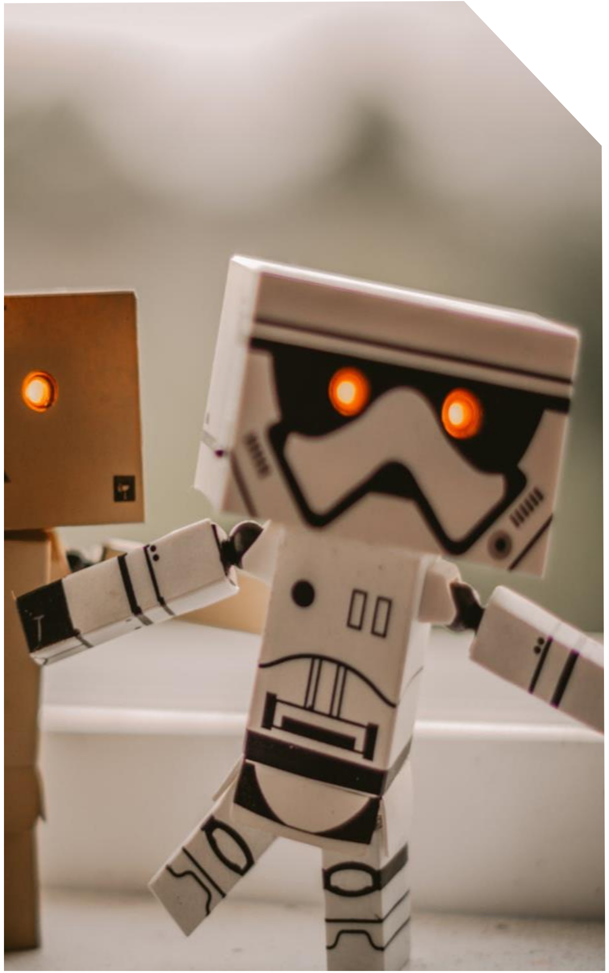
TRUST & TECHNOLOGY INITIATIVE

Machine learning: risks and bias

Dr Jennifer Cobbe
Trust & Technology Initiative
Department of Computer Science and Technology

Web: www.trusttech.cam.ac.uk
Email: jennifer.cobbe@cl.cam.ac.uk
Twitter: @jennifercobbe | @CamTrustTech

UNIVERSITY OF CAMBRIDGE



TRUST & TECHNOLOGY INITIATIVE

Trust & Technology Initiative

- Multi-disciplinary research initiative exploring the dynamics of trust and distrust around internet technologies, society, and power.

- Website: www.trusttech.cam.ac.uk
- Twitter: @CamTrustTech
- Mailing list: www.bit.ly/CamTrustTechList
- Zotero: www.bit.ly/CamTrustTechLibrary

UNIVERSITY OF CAMBRIDGE

TRUST & TECHNOLOGY INITIATIVE

ANNALS OF TECHNOLOGY MARCH 4, 2019 ISSUE

ARE ROBOTS COMPETING FOR YOUR JOB?

Probably, but don't count yourself out.

By Jill Lepore

UNIVERSITY OF CAMBRIDGE

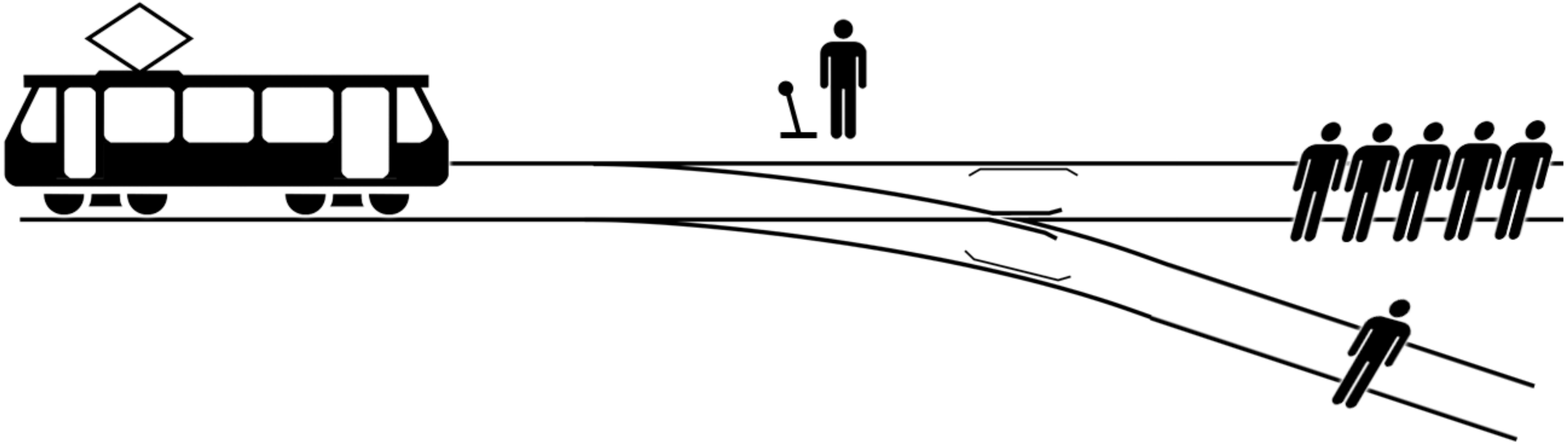
TRUST &
TECHNOLOGY
INITIATIVE

The Troubling Trajectory Of Technological Singularity

Jayshree Pandya Contributor
COGNITIVE WORLD Contributor Group
AI & Big Data
Jayshree Pandya is Founder of Risk Group & Host of Risk Roundup.

UNIVERSITY OF
CAMBRIDGE

TRUST &
TECHNOLOGY
INITIATIVE



UNIVERSITY OF
CAMBRIDGE

TRUST & TECHNOLOGY INITIATIVE

Introduction

- Automation bias
- Opacity
- Normativity
- Errors
- Bias
- Discrimination
- Predictive privacy harms
- Surveillance
- Solutionism

UNIVERSITY OF CAMBRIDGE

Automation bias

- People are…
    - More likely to trust decisions made by machines than by other people
    - Less likely to exercise meaningful review of or identify problems with automated systems

- Problem for…
    - Engineers
    - Users
    - Reviewers

TRUST & TECHNOLOGY INITIATIVE

Opacity

- Problems for
  - Design and engineering
  - Problems for accountability and oversight

UNIVERSITY OF CAMBRIDGE

TRUST & TECHNOLOGY INITIATIVE

Normativity

- Technology is neither good nor bad – but also not neutral
- Algorithm: “a series of steps undertaken in order to solve a particular problem or accomplish a defined outcome” (Diakopoulos 2015).
- Technology – including ML – is inherently normative
- In what context could a given ML system be used and for what purpose?

UNIVERSITY OF CAMBRIDGE

Errors

- All predictive systems have margins of error
  - Training = to within an acceptable margin of error

- ML systems will make mistakes and these mistakes will have consequences

- Engineers need to think about
  - Detecting errors
  - Rectifying of errors
  - Accommodating errors

Technology Is Biased Too. How Do We Fix It?

Algorithms were supposed to free us from our unconscious mistakes. But now there's a new set of problems to solve.

By Laura Hudson

Filed under If Then Next

Published Jul. 20, 2017

TRUST &
TECHNOLOGY
INITIATIVE

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

UNIVERSITY OF
CAMBRIDGE

TRUST & TECHNOLOGY INITIATIVE

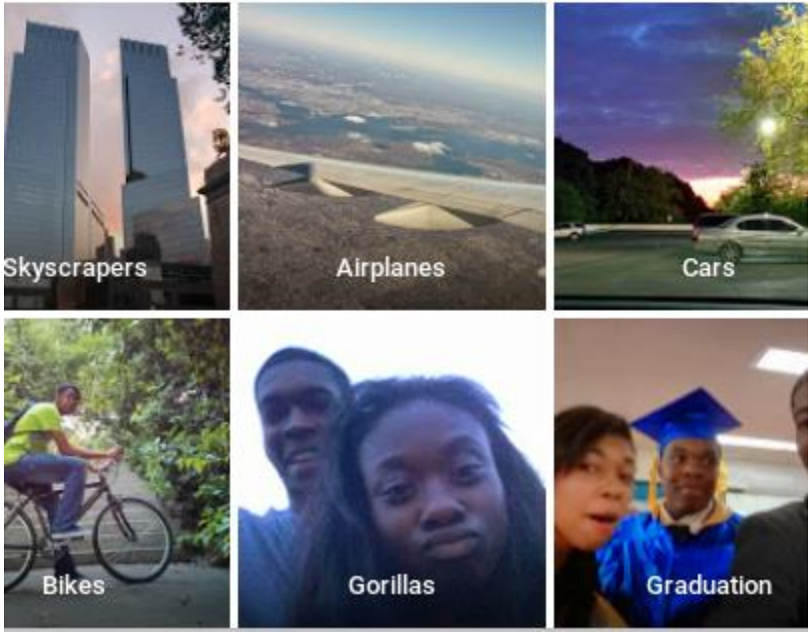
Bias

- Sentencing algorithm used in US criminal justice system

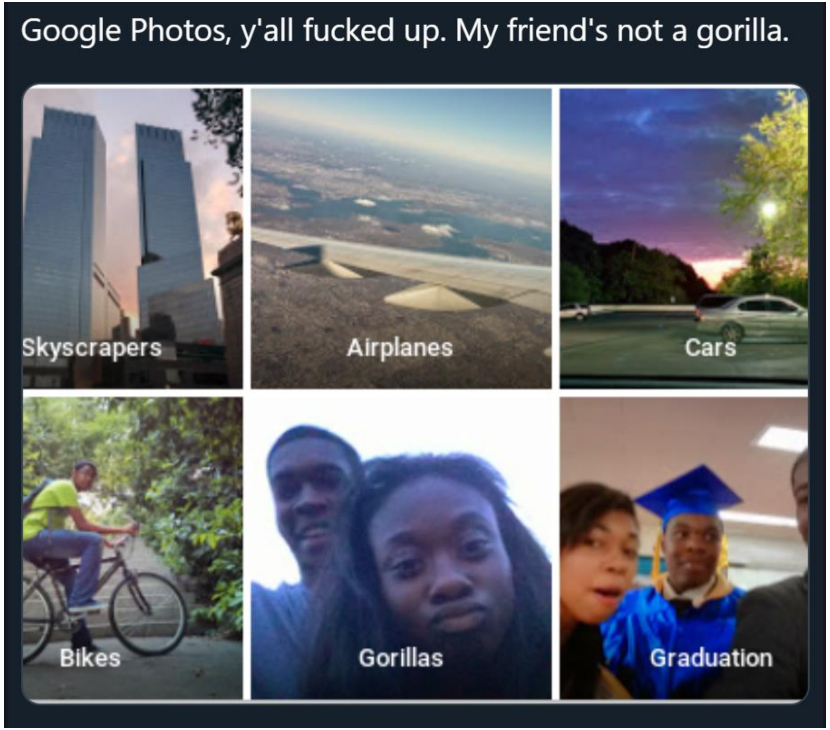
- Of those predicted to reoffend, 61% were subsequently arrested

- But not equal – when other factors controlled for:
  - White defendants routinely mislabelled as low risk
  - Black defendants 77% more likely to be rated at a higher risk of committing a violent crime
  - Black defendants 45% more like to be predicted to commit any crime

UNIVERSITY OF CAMBRIDGE


Skyscrapers
Airplanes
Cars
Bikes
Gorillas
Graduation

TRUST & TECHNOLOGY INITIATIVE



UNIVERSITY OF CAMBRIDGE

TRUST & TECHNOLOGY INITIATIVE

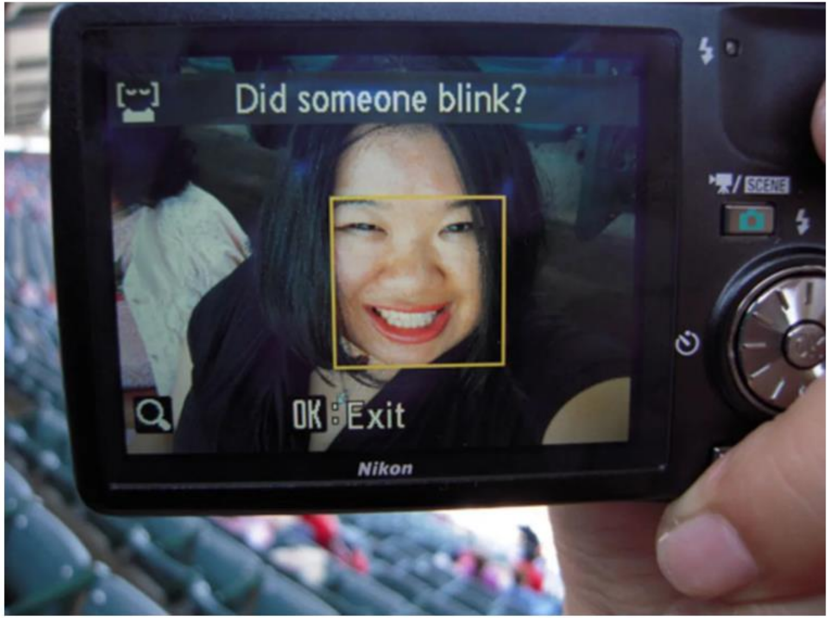
UNIVERSITY OF CAMBRIDGE

TRUST & TECHNOLOGY INITIATIVE

Camera Misses the Mark on Racial Sensitivity

Odelia Lee
5/15/09 7:40pm • Filed to: BLINK DETECTION

87.8K 14 Save



UNIVERSITY OF CAMBRIDGE

TRUST & TECHNOLOGY INITIATIVE

BUSINESS NEWS OCTOBER 10, 2018 / 4:12 AM / 4 MONTHS AGO

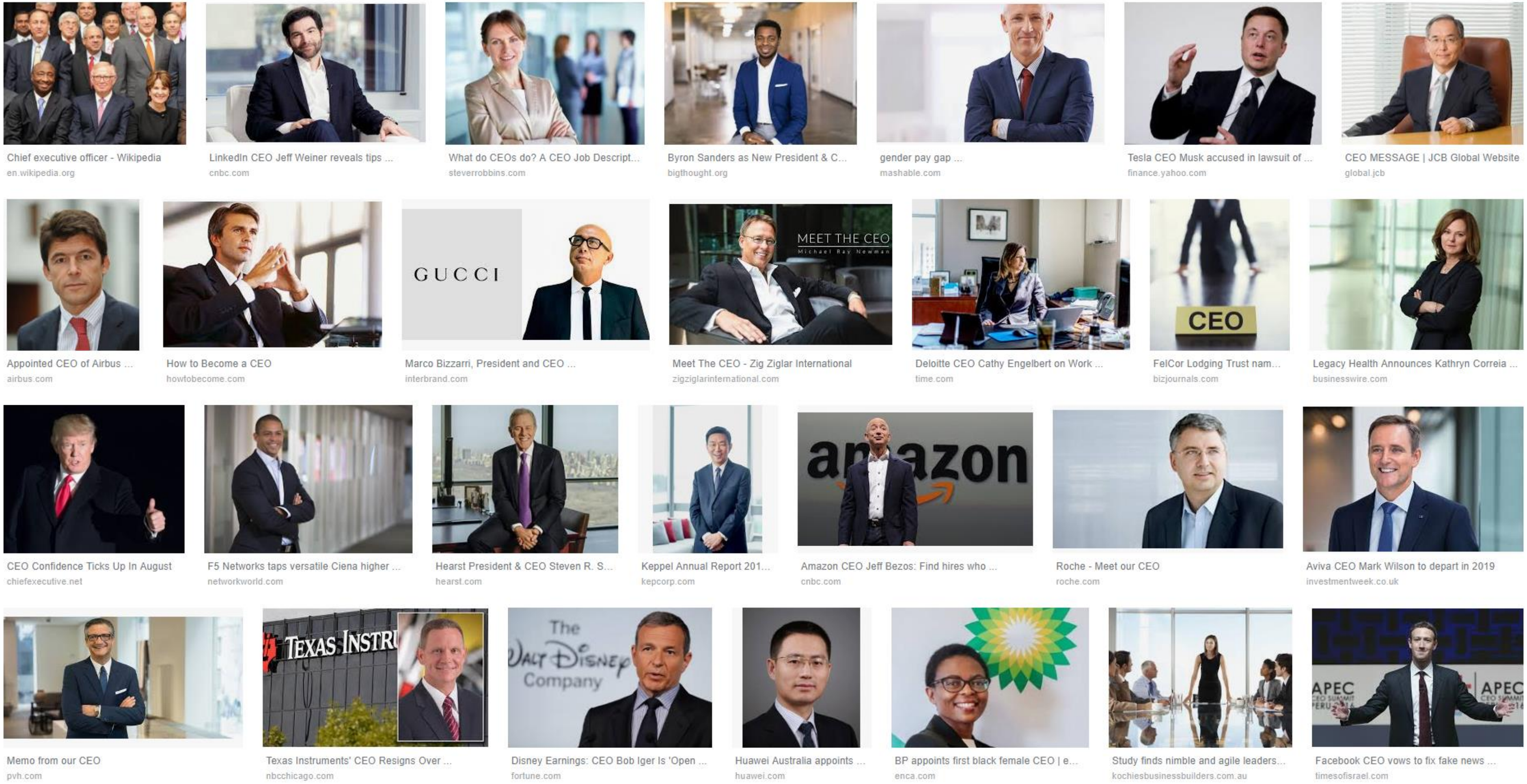
Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

UNIVERSITY OF CAMBRIDGE


Chief executive officer - Wikipedia
en.wikipedia.org
LinkedIn CEO Jeff Weiner reveals tips ...
cnbc.com
What do CEOs do? A CEO Job Descript...
steverrobbins.com
Byron Sanders as New President & C...
bigthought.org
gender pay gap ...
mashable.com
Tesla CEO Musk accused in lawsuit of ...
finance.yahoo.com
CEO MESSAGE | JCB Global Website
global.jcb
Appointed CEO of Airbus ...
airbus.com
How to Become a CEO
howtobecome.com
Marco Bizzarri, President and CEO ...
interbrand.com
Meet The CEO - Zig Ziglar International
zigziglarinternational.com
Deloitte CEO Cathy Engelbert on Work ...
time.com
FelCor Lodging Trust nam...
bizjournals.com
Legacy Health Announces Kathryn Correia ...
businesswire.com
CEO Confidence Ticks Up In August
chiefexecutive.net
F5 Networks taps versatile Ciena higher ...
networkworld.com
Hearst President & CEO Steven R. S...
hearst.com
Keppel Annual Report 201...
kepcorp.com
Amazon CEO Jeff Bezos: Find hires who ...
cnbc.com
Roche - Meet our CEO
roche.com
Aviva CEO Mark Wilson to depart in 2019
investmentweek.co.uk
Memo from our CEO
pvh.com
Texas Instruments' CEO Resigns Over ...
nbcchicago.com
Disney Earnings: CEO Bob Iger Is 'Open ...
fortune.com
Huawei Australia appoints ...
huawei.com
BP appoints first black female CEO | e...
enca.com
Study finds nimble and agile leaders...
kochiesbusinessbuilders.com.au
Facebook CEO vows to fix fake news ...
timesofisrael.com

TRUST &
TECHNOLOGY
INITIATIVE

TOM SIMONITE BUSINESS 08.21.17 09:00 AM

MACHINES TAUGHT BY PHOTOS LEARN A SEXIST VIEW OF WOMEN

UNIVERSITY OF CAMBRIDGE

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

By James Vincent | Mar 24, 2016, 6:43am EDT

Microsoft 'deeply sorry' for racist and sexist tweets by AI chatbot

Company finally apologises after 'Tay' quickly learned to produce offensive posts, forcing the tech giant to shut it down after just 16 hours

Microsoft's racist chatbot returns with drug-smoking Twitter meltdown

Short-lived return saw Tay tweet about smoking drugs in front of the police before suffering a meltdown and being taken offline

TRUST & TECHNOLOGY INITIATIVE

https://www.youtube.com/watch?v=TWWsW1w-BVo&t=74s

UNIVERSITY OF CAMBRIDGE

Bias

- Particular groups are or historically were treated less favourably -> model which repeats this difference in treatment

- Particular groups are or were societally disadvantaged -> model which repeats the disadvantage

- Training data not sufficiently varied for the system to have been trained to adequately handle all possible inputs -> model which is incapable of dealing with certain inputs equally to others

TRUST & TECHNOLOGY INITIATIVE

Bias

- ML can encode historical practices into predictions about the future
- ML systems are limited by their training data
- ML trained on data about society will reflect society’s biases and prejudices
- Poorest, most marginalised, and most vulnerable are most likely to be affected

UNIVERSITY OF CAMBRIDGE

TRUST & TECHNOLOGY INITIATIVE

Discrimination

- ‘Fair’ systems can still be discriminatory

- Discrimination is a legal term (Equality Act 2010)
    - Direct discrimination: *where people are treated less favourably on the basis of a protected characteristic*
    - Indirect discrimination: *where rules that appear to treat everyone equally have the practical effect of excluding, placing onerous requirements on, or disadvantaging people who share a protected characteristic*

UNIVERSITY OF CAMBRIDGE

TRUST & TECHNOLOGY INITIATIVE

Predictive privacy harms

- Inaccurate predictions with consequences for individual
- Accurate predictions disclosed to wrong person
- Predictive privacy harms can feed into discriminatory actions and other problems

TRUST & TECHNOLOGY INITIATIVE

Predictive privacy harms

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Facebook recommended that this psychiatrist's patients friend each other

Kashmir Hill
8/29/16 4:21pm • Filed to: REAL FUTURE
27.4K 6 5

FACEBOOK

How Facebook Outs Sex Workers

Kashmir Hill
10/11/17 2:20pm • Filed to: PEOPLE YOU MAY KNOW
663.9K 659 46

How Facebook's Targeted Ads Revealed One User's Sexuality

Adrian Chen
10/23/10 11:57AM Filed to: PRIVACY
27.09K

UNIVERSITY OF CAMBRIDGE

TRUST &
TECHNOLOGY
INITIATIVE

Predictive privacy harms

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

AI that can determine a person's sexuality from photos shows the dark side of the data age


Machine gaydar: AI is reinforcing stereotypes that liberal societies are trying to get rid of

September 13, 2017 11.35am BST

UNIVERSITY OF
CAMBRIDGE

TRUST &
TECHNOLOGY
INITIATIVE

Surveillance

- ML is increasingly used in surveillance
    - Predictive analytics
    - Biometric identification
- Surveillance capitalism
- State security and intelligence agencies
- Voter surveillance and microtargeting

UNIVERSITY OF
CAMBRIDGE

TRUST & TECHNOLOGY INITIATIVE

Solutionism

- Technology is often presented as an obvious solution to difficult problems

- But: socio-economic problems are rarely solved by technology

- Questions:
  - What problem are we trying to solve?
  - Is the best solution to that problem a technical one?
  - If so, is machine learning the correct technical solution to that problem?

UNIVERSITY OF CAMBRIDGE

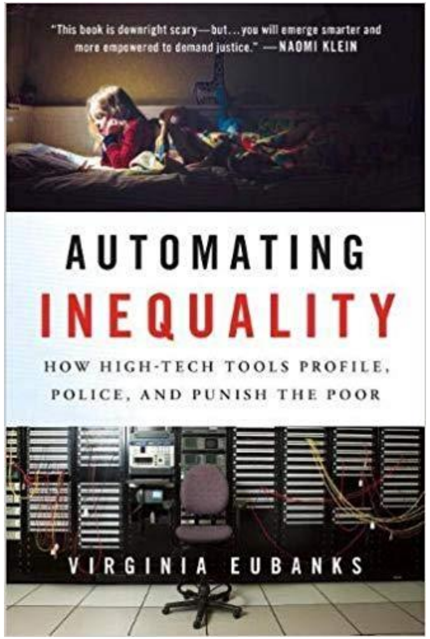
Conclusions

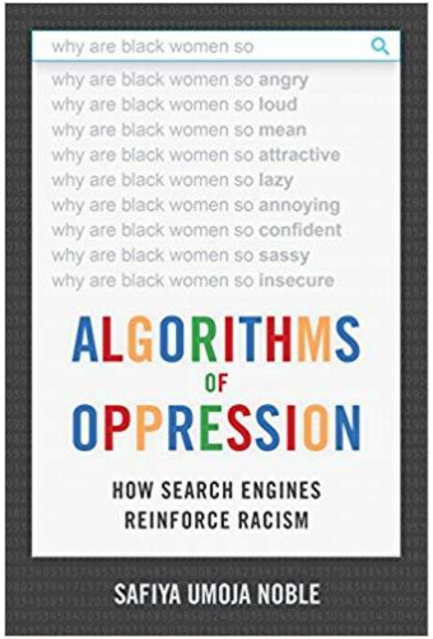
- Machine learning problems are human problems with human responsibility
    - Training datasets compiled by people
    - Models constructed by people
    - Models trained and tested by people
    - Systems used for purposes determined by people to achieve outcomes desired by people

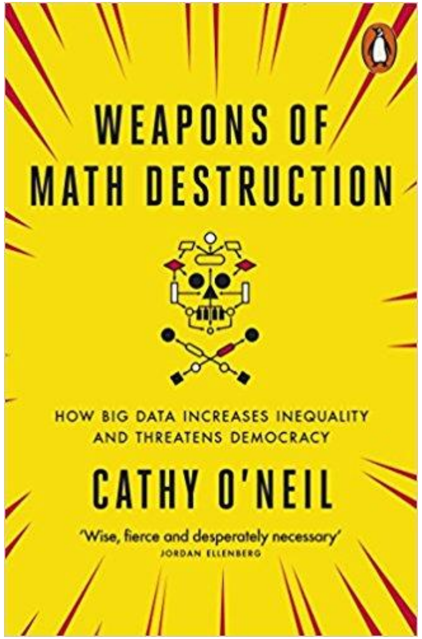
- Replicating human bias is an engineering failure

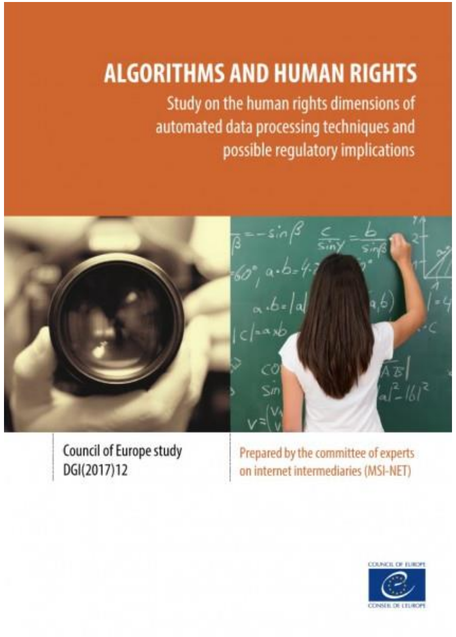
- Problems can only be avoided if you know about the risks and proactively take steps to avoid them

TRUST & TECHNOLOGY INITIATIVE









UNIVERSITY OF CAMBRIDGE

TRUST &
TECHNOLOGY
INITIATIVE

End

Dr Jennifer Cobbe

Trust & Technology Initiative

Department of Computer Science and Technology

Web: www.trusttech.cam.ac.uk

Email: jennifer.cobbe@cl.cam.ac.uk

Twitter: @jennifercobbe | @CamTrustTech

UNIVERSITY OF
CAMBRIDGE