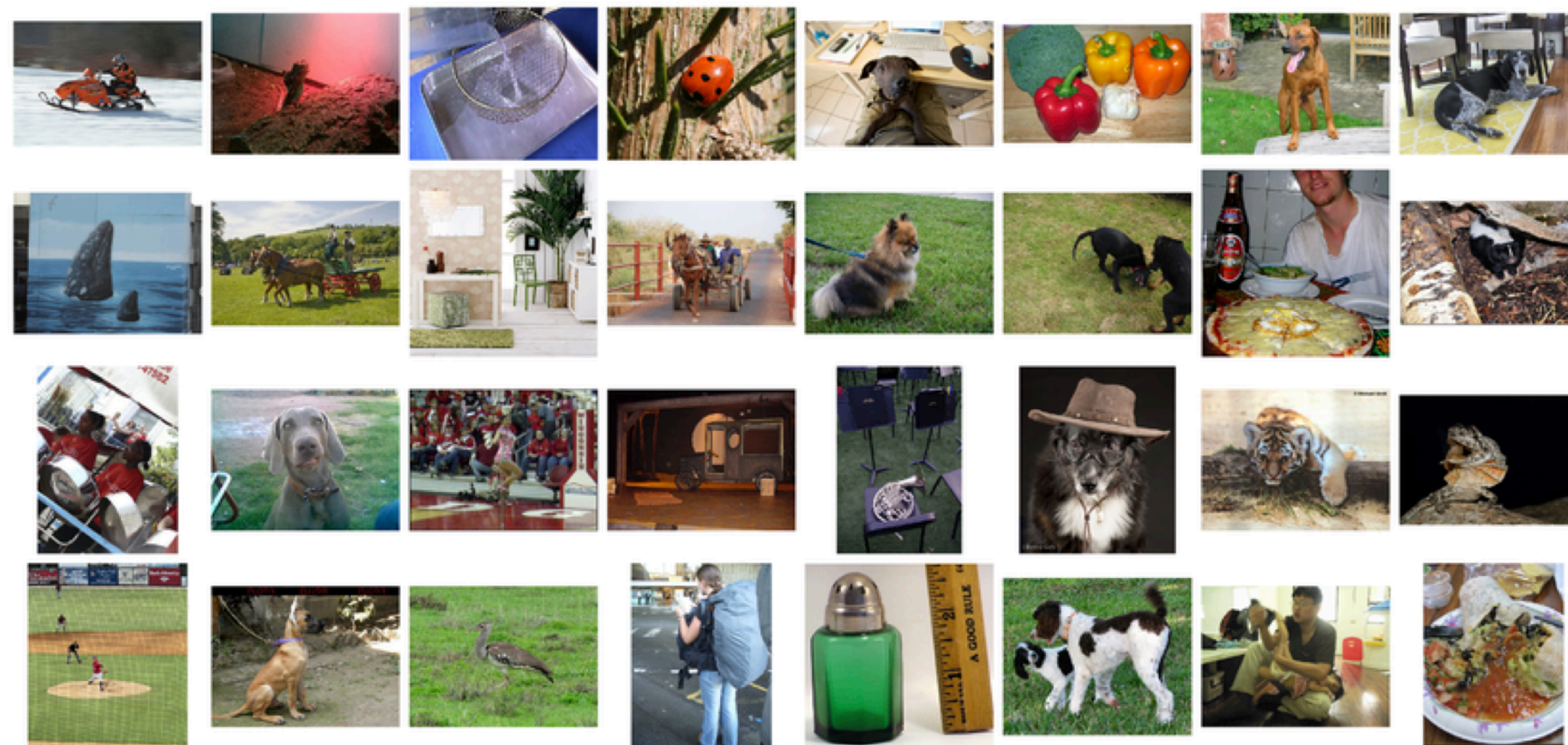
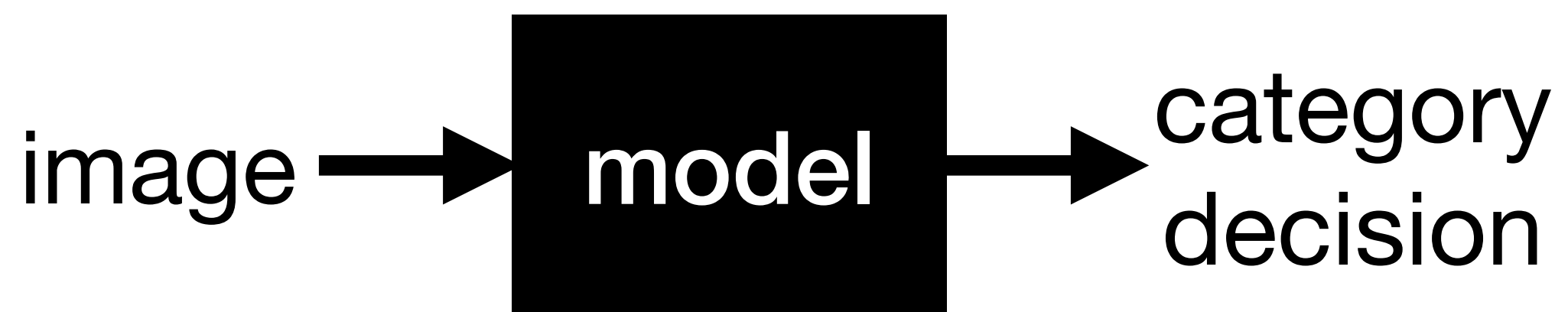
Example: Classifying cats and dogs

- We want a model that takes in a new image $\mathbf{x} \in \mathbb{R}^{H \times W \times C}$ and returns a prediction $y \in \mathbb{Z}_{<2}^+ = \{0,1\}$ where 0 is CAT and 1 is DOG
- Our training data consists of images of cats and dogs that are labelled 0/1



Example: Multiway classification

- We want a model that takes in a new image $\mathbf{x} \in \mathbb{R}^{H \times W \times C}$ and returns a prediction $y \in \mathbb{Z}_{<1000}^+ = \{0, 1, \dots, 999\}$ corresponding to 1 of 1000 categories
- Our training data consists of labelled images from those 1000 categories



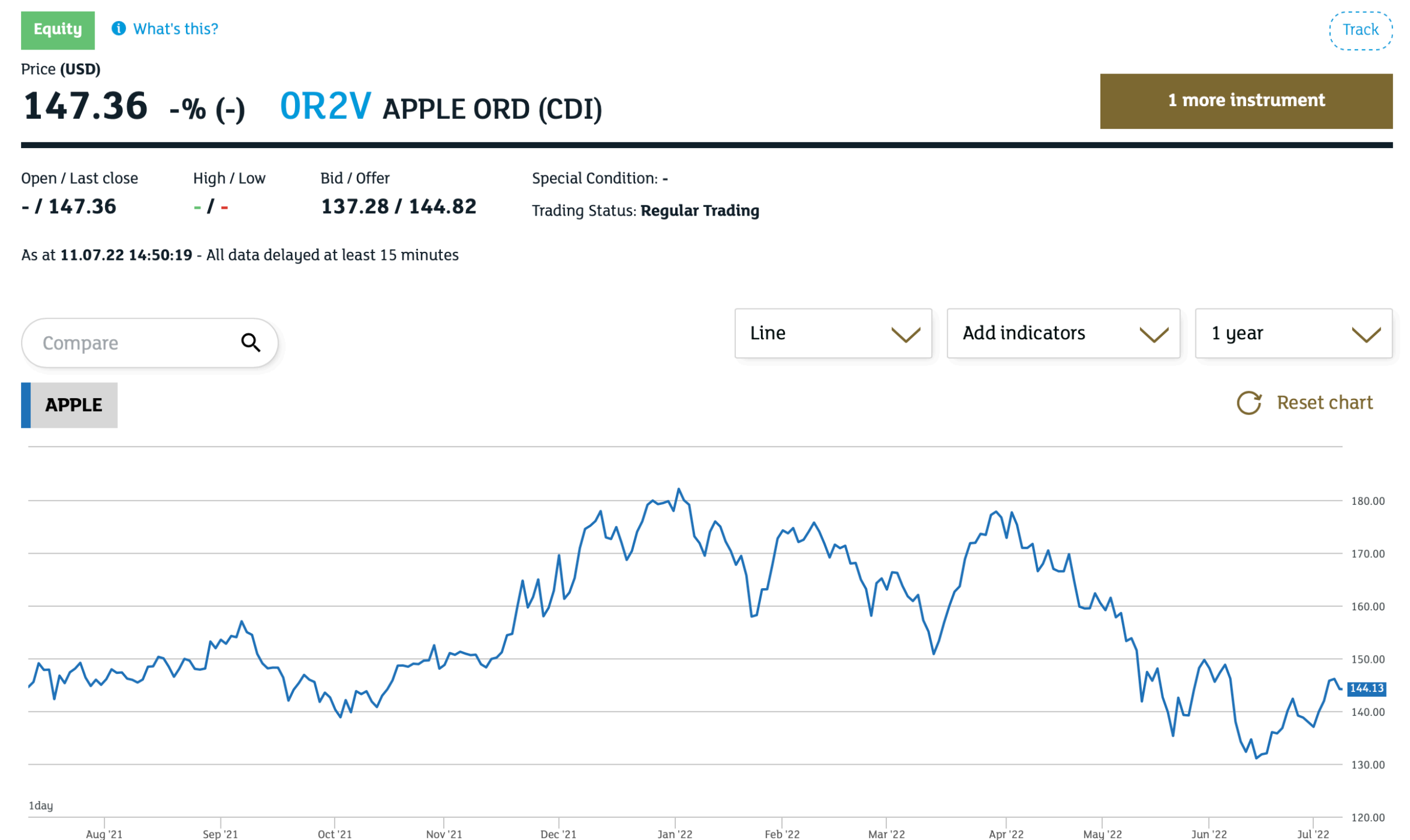
Example: Semantic Segmentation

- We want a model that takes in a new image \mathbf{x} and predicts a segmentation map \mathbf{y}
- Our training data consists of images and hand-labelled maps



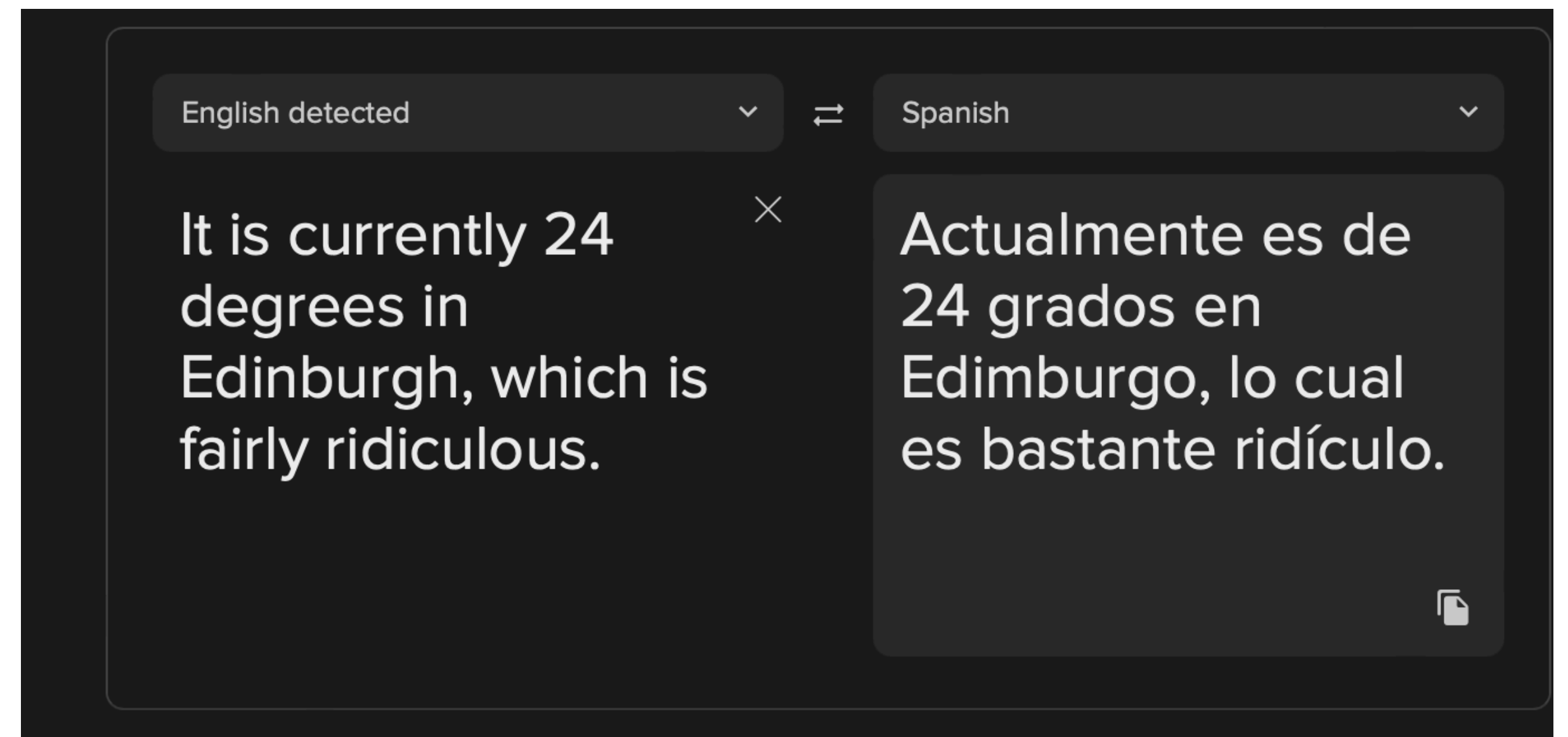
Example: Stock prediction

- We want a model that predicts stock price $y \in \mathbb{R}^1$ at a future date $x \in \mathbb{R}^1$
- Our training data consists of stock prices at previous dates



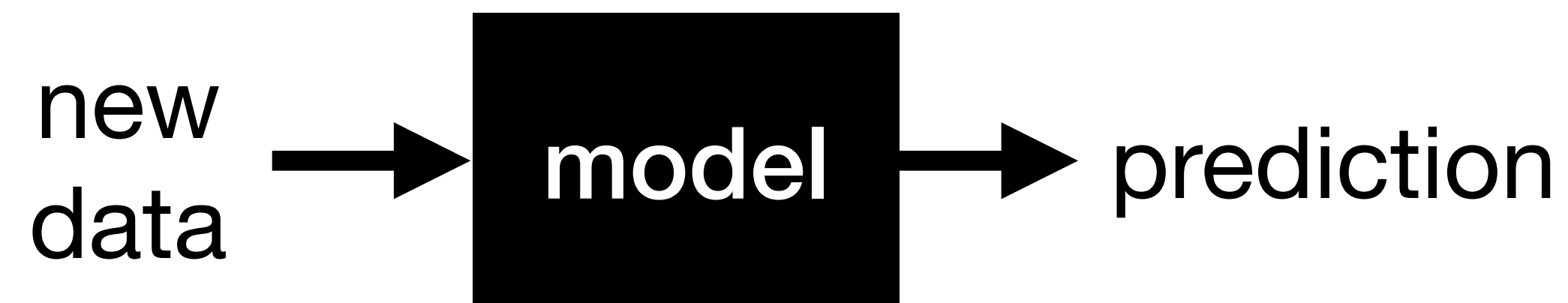
Example: Machine translation

- We want a model that translates some sentence in English \mathbf{x} to Spanish \mathbf{y}
- Our training data consists of English-Spanish sentence pairs



Machine Learning models

- Most of the time these models are mathematical functions
- You will find out what these models look like from the next lecture onwards
- Assume for now that given enough training data, they “work”




Ethical Issues

Be wary of hype

Robots could soon think like HUMANS: Scientists develop AI that can learn the basic common sense rules of the physical world – just like a baby

- AI can be taught 'intuitive physics' - common sense rules of how the world works
- Researchers at DeepMind trained an AI called PLATO with slides of a ball moving
- PLATO demonstrated learning and 'surprise' if a ball moved in an impossible way

By [JONATHAN CHADWICK FOR MAILONLINE](#) 
PUBLISHED: 16:11, 11 July 2022 | UPDATED: 17:05, 11 July 2022

Community

Debate over AI sentience marks a watershed moment



**Elon Musk**  @elonmusk · Sep 4, 2017
It begins ...

**The Verge**  @verge
Putin says the nation that leads in AI 'will be the ruler of the world' [theverge.com/2017/9/4/16251..](#)



**Elon Musk**  @elonmusk · [Follow](#)

China, Russia, soon all countries w strong computer science. Competition for AI superiority at national level most likely cause of WW3 imo.

10:33 AM · Sep 4, 2017 

 39.9K  Reply  Share

[Read 3.5K replies](#)

Artificial Intelligence Market Hits USD 35,870 Million By 2025: By Grand View Research, Inc.

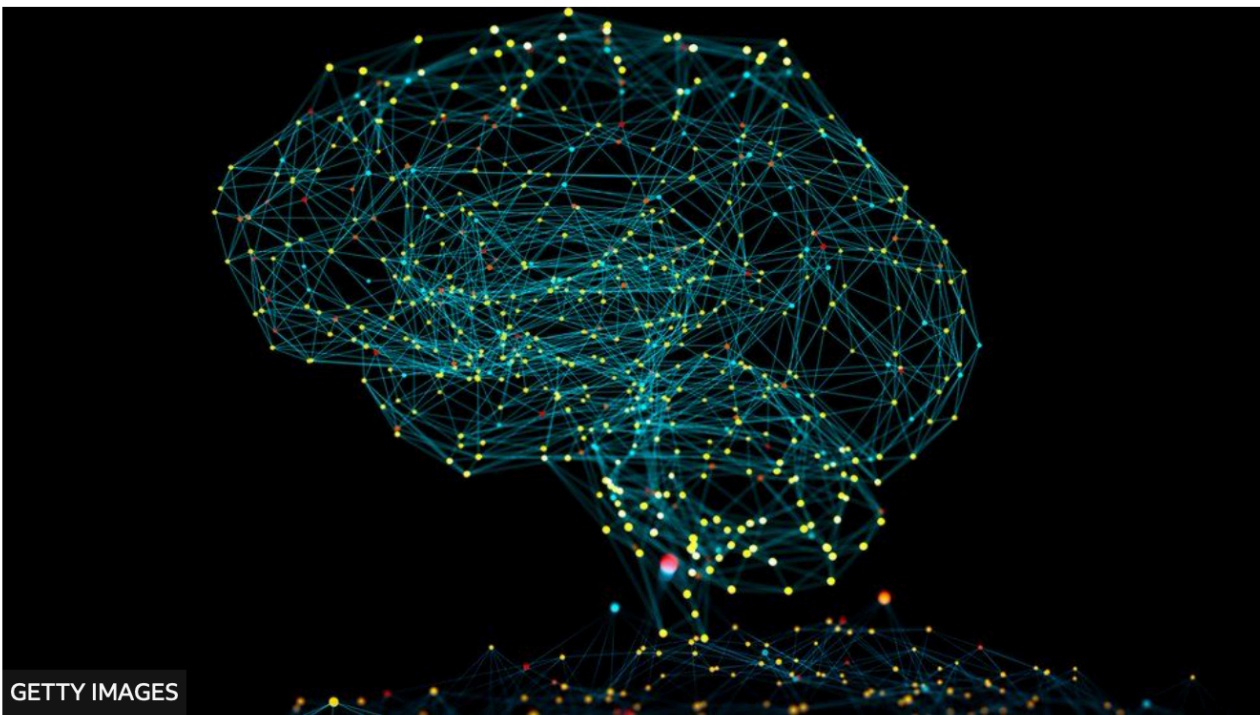
The global [artificial intelligence market](#) is expected to reach USD 35,870.0 million by 2025 from its direct revenue sources, growing at a CAGR of 57.2% from 2017 to 2025, whereas it is expected to garner around USD 58,975.4 million by 2025 from its enabled revenue arenas, according to a new report by Grand View Research, Inc.

Artificial Intelligence (AI) is considered to be the next stupendous technological development, alike past developments such as the revolution of industries, the computer era, and the emergence of smartphone technology. The North American region is

Google engineer says Lamda AI system may have its own feelings

By **Chris Vallance**
Technology reporter

🕒 13 June

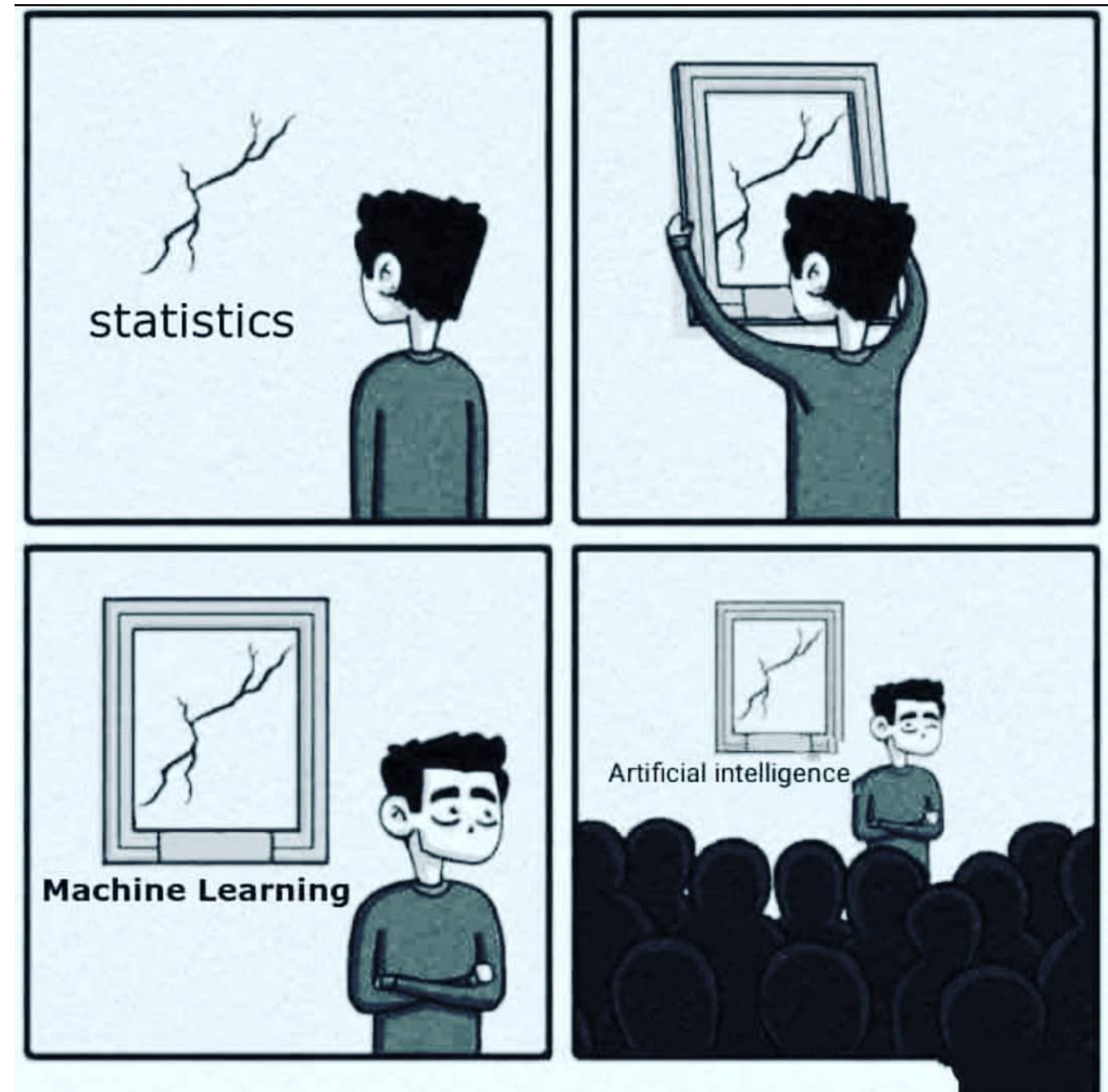


A Google engineer says one of the firm's artificial intelligence (AI) systems might have its own feelings and says its "wants" should be respected.

Reality



<https://xkcd.com/1838/>




<https://www.pinterest.co.uk/pin/591449363544632526/>

Reality

Analysis | [Open Access](#) | [Published: 15 March 2021](#)

Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans

[Michael Roberts](#) , [Derek Driggs](#), [Matthew Thorpe](#), [Julian Gilbey](#), [Michael Yeung](#), [Stephan Ursprung](#), [Angelica I. Aviles-Rivero](#), [Christian Etmann](#), [Cathal McCague](#), [Lucian Beer](#), [Jonathan R. Weir-McCall](#), [Zhongzhao Teng](#), [Effrossyni Gkrania-Klotsas](#), [AIX-COVNET](#), [James H. F. Rudd](#), [Evis Sala](#) & [Carola-Bibiane Schönlieb](#)

[Nature Machine Intelligence](#) **3**, 199–217 (2021) | [Cite this article](#)

in EMBASE and MEDLINE in this timeframe are considered. Our search identified 2,212 studies, of which 415 were included after initial screening and, after quality screening, 62 studies were included in this systematic review. Our review finds that none of the models identified are of potential clinical use due to methodological flaws and/or underlying biases.

<https://www.nature.com/articles/s42256-021-00307-0>



<https://me.me/i/me-wondering-why-my-ai-is-classifying-my-cat-as-67127de0453d4583b9e4d88305fa4779>

ML can do amazing things



A photo of a Corgi dog riding a bike in Times Square. It is wearing sunglasses and a beach hat.

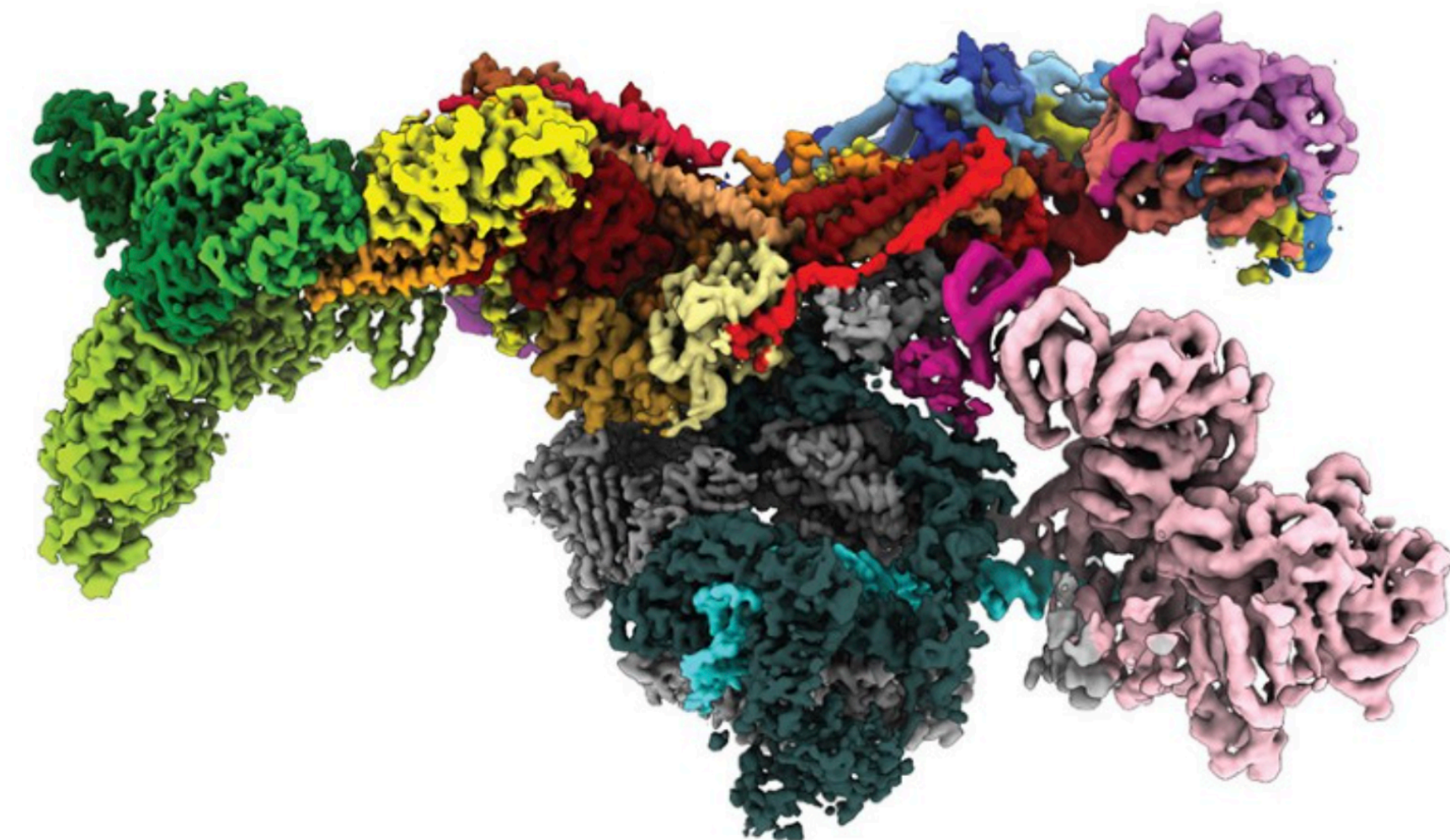
<https://imagen.research.google>

NEWS | 22 July 2021

DeepMind's AI predicts structures for a vast trove of proteins

AlphaFold neural network produced a 'totally transformative' database of more than 350,000 structures from *Homo sapiens* and 20 model organisms.

[Ewen Callaway](#)



<https://www.nature.com/articles/d41586-021-02025-4>

But things can go wrong

- Machine learning models are **dumb**
- **They are not human!**
- They just learn to map inputs to outputs
- They don't care what the data is
- They don't care where you got the data
- They don't care what the task is



Data privacy

Facebook sued for 'losing control' of users' data

🕒 9 February 2021



Facebook-Cambridge Analytica scandal

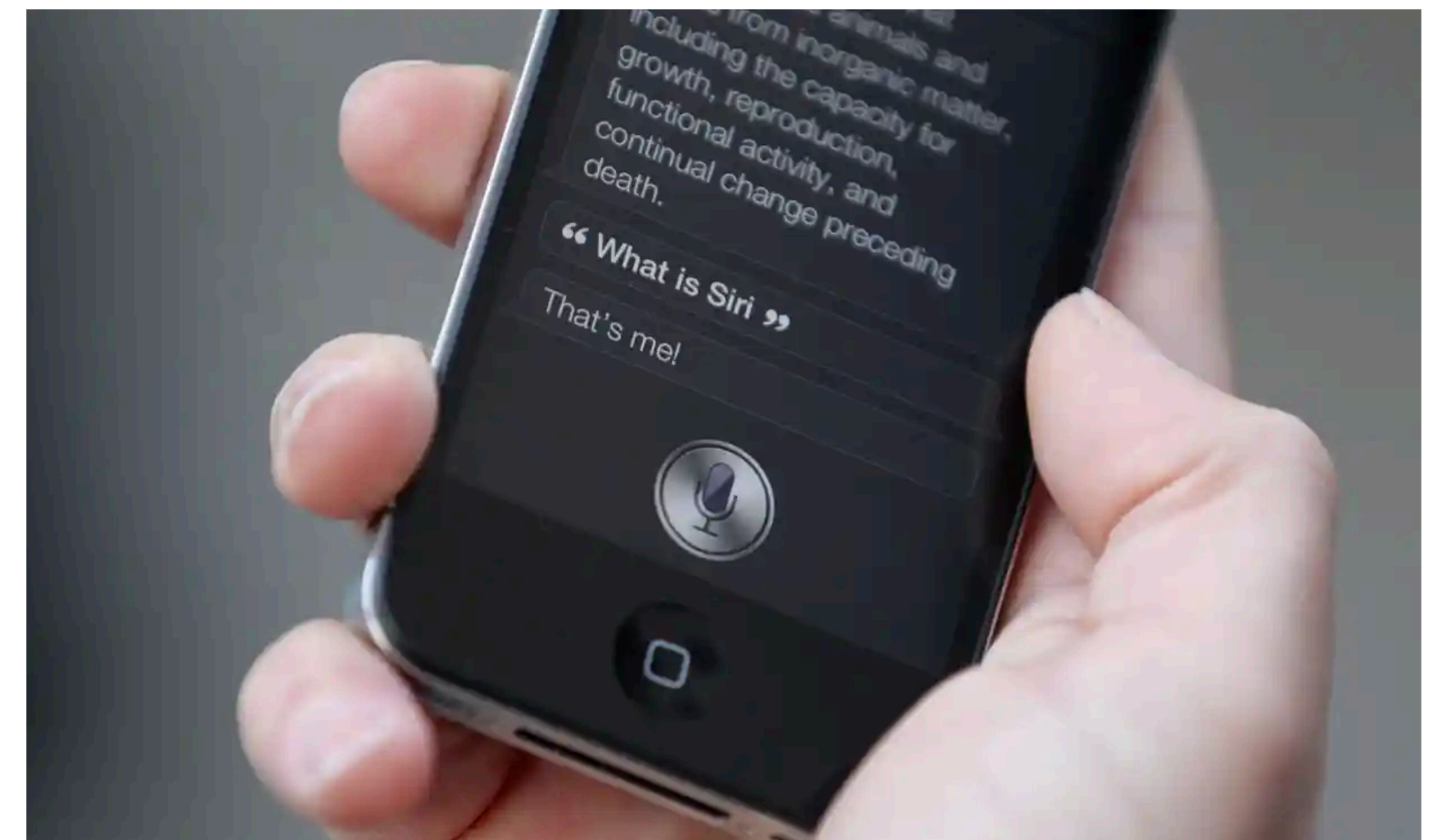


Facebook is being sued for "losing control" of the data of about a million users in England and Wales.

<https://www.bbc.co.uk/news/technology-55998588>

Apple contractors 'regularly hear confidential details' on Siri recordings

Workers hear drug deals, medical details and people having sex, says whistleblower



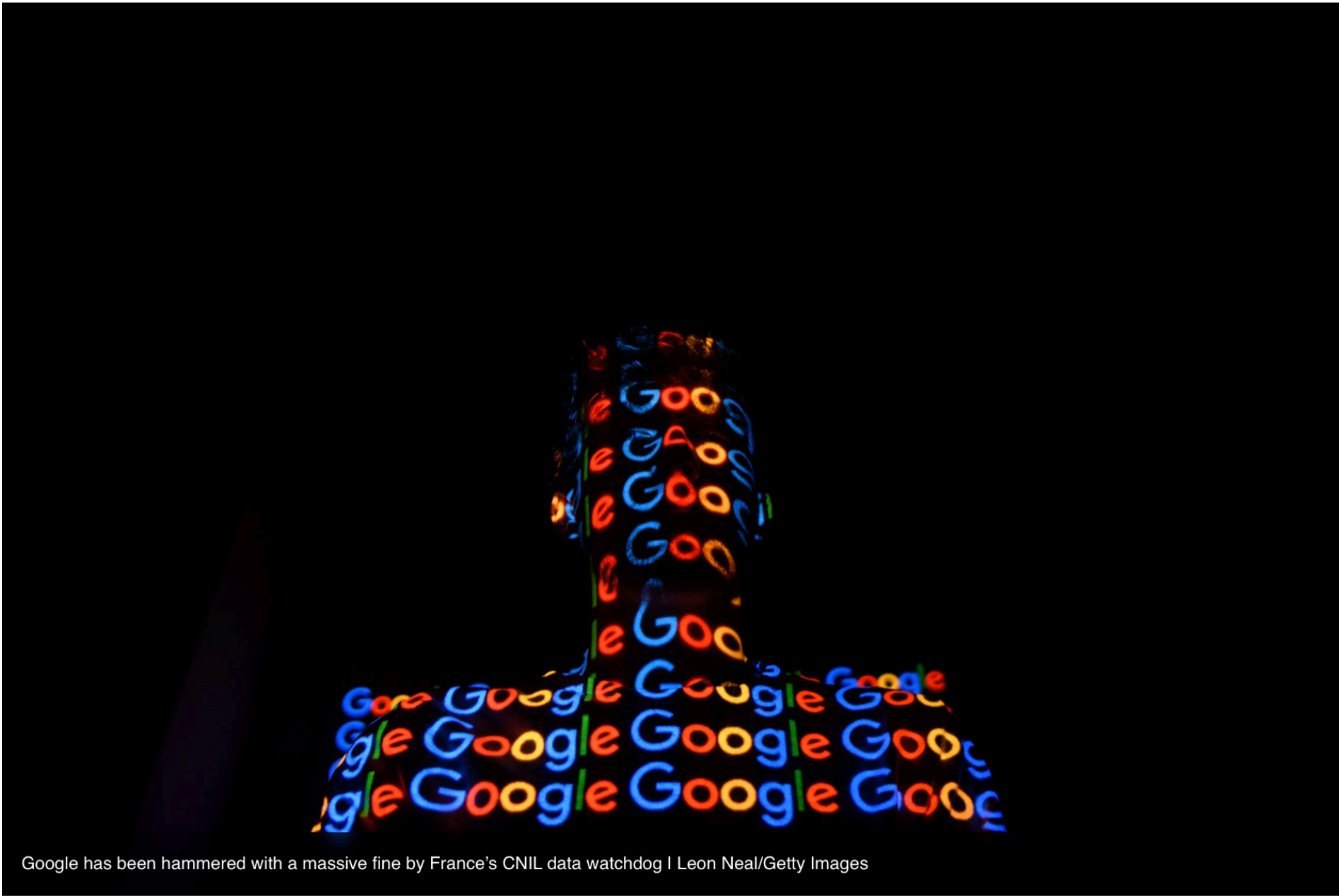
📷 Workers heard the information when or providing quality control for Apple's Siri voice assistant. Photograph: Oli Scarff/Getty Images

<https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings>

GDPR

Google fine launches new era in privacy enforcement

The search giant is the first big tech company to feel the full brunt of GDPR enforcement. It won't be the last.




Fines and notices [\[edit \]](#)

Date ↕	Organisation ↕	Amount ▼	Issued by ↕	Reason(s)
2021-06-16	Amazon Europe Core Srl	€746,000,000	Luxembourg (CNPD)	The largest fine for violating GDPR to date. ^[72] ^[73]
2021-09-02	WhatsApp Ireland Ltd	€225 M	Ireland	^[74]
2019-07-08	British Airways	£183,000,000	UK (ICO)	Use of poor security arrangements that resulted in a 2018 web skimming attack affecting 500,000 consumers. ^[24] ^[25] ^[26] Was later reduced to £20 million ^[27]
2020-12-10	Google LLC	€60,000,000	France (CNIL)	Deposit of cookies without obtaining consent, lack of information provided to users and defective "opposition" mechanism ^[49]
2019-01-21	Google LLC	€50,000,000	France (CNIL)	Insufficient transparency, control, and consent over the processing of personal data for the purposes of behavioural advertising . ^[4] ^[5]
2020-12-10	Google Ireland Limited	€40,000,000	France (CNIL)	Deposit of cookies without obtaining consent, lack of information provided to users and defective "opposition" mechanism ^[49]
2020-10-01	H&M	€35,300,000	Germany (HmbBfDI)	Illegal surveillance of several hundred employees ^[47]

Data bias

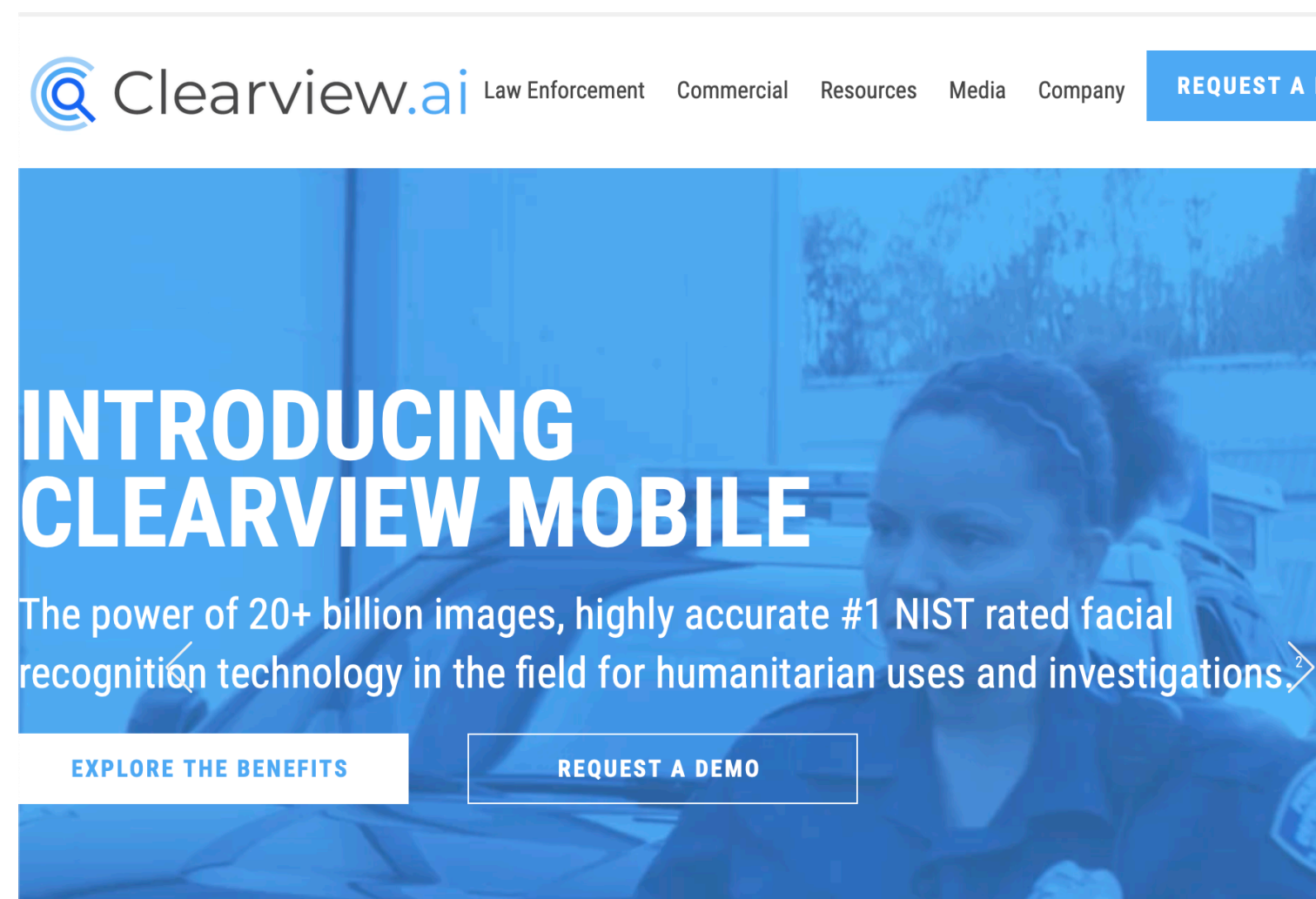
- A machine learning model learns from a training set
- The data in that training set may not be representative of the world
- It may contain biases e.g. reinforce antiquated stereotypes
- Biased data gives a biased model

A green Muppet character with a speech bubble pointing to it.

Remember that rubbish
in means rubbish out!

Applications

- ML models can be used to spread misinformation
- ML models can be used for surveillance
- ML models can be used to kill people



Mohsen Fakhrizadeh: 'Machine-gun with AI' used to kill Iran scientist

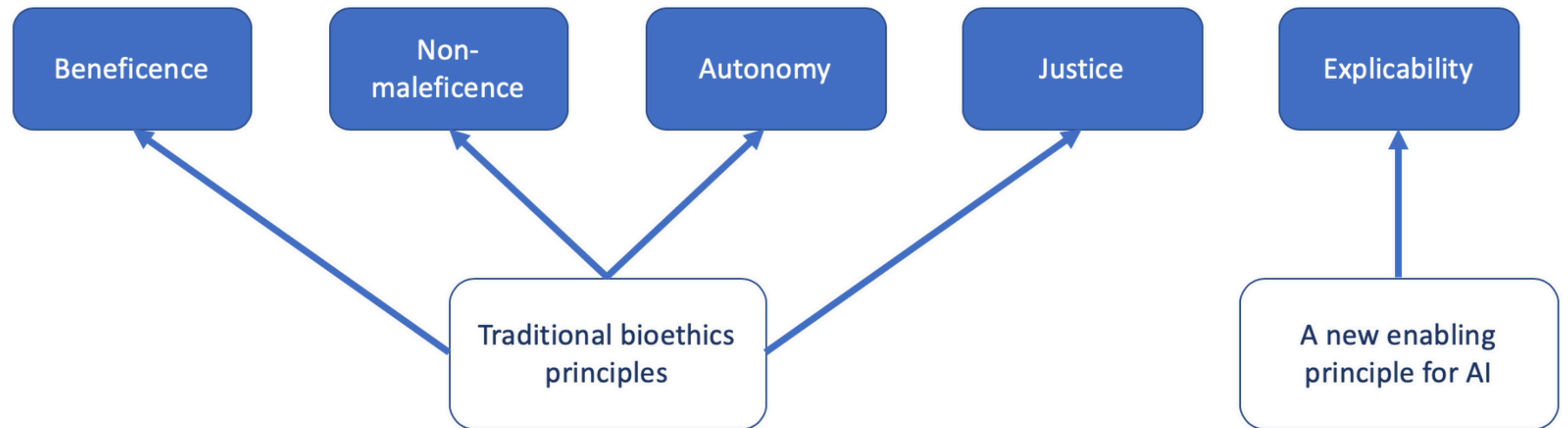
7 December 2020



REUTERS

The Iranian authorities have put out conflicting accounts of how the scientist was killed

Machine learning ethics



**A Unified Framework of Five Principles
for AI in Society**

by Luciano Floridi and Josh Cowls

Published on Jul 01, 2019

<https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/8>

Recruitment tools

- A machine learning model trained on predominantly male applicants
- Penalised CVs that included the word “women”

About 55% of US human resources managers said that AI would play a role in recruitment within the next five years, according to a survey by software firm CareerBuilder.

Amazon scrapped 'sexist AI' tool

🕒 10 October 2018

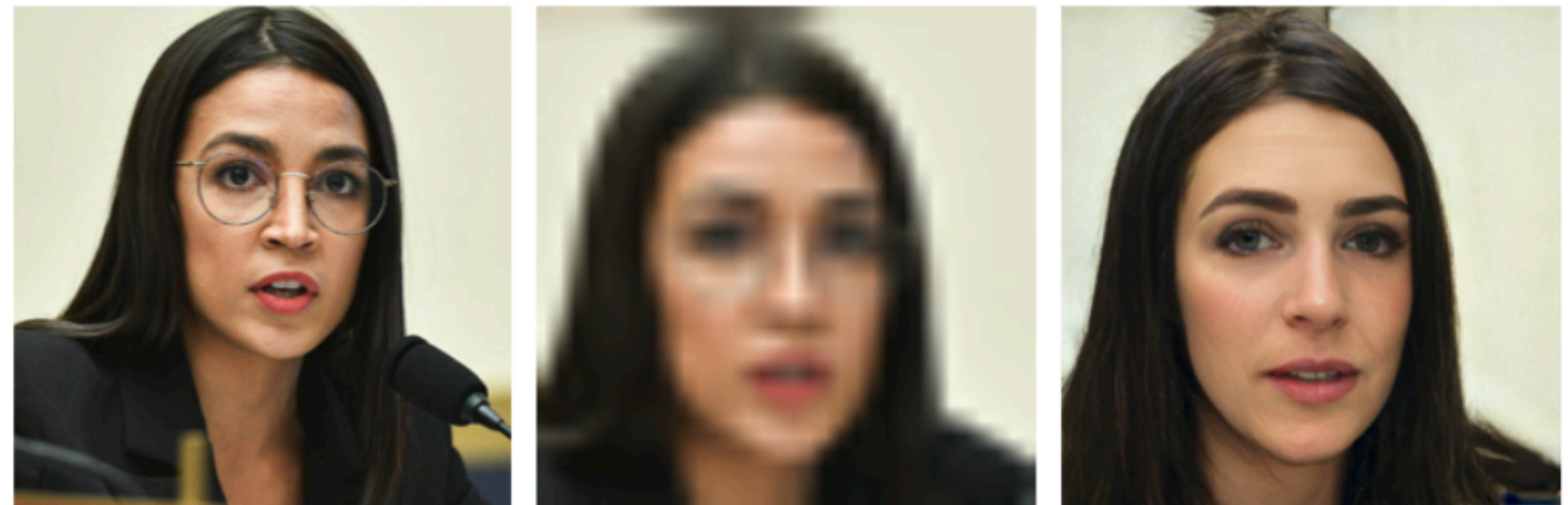
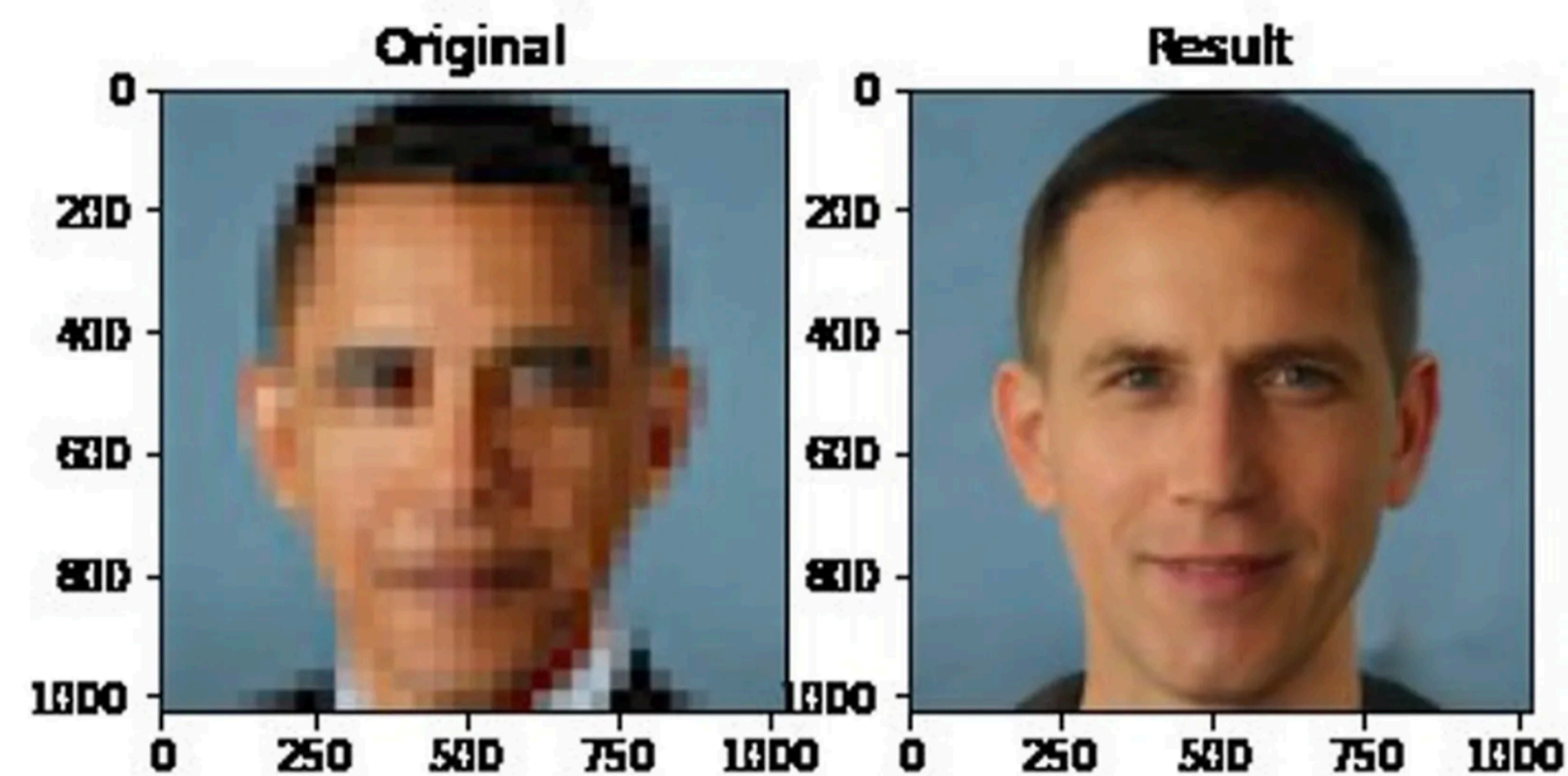


| The algorithm repeated bias towards men, reflected in the technology industry

<https://www.bbc.co.uk/news/technology-45809919>

Generative models

- PULSE depixelizer generates high-res images of faces from low-res inputs
- It tends to generate white faces
- This could be because its training data is dominated by white faces



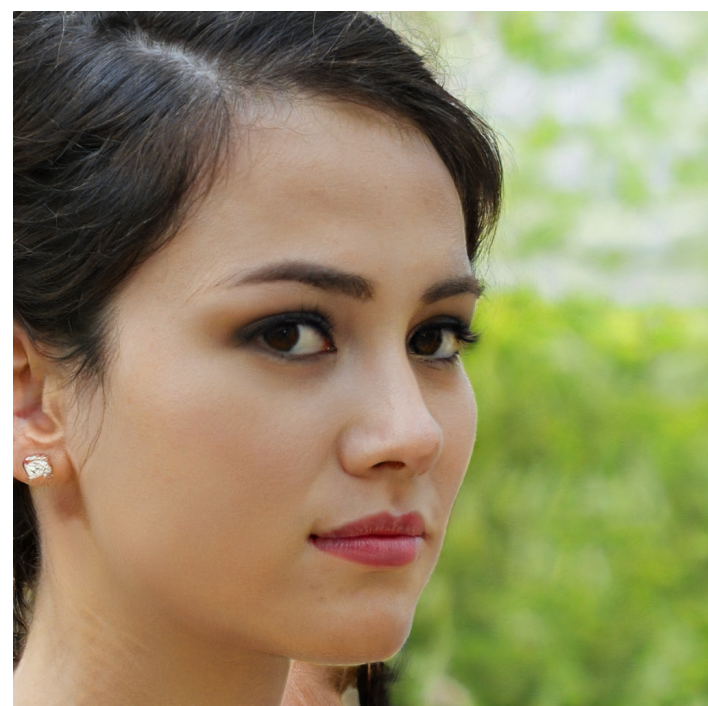
Auto-grading

- A level results in England and Wales were (initially) algorithmically generated
- This disproportionately benefited privately educated students
- High performing students at underperforming schools lost out
- Weight was placed on a school's historical importance



Deepfakes

- ML models that generate realistic faces in photos and videos
- They have be used to create hoaxes and revenge porn



Advertising

- Google and Facebook (/Meta) control most of the world's digital advertising
- They have both invested heavily in machine learning
- An ML model can use **your** internet activity to profile **you!**
- This will allow better targeting of adverts



GOOGLE ADS

Putting machine learning into the hands of every advertiser

Jul 10, 2018 · 4 min read



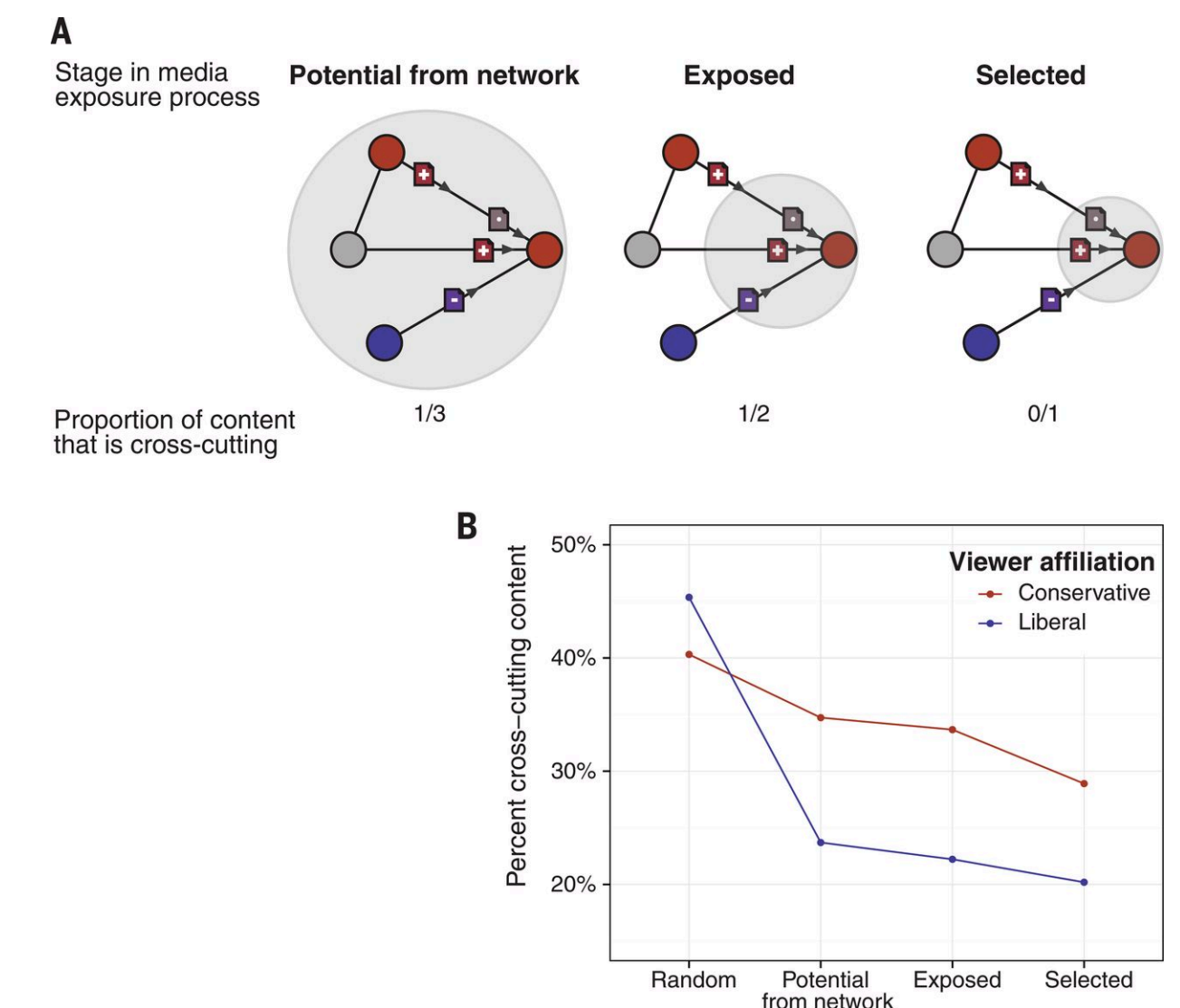
Jerry Dischler
Vice President / General
Manager, Ads

Share



Recommender systems

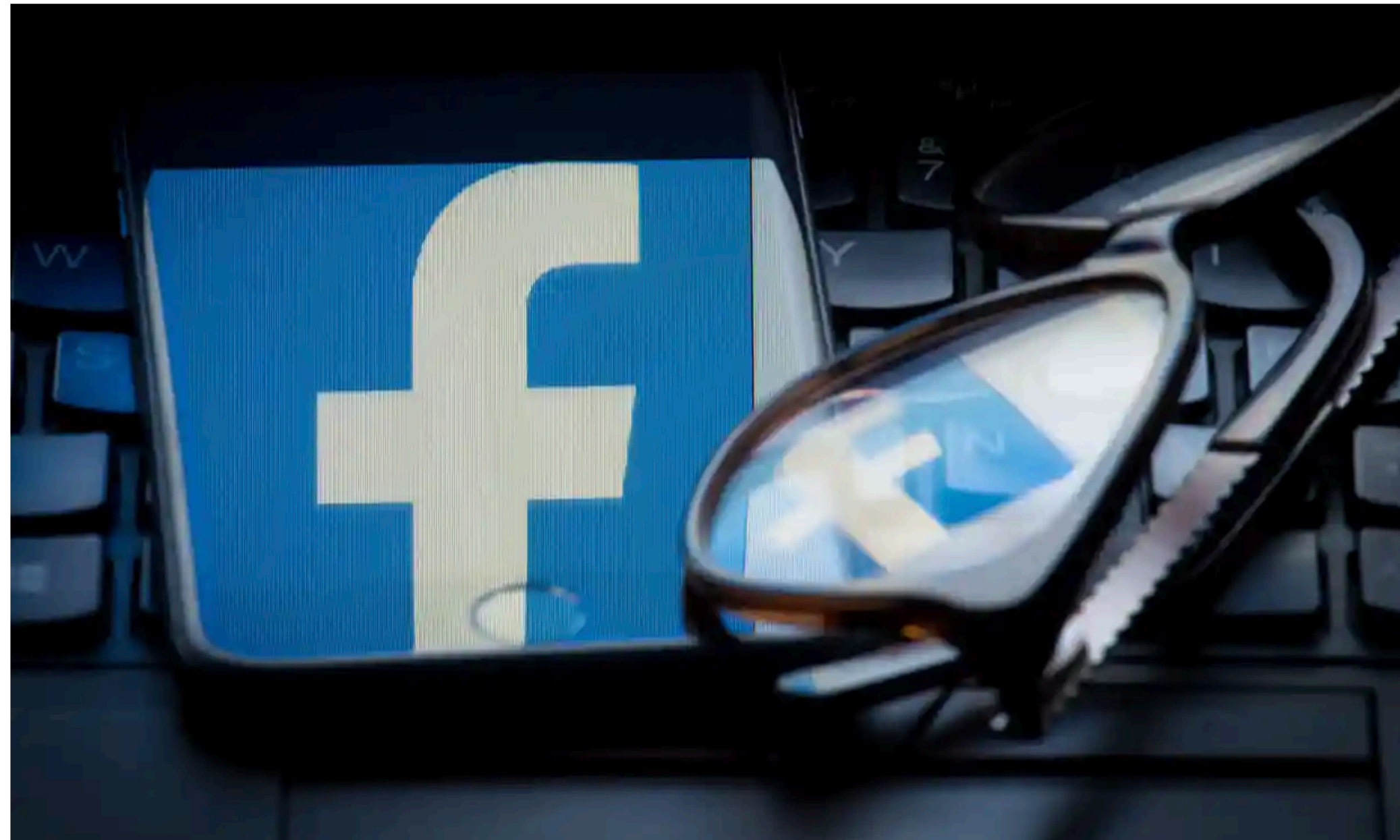
- Lots of people get their news from social media
- ML models can decide which content to present to users
- Users are presented with content that aligns with their views
- Is this ideal?
- Could this amplify radical views?



Cambridge Analytica

The Cambridge Analytica files: the story so far

What is the company accused of, how is Facebook involved and what is the Brexit link?



Facebook ran adverts on Sunday in several UK and US newspapers apologising for the data breach. Photograph: Dominic Lipinski/PA

What are the allegations against Cambridge Analytica?

The data analytics firm used personal information harvested from more than 50 million Facebook profiles without permission to build a system that could target US voters with personalised political advertisements based on their psychological profile, according to Christopher Wylie, a former Cambridge Analytica contractor who helped build the algorithm. Employees of Cambridge Analytica, including the suspended CEO Alexander Nix, were also filmed boasting of using manufactured sex scandals, fake news and dirty tricks to swing elections around the world.

Lighting the fuse?

‘Storm the Capitol’: Violence organised on social media as warnings of far-right post-election went unheard

Federal law enforcement has warned for years about far-right threats. Trump supporters openly planned armed insurrection for weeks, writes **Alex Woodward**

Friday 08 January 2021 21:31 • [Comments](#)



Dan Milmo *Global technology correspondent*

Mon 6 Dec 2021 17.03 GMT



Rohingya sue Facebook for £150bn over Myanmar genocide

Victims in US and UK legal action accuse social media firm of failing to prevent incitement of violence



Residents of the Rohingya refugee camp in Cox's Bazar, Bangladesh. Photograph: Tanbirul Miraj Ripon/EPA

Facebook's negligence facilitated the [genocide of Rohingya Muslims in Myanmar](#) after the social media network's algorithms amplified hate speech and the platform failed to take down inflammatory posts, according to legal action launched in the US and the UK.

<https://www.theguardian.com/technology/2021/dec/06/rohingya-sue-facebook-myanmar-genocide-us-uk-legal-action-social-media-violence>

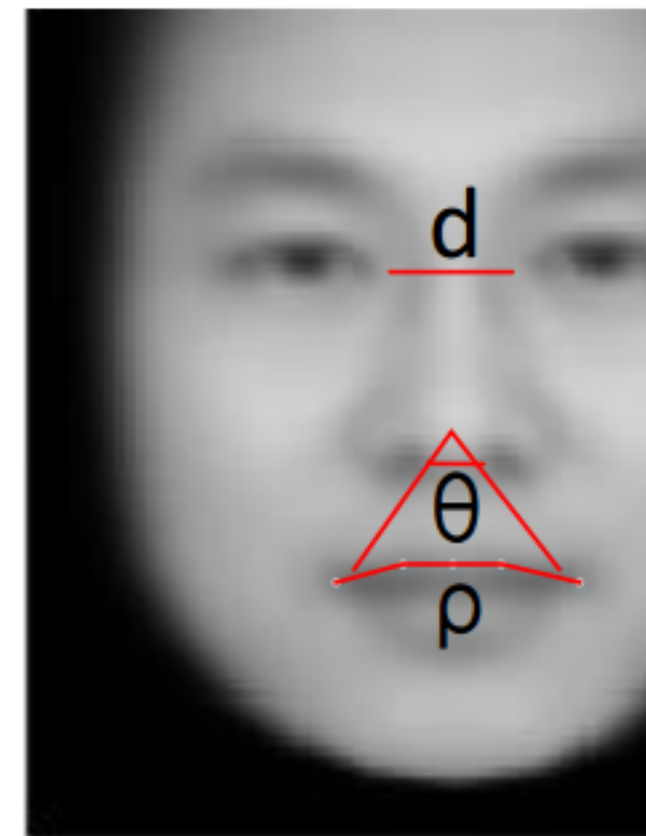
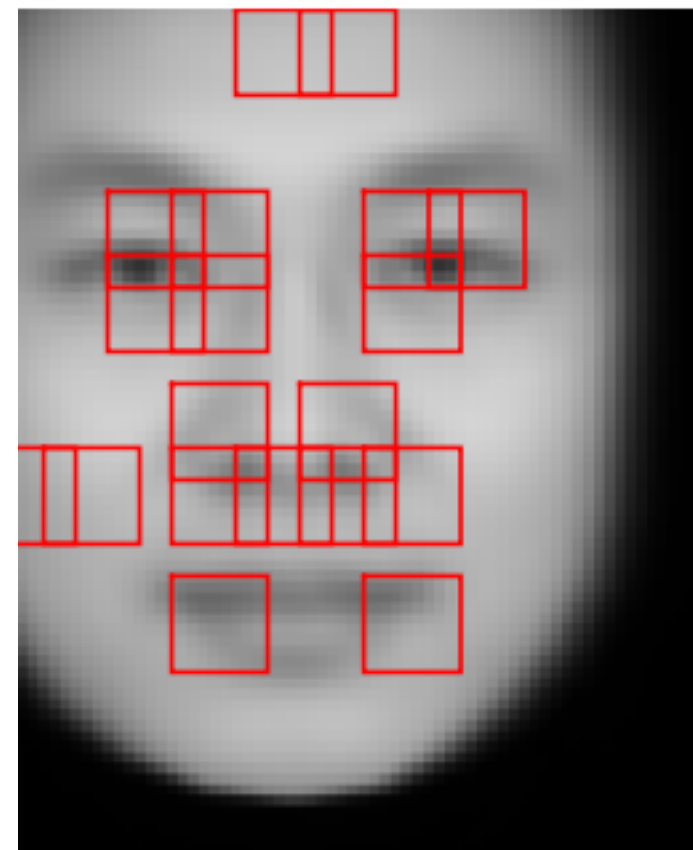
<https://www.independent.co.uk/news/world/americas/us-politics/capitol-riot-was-openly-organized-on-mainstream-social-media-b1784703.html>

Contentious applications

ARTIFICIAL INTELLIGENCE

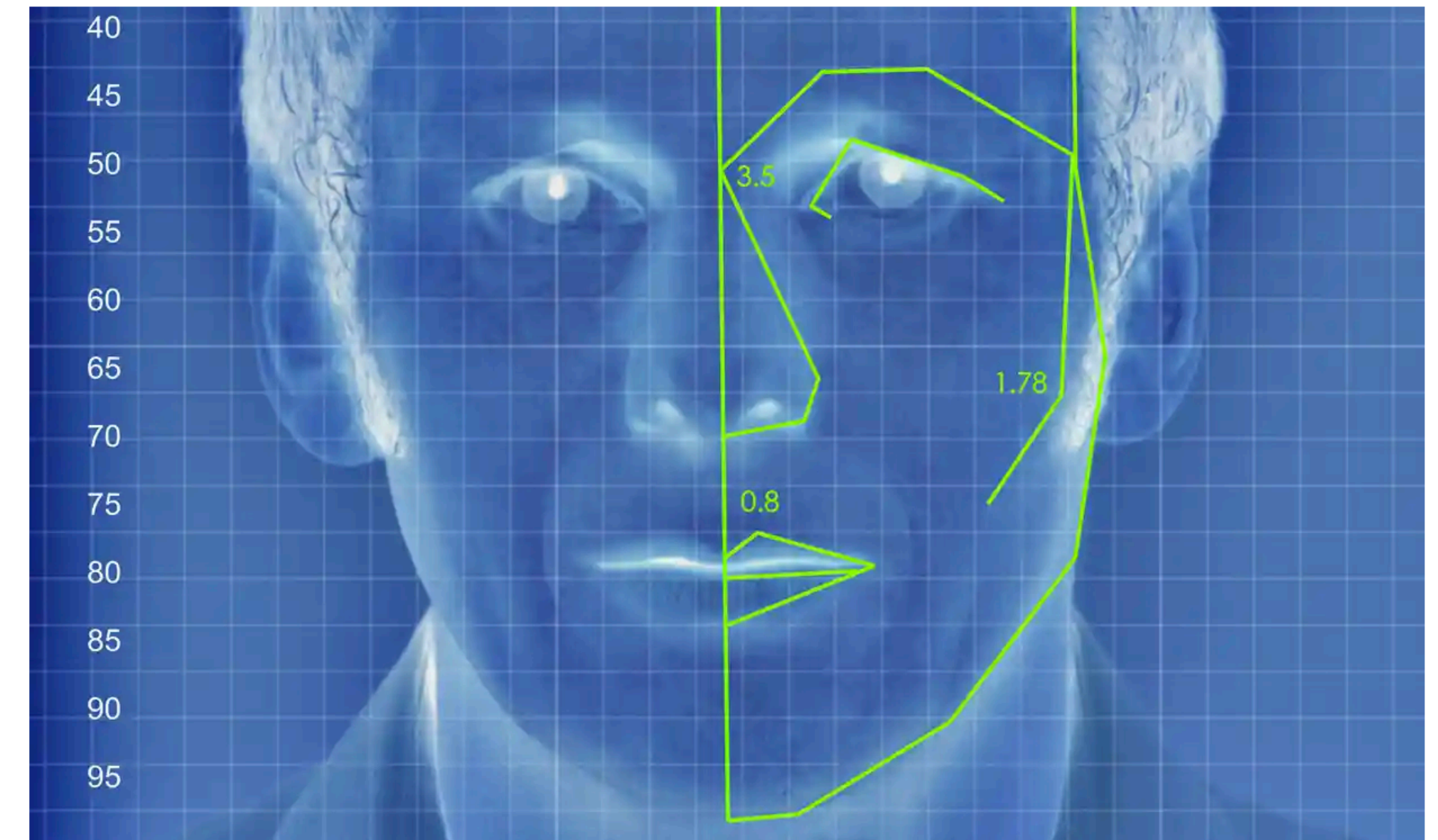
Neural Network Learns to Identify Criminals by Their Faces

The effort aimed at identifying criminals from their mugshots raises serious ethical issues about how we should use artificial intelligence.



New AI can guess whether you're gay or straight from a photograph

An algorithm deduced the sexuality of people on a dating site with up to 91% accuracy, raising tricky ethical questions



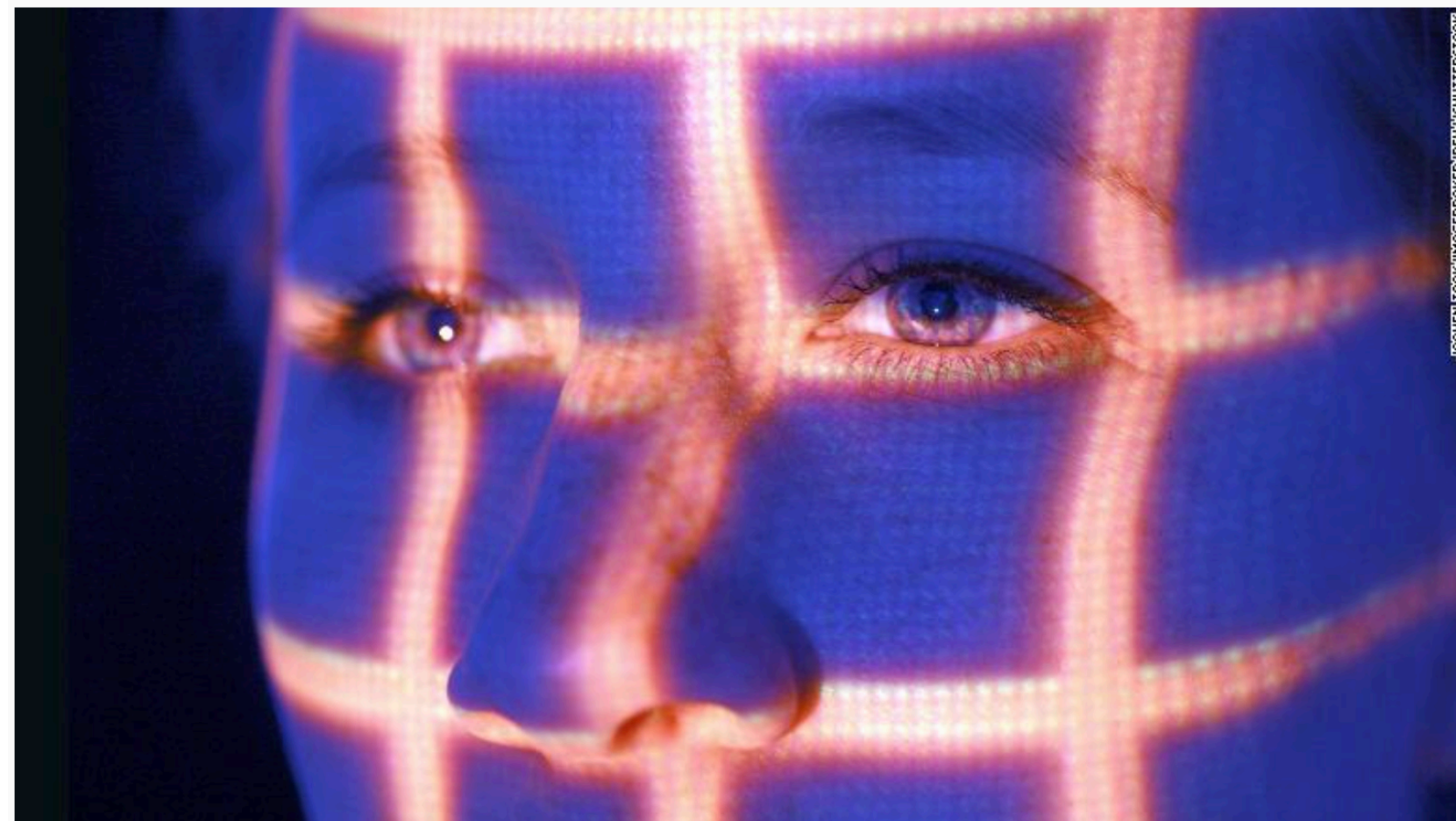
📷 An illustrated depiction of facial analysis technology similar to that used in the experiment.
Illustration: Alamy

Contentious applications

AI technology can identify genetic diseases by looking at your face, study says

By Nina Avramova, CNN

🕒 Updated 2117 GMT (0517 HKT) January 8, 2019



New AI technology could identify rare genetic diseases from patients' facial images.

(CNN) — A new artificial intelligence technology can accurately identify some rare genetic disorders using a photograph of a patient's face, according to a new study.

The AI technology, called DeepGestalt, outperformed clinicians in identifying a range of syndromes in three trials and could add significant value in personalized care, according to the [study](#) published Monday in the journal Nature Medicine.



FACEPTION IS A FACIAL PERSONALITY ANALYTICS TECHNOLOGY COMPANY

We reveal personality from facial images at scale to revolutionize how companies, organizations and even robots understand people and dramatically improve public safety, communications, decision-making, and experiences.

Plagiarism?

Artificial
intelligence (AI)

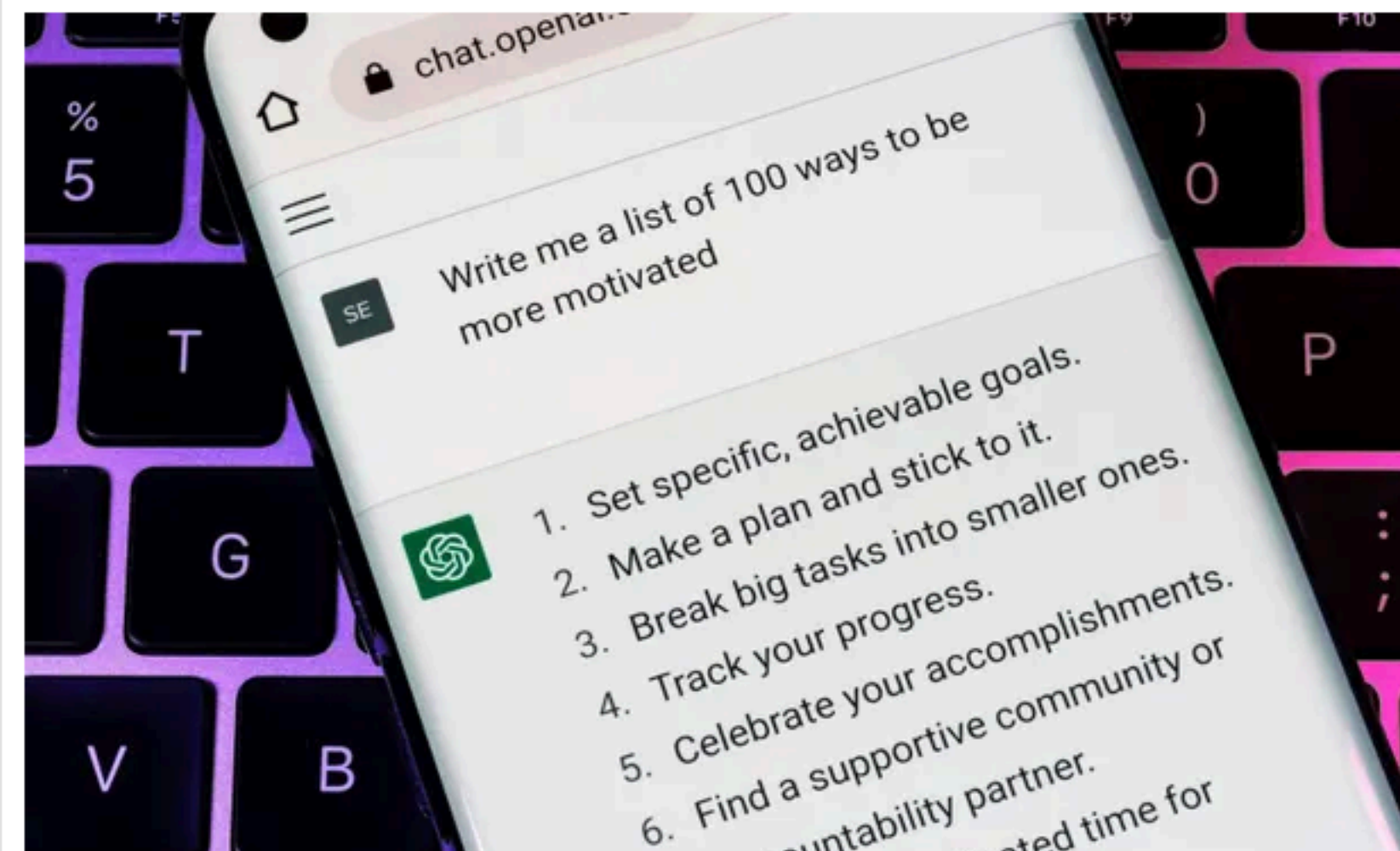
Lecturers urged to review assessments in UK amid concerns over new AI tool

ChatGPT is capable of producing high-quality essays with minimal human input

● **ChatGPT: what can the extraordinary artificial intelligence chatbot do?**

Sally Weale
Education
correspondent

Fri 13 Jan 2023 16.23 GMT



📷 ChatGPT has already triggered concerns about the potential for hard-to-detect plagiarism and questions about the validity of the essay as a future form of assessment. Photograph: Ascannio/Alamy

Lecturers at UK universities have been urged to review the way in which their courses are assessed amid concerns that students are already using a potent new AI tool capable of **producing high-quality essays** with minimal human input.

Weaponry



SLAUGHTERBOTS ARE HERE.

The era in which algorithms decide who lives and who dies is upon us.
We must act now to prohibit these weapons.



LEARN MORE

Accelerating the climate crisis

Oil in the Cloud

How Tech Companies are Helping Big Oil Profit from Climate Destruction

Published: 05-19-2020

Executive Summary

As the oil and gas industry confronts the end of the oil age and deteriorating earnings, major oil corporations such as Shell, BP, Chevron, ExxonMobil and others have turned to the cloud giants and their high powered computing capabilities to find and extract more oil and gas and reduce production costs. Despite the biggest cloud companies' commitments to address climate change, Microsoft, Google, and Amazon all have connections to some of the world's dirtiest oil companies for the explicit purpose of getting more oil and gas out of the ground and onto the market faster and cheaper.

Coursework 1 (25% of course mark)

- You will submit **slides** and a **video presentation** using those slides
- The video should be **5-10 minutes** and **must include your face**
- You should:
 1. introduce a real-world machine learning application
 2. Critique it according to the 5 ethical principles introduced earlier
 3. Recommend what can be done differently
- The full brief, submission instructions, and the marking rubric are available on Learn under the “Assessment” tab (after 0950 today!). **Deadline: 21/2 @ 1600**

Summary

- We have motivated machine learning
- We have looked at some examples of supervised machine learning
- We have looked at ethical issues that arise when applying machine learning