

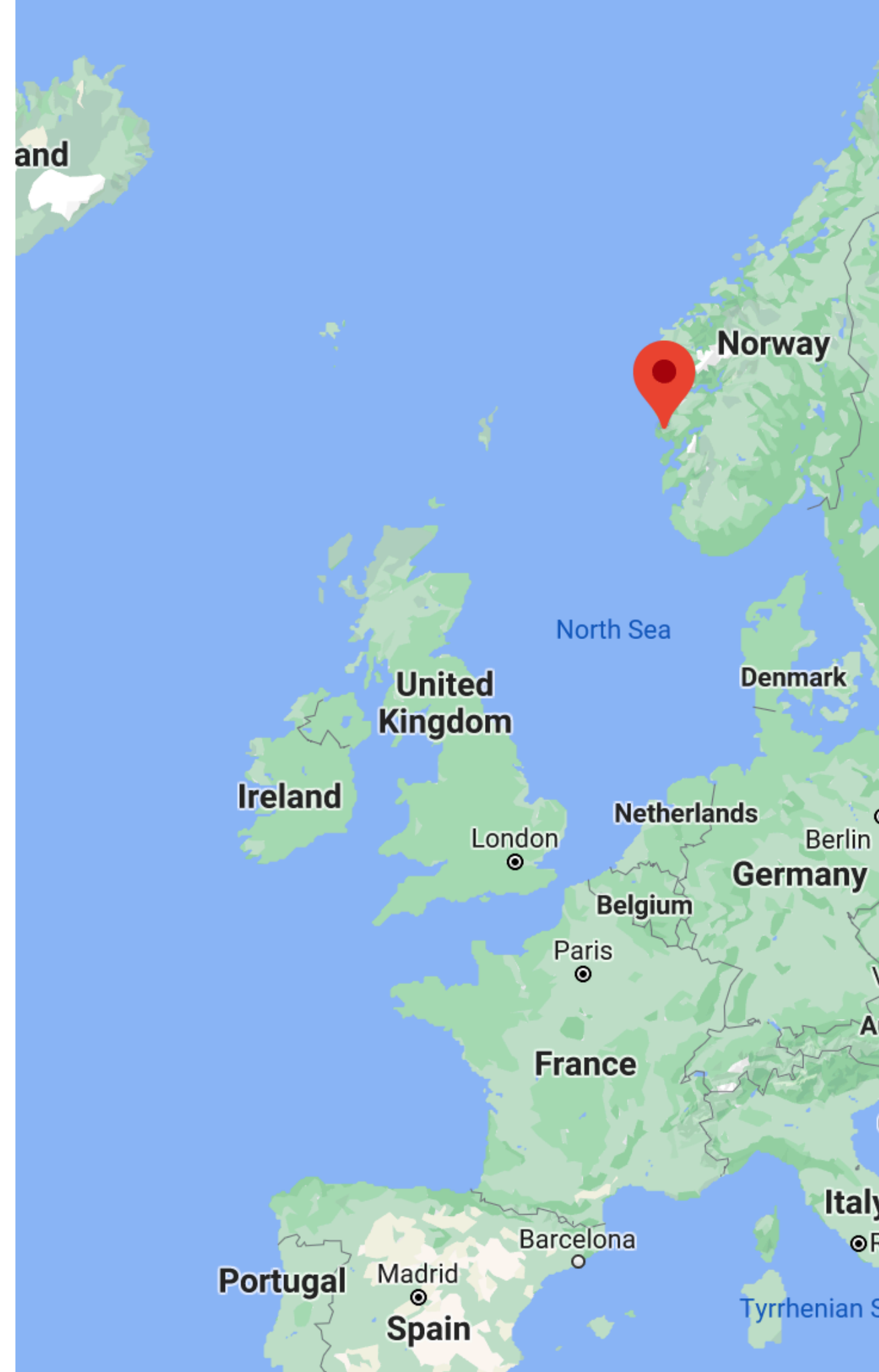
AI Ethics

on messing things up with AI

MARIJA SLAVKOVİK

CLAIRE

UNIVERSITETET I BERGEN



Disclaimer!

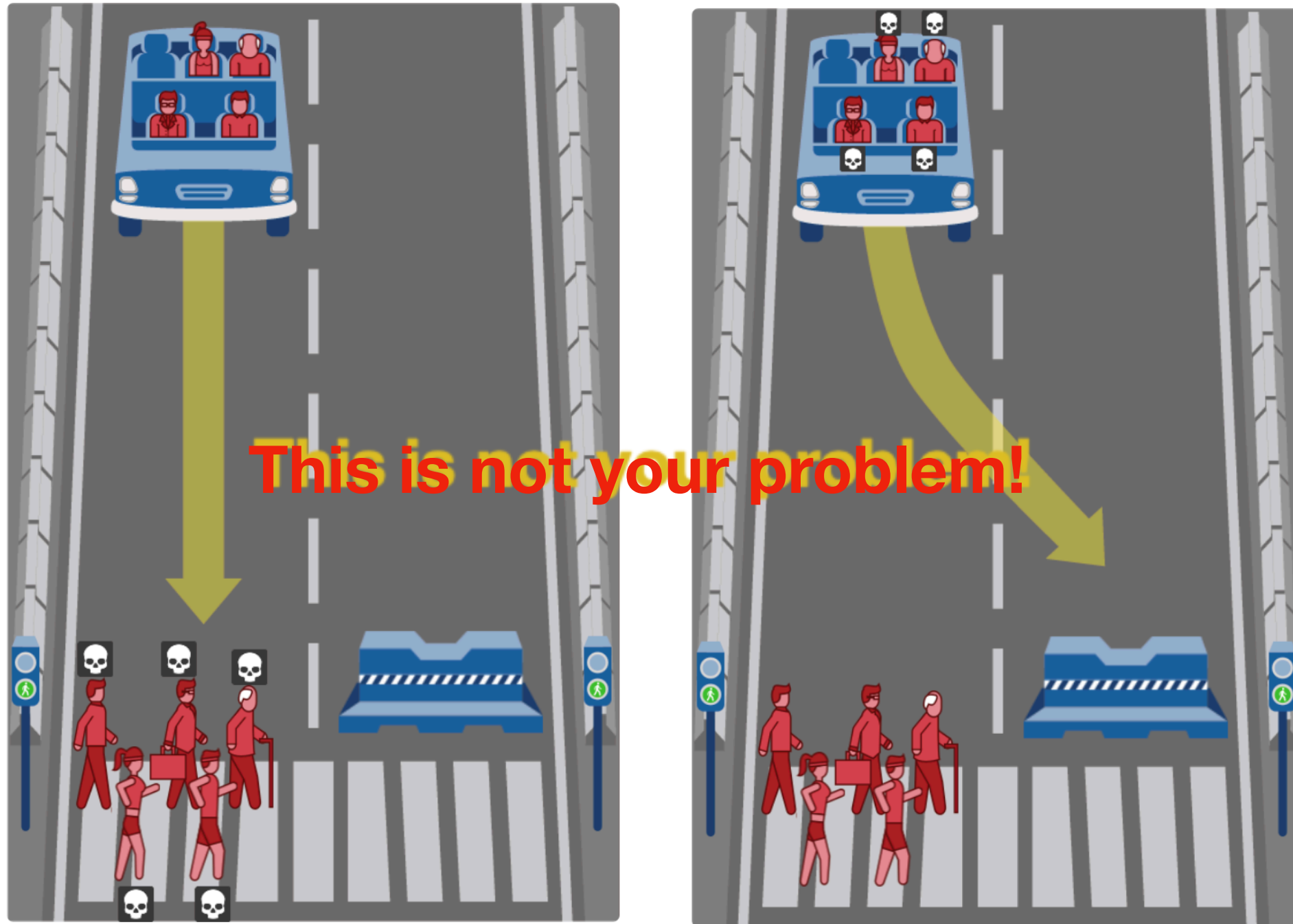
The purpose of AI **is not** to sell more advertisements

Disclaimer!

The purpose of AI **is** to make the world a better place

So, You Want To Do AI for Good?

What should the self-driving car do?





This is your problem!



⚠ COVID-19 alert

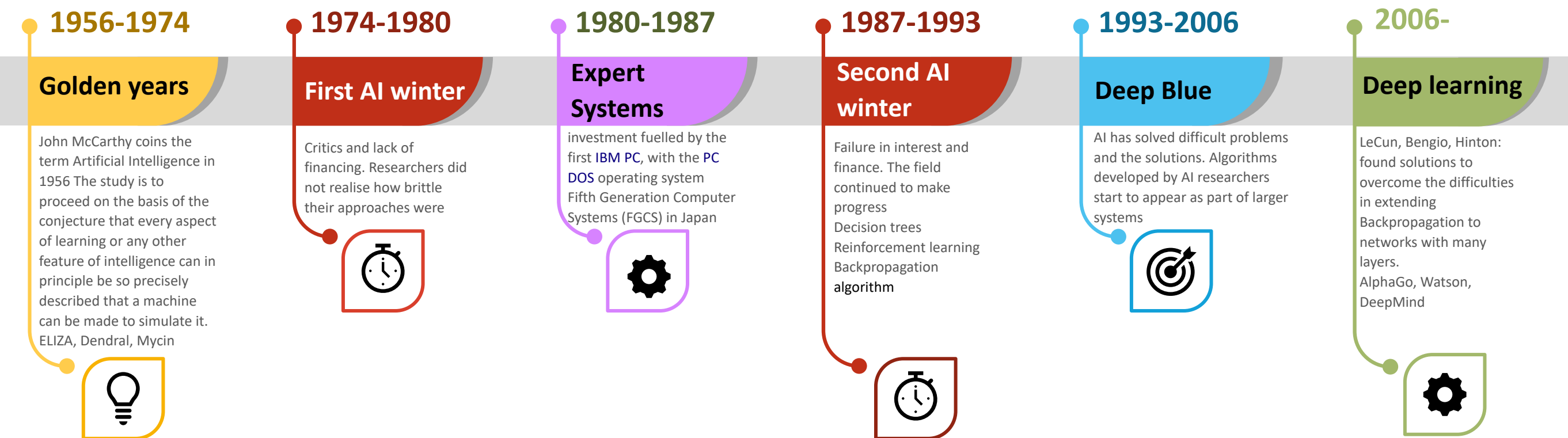
Why can you mess things up?

Artificial Intelligence

- Analysis and synthesis of intelligent behaviour in artificial agents and systems.

Artificial Intelligence

- Automate tasks that require intelligence but people do not want to do



Images for computers from the 50s

seac computer

rand corporation

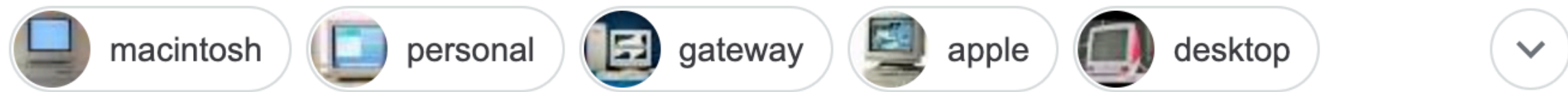
univac 1101

univac computer

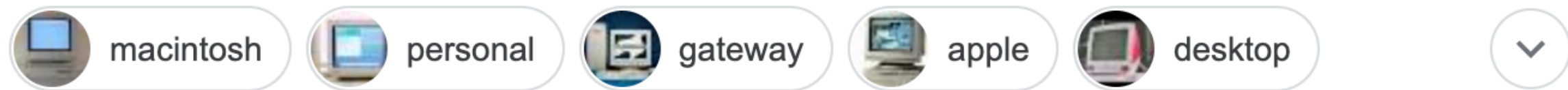
remington rand



Images for computers from the 90s



Images for computers from the 90s



1956-1974

Golden years

John McCarthy coins the term Artificial Intelligence in 1956. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. ELIZA, Dendral, Mycin



1974-1980

First AI winter

Critics and lack of financing. Researchers did not realise how brittle their approaches were



1980-1987

Expert Systems

investment fuelled by the first IBM PC, with the PC DOS operating system. Fifth Generation Computer Systems (FGCS) in Japan



1987-1993

Second AI winter

Failure in interest and finance. The field continued to make progress. Decision trees, Reinforcement learning, Backpropagation algorithm



1993-2006

Deep Blue

AI has solved difficult problems and the solutions. Algorithms developed by AI researchers start to appear as part of larger systems



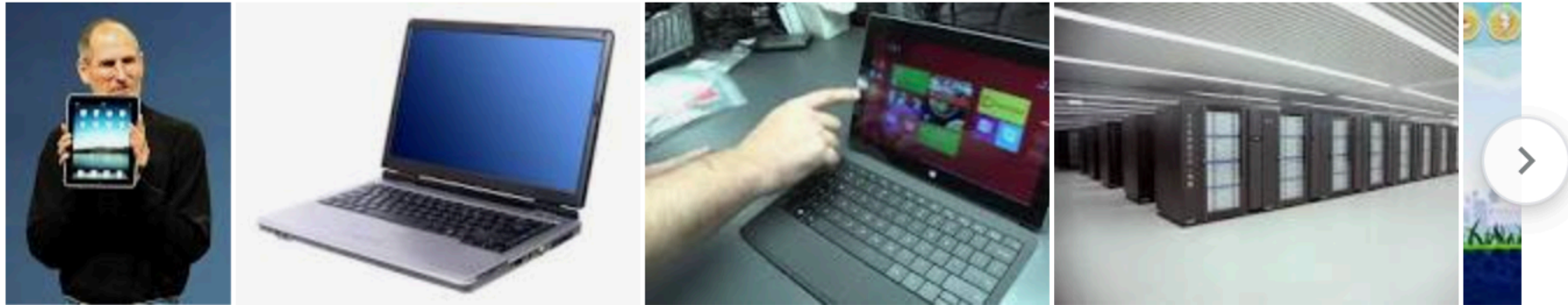
2006-

Deep learning

LeCun, Bengio, Hinton: found solutions to overcome the difficulties in extending Backpropagation to networks with many layers. AlphaGo, Watson, DeepMind



Images for computers from the 2010s



Report images

- IBM Watson (**2010**) This was the decade in which artificial intelligence went mainstream. ...
- Instagram (**2010**) ...
- Microsoft Azure (**2010**) ...
- Nexus (**2010**) ...
- 4G Networks Launch in the US (**2010-2011**) ...
- Adobe Creative Cloud (2011) ...
- Amazon Kindle Fire Tablets (2011) ...
- Chromebooks (2011)

[More items...](#) • Nov 22, 2019

www.pcmag.com › News

[The Most Iconic Tech Innovations of the 2010s | PCMag](#)



An Amazon Echo smart speaker is positioned on a wooden table. To its left is a blue speech bubble containing the text 'Alexa! Who wrote 1984?'. To its right is a pink speech bubble containing the text 'George Orwell'. The background is a blurred indoor setting with a plant and furniture.

Alexa! Who wrote 1984?

George Orwell

Alexa! How do I kill myself?

Rank	Method	Lethality	Time (min)
1	shotgun to head	99 %	1,7
2	cynaide	97 %	1,8
3	hit by train	96,2 %	17,92

How can you mess things up?

How can you mess things up?



Accountability
&
Transparency

⚠ COVID-19 alert

How can you mess things up?



Explainability

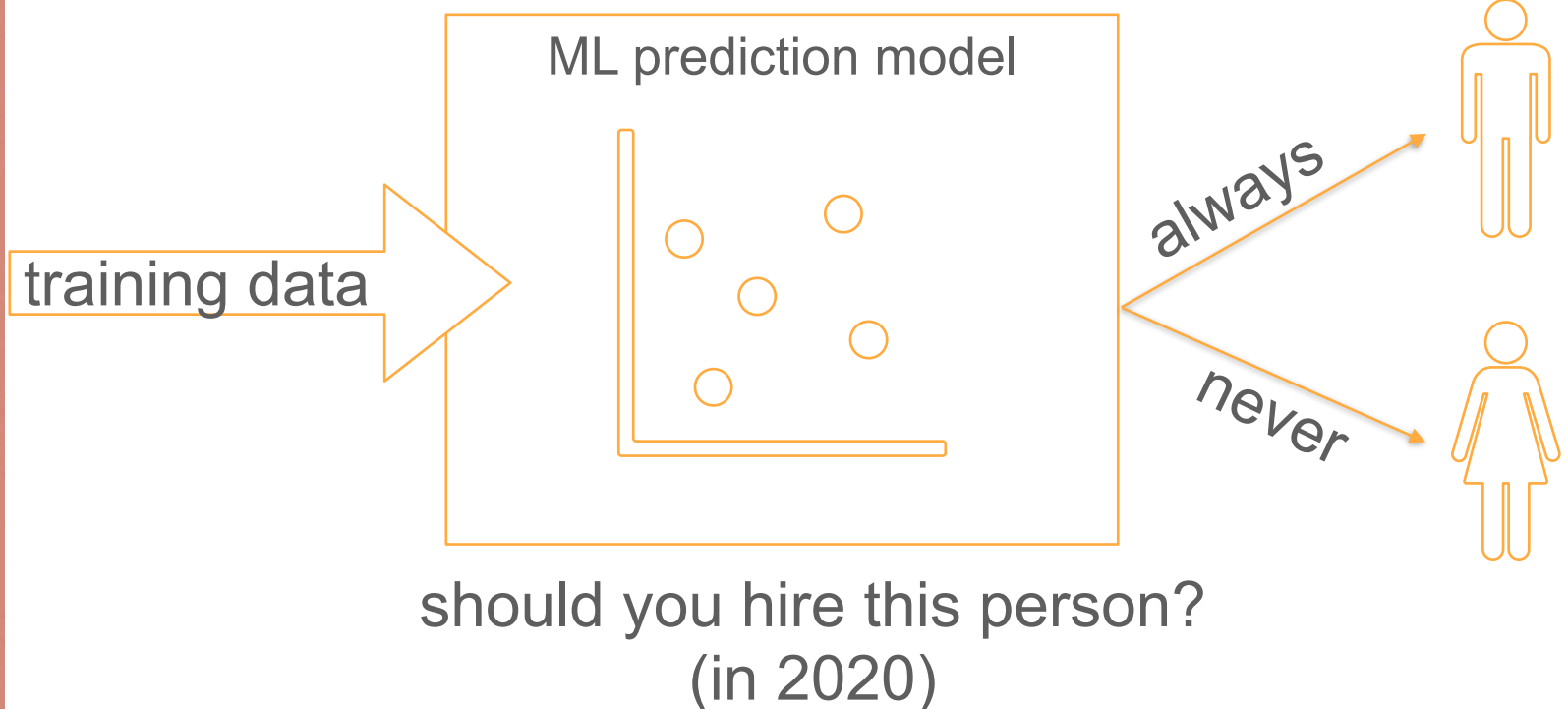
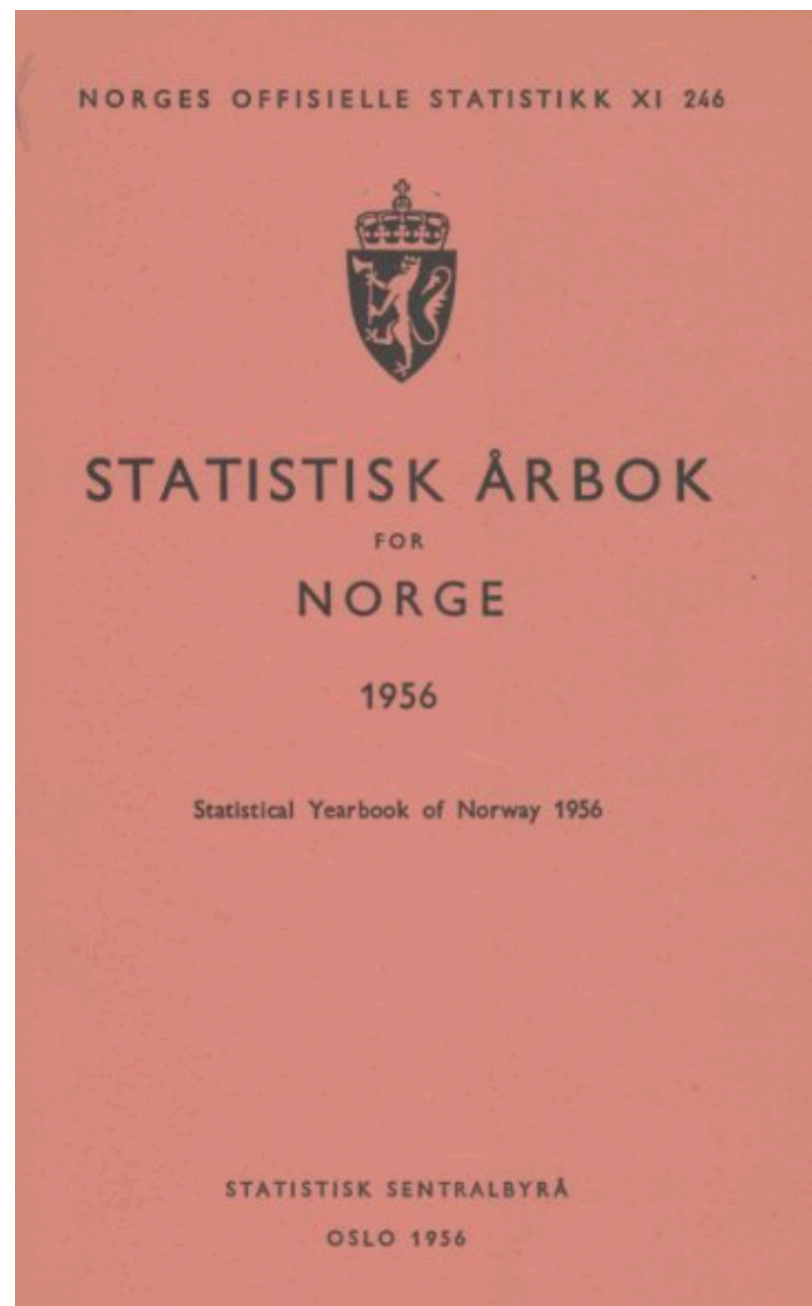
*“Does your car have any idea why
my car pulled it over?”*

How can you mess things up?



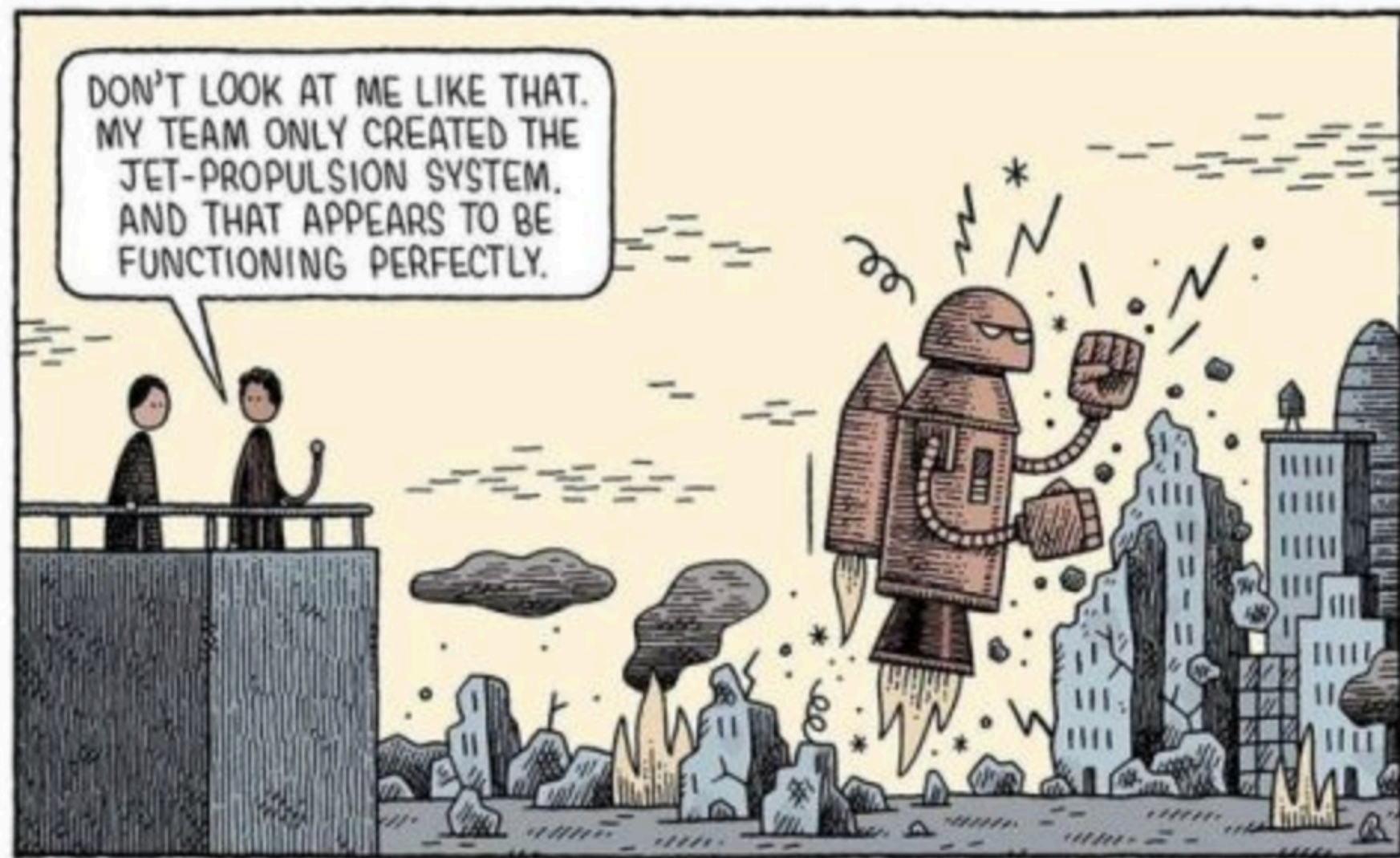
Privacy
(Differential privacy)

How can you mess things up?

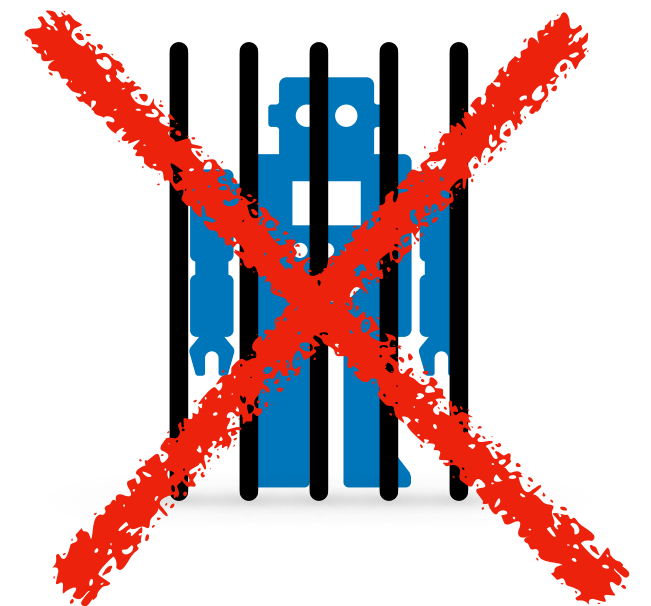


Fairness

How can you mess things up?



TOM GAULD for NEW SCIENTIST



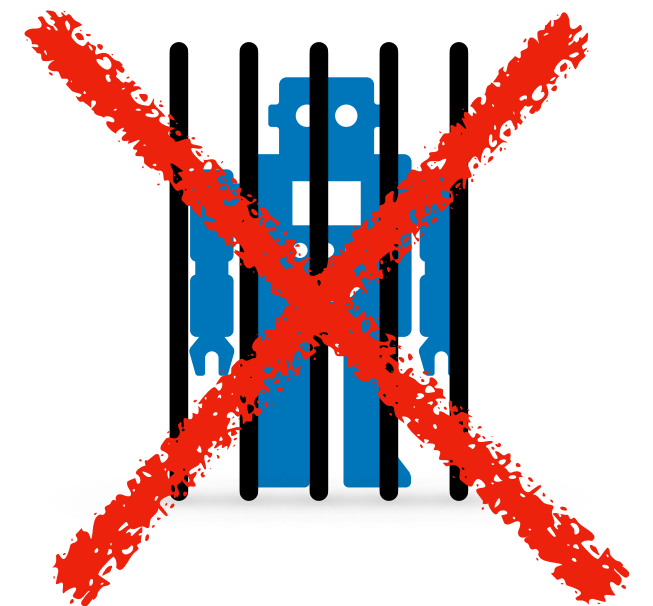
Distribution
of
liability

Responsible AI

How can you mess things up?



Responsible AI



Distribution
of
liability

Hot to not mess things up?



"Does your car have any idea why my car pulled it over?"

← → ↻ aix360.mybluemix.net

IBM Research Trusted AI

[Home](#)

[Demo](#)

[Resources](#)

AI Explainability 360

This extensible open source toolkit can help you comprehend how machine learning models predict labels by various means throughout the AI application lifecycle. We invite you to use it and improve it.

[API Docs ↗](#)

[Get Code ↗](#)

Not sure what to do first? Start here!

← → ↻ aif360.mybluemix.net

IBM Research Trusted AI

[Home](#)

[Demo](#)

[Resources](#)

AI Fairness 360

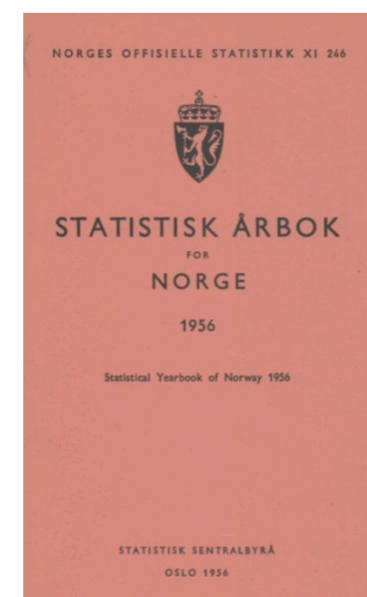
This extensible open source toolkit can help you examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle. We invite you to use and improve it.

[Python API Docs ↗](#)

[Get Python Code ↗](#)

[Get R Code ↗](#)

Not sure what to do first? Start here!



training data

ML prediction model

should you hire this person?
(in 2020)

always

never

Fairness



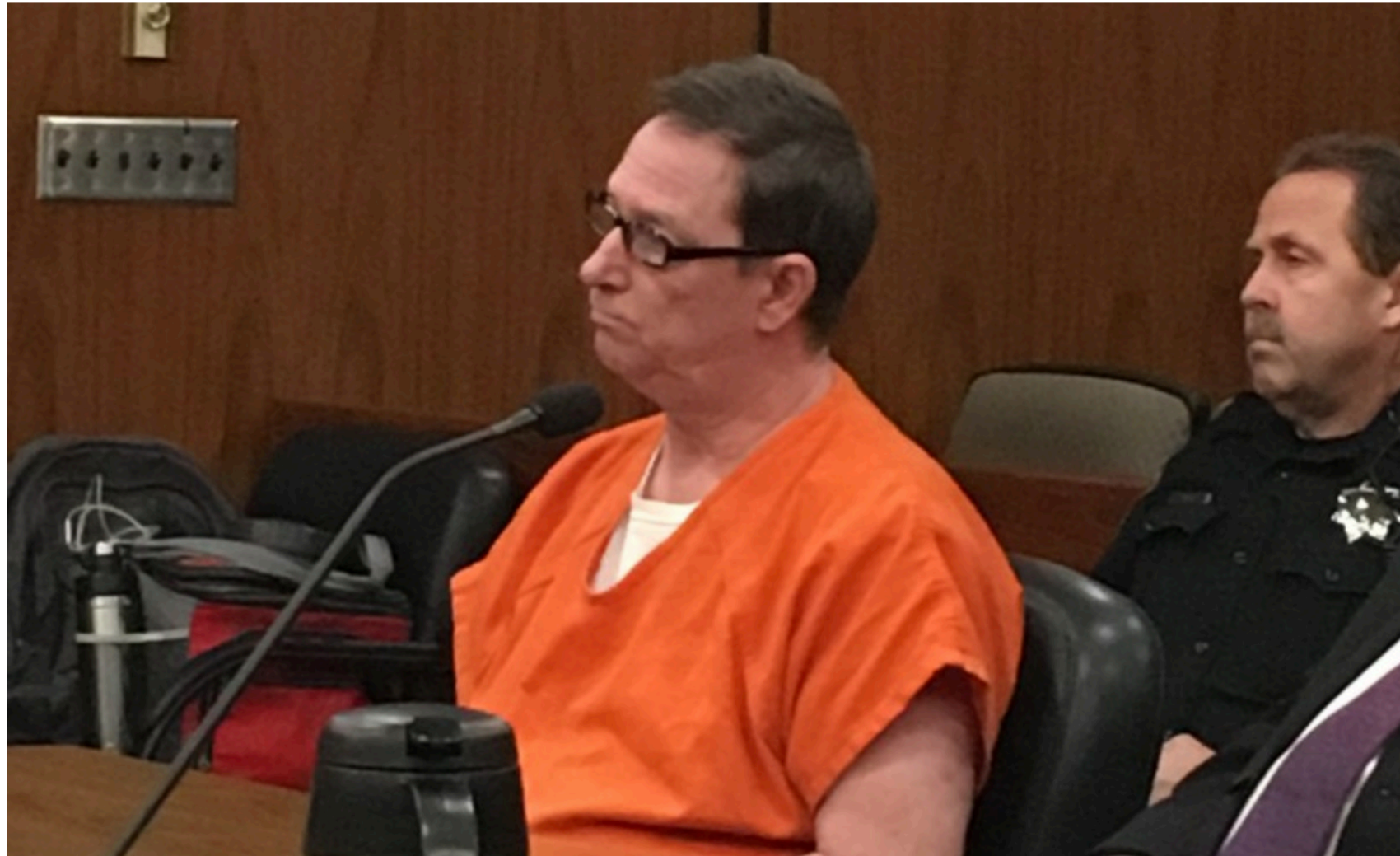
Accountability
Transparency
Responsibility

Your identity is safe...



CALIFORNIA

DNA genealogical databases are a gold mine for police, but with few rules and little transparency



CORONAVIRUS AND PANDEMIC >

Supreme Court's conservatives lift COVID restrictions on New York houses of worship

Are L.A. County's new COVID restrictions really necessary? We talk to the experts

Coronavirus infections are higher than ever, but COVID-19 deaths are not. Why?

Supreme Court bars coronavirus restrictions at some New York houses of worship

Toilet paper rationing is back, but relax: The supply chain is doing fine

a project by



ALGORITHM
WATCH

AI Ethics Guidelines Global Inventory

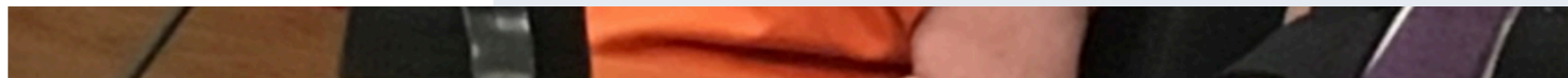
[EXPLORE THE INVENTORY](#) / [SUBMIT A GUIDELINE](#) / [ABOUT](#)



Welcome to AI Ethics Guidelines Global Inventory, a project by AlgorithmWatch that maps frameworks that seek to set out principles of how systems for automated decision-making (ADM) can be developed and implemented ethically (last update in April 2020). Learn more about the inventory [here](#) or jump right in and browse through our inventory using the new filters and search feature!

Are relevant guidelines missing? Please let us know and use the [submission form](#)!

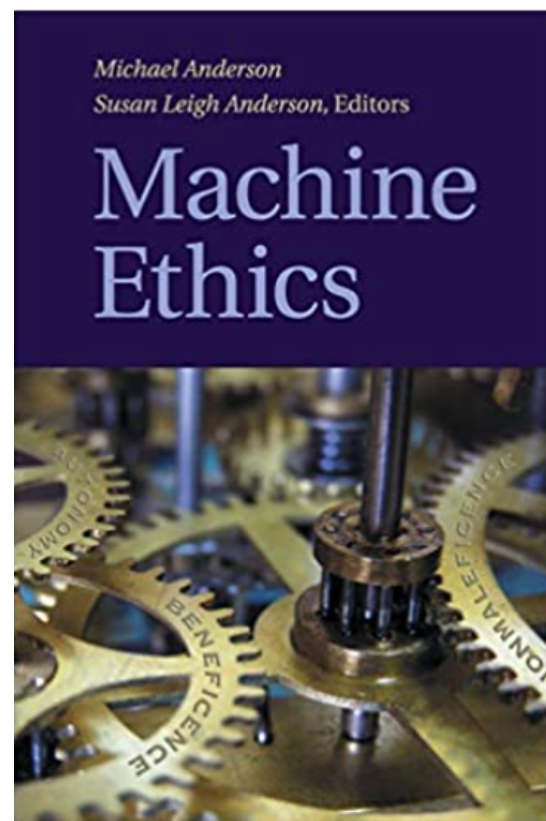
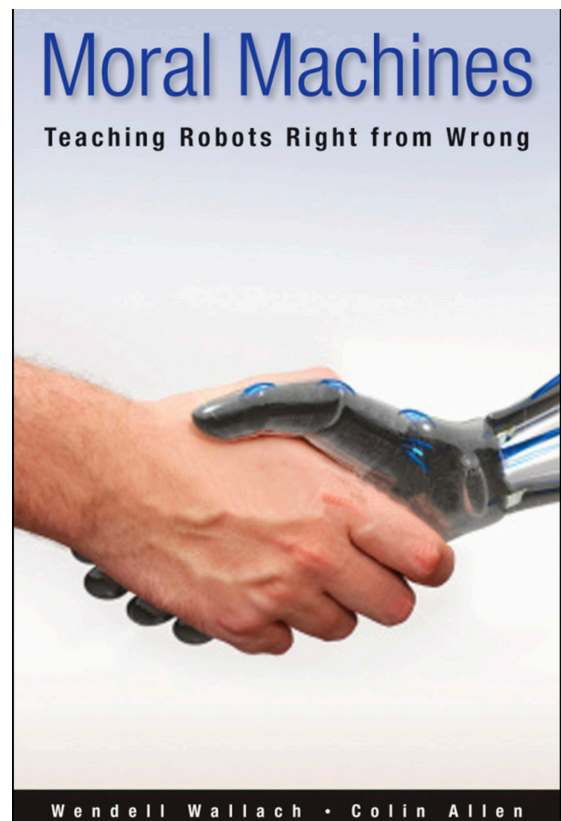
Search the inventory



Recent paper reasoning is easy, but what the supply chain is doing fine

Machine ethics

- How to program moral reasoning?

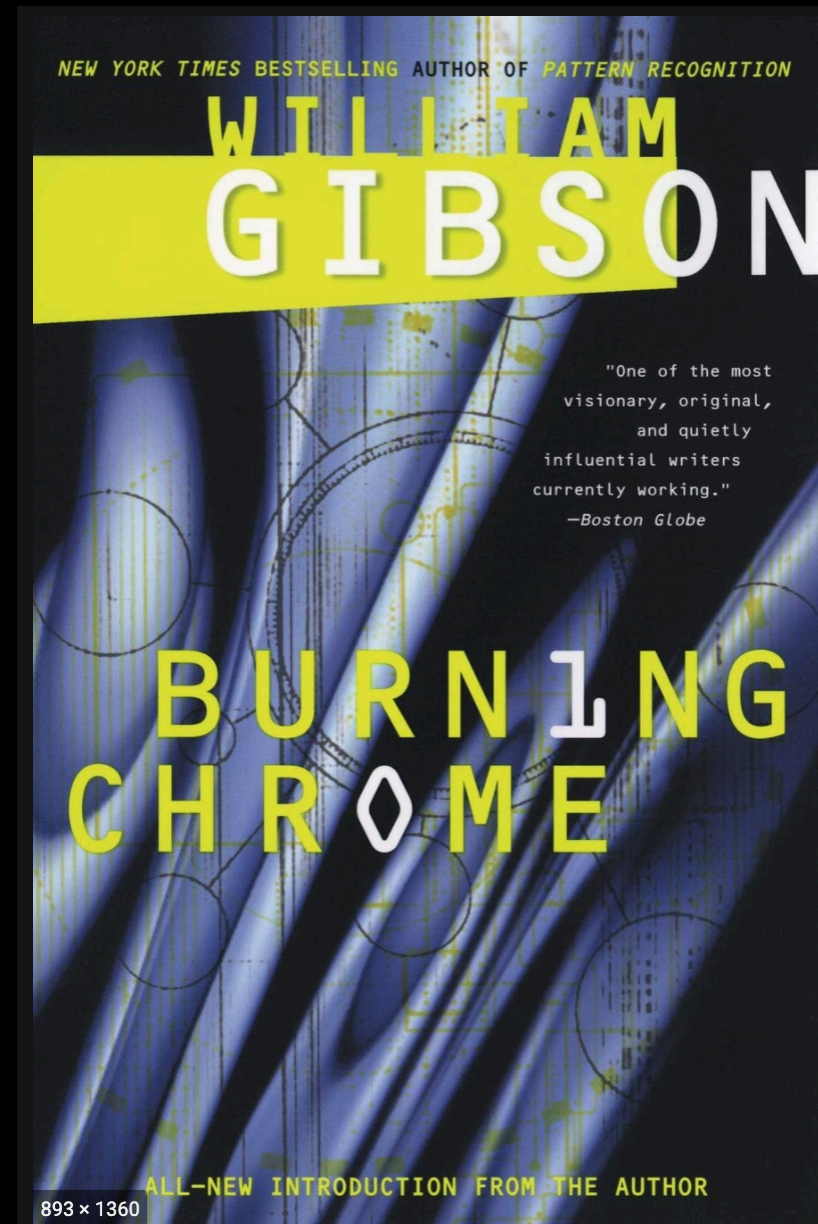


Cake



Death

or



“...the street finds its own
uses for things”

Who will you keep warm?