Machine Learning & Society

Daniela Huppenkothen

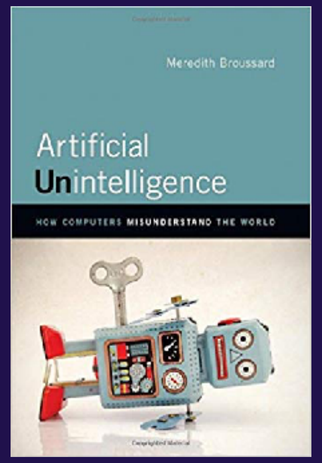
DIRAC, University of Washington

Tiana_Athriel

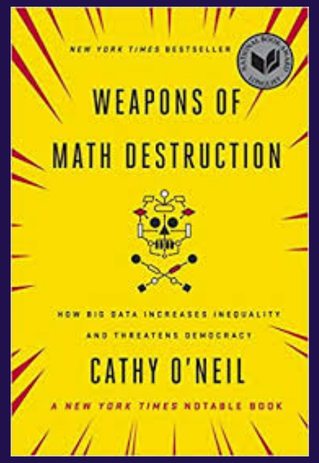
dhuppenkothen

dhuppenk@uw.edu

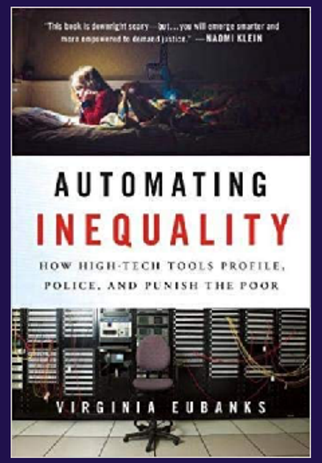
Resources



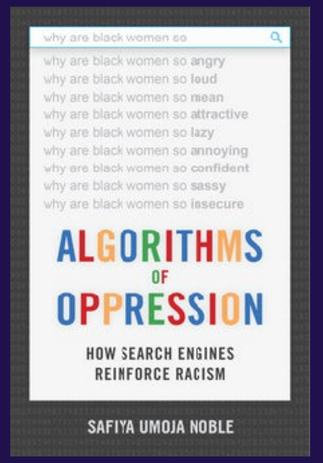
Meredith Broussard



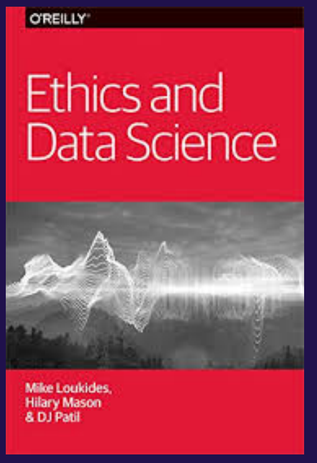
Cathy O'Neil



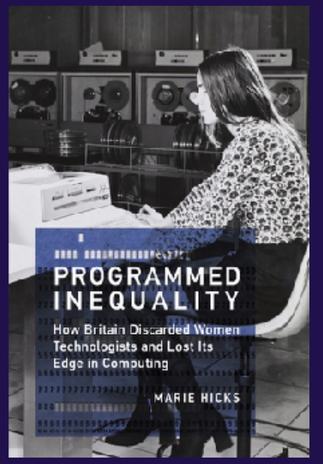
Virginia Eubanks



Safiya Umoja Noble



M Loudikes, H Mason, DJ Patil



Marie Hicks

But we are astronomers! Why should we care about ethics in machine learning?

- Our lives are affected every day by machine learning predictions.
- Technology and algorithms can be misused.
- There is a growing number of studies about the astronomy community
- You might have a future job or side project that is not in astronomy

Can you think of a situation where machine learning has played a role in your life? Was that interaction a positive or negative experience?

1 minute: think quietly
2 minutes: share with partner
5 minutes: share with group

Can you think of situations where an algorithm should not be used to support human decision making?

Algorithms are designed by people, and people embed (unconscious) biases in algorithms, data collection and interpretation of results.

Meredith Broussard: Artificial Unintelligence

Ethical Concerns in Machine Learning

- purposeful harm or manipulation
- technological solutionism
- ubiquity, consolidation, trust
- limited perspective of algorithm writer
- errors in algorithm's implementation
- unrepresentative data sets
- insufficient evaluation metrics
- faulty representations (e.g. in langauge)
- fairness
- privacy and data rights
- ...

Bernease Herman, Software Engineering for Data Scientists, http://uwseds.github.io

Deepfake Videos Are Getting Real and That's a Problem

The damaging impact and destabilizing potential of sophisticated video fakery

By Hilke Schellmann
Oct. 15, 2018 11:30 am

Computer-generated videos are getting more realistic and even harder to detect thanks to deep learning and artificial intelligence. As WSJ's Jason Bellini finds in this episode of Moving Upstream, these so-called deepfakes can be playful, but can also have real, damaging consequences for people's lives.

- **whichfaceisreal.com**

'I was shocked it was so easy': meet the professor who says facial recognition can tell if you're gay

Michal Kosinski: 'I don't believe in free will.' Photograph: Jason Henry/The Guardian

Psychologist Michal Kosinski says artificial intelligence can detect your sexuality and politics just by looking at your face. What if he's right?

How a Self-Driving Uber Killed a Pedestrian in Arizona

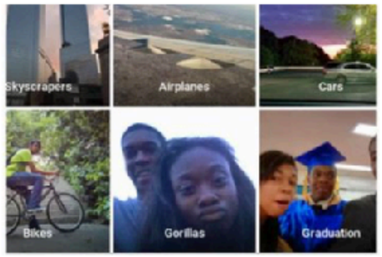
By **TROY GRIGGS** and **DAISUKE WAKABAYASHI** UPDATED MARCH 21, 2018

A woman was struck and killed on Sunday night by an autonomous car operated by Uber in Tempe, Ariz. It was believed to be the first pedestrian death associated with self-driving technology.

DIGITS

Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms

By Alistair Barr
Jul 1, 2015 3:40 pm ET



Black programmer Jacky Alciné said on Twitter that the new Google Photos app had tagged photos of him and a friend as gorillas. *JACKY ALCINÉ AND TWITTER*

Andrej Karpathy blog

About Hacker's guide to Neural Networks

What a Deep Neural Network thinks about your #selfie

Oct 25, 2015

Technochauvinism (n): a mindset that says that algorithms are superior to human judgment. The same mindset argues using technology is always the best strategy.

Automated Inference on Criminality using Face Images

Xiaolin Wu, Xi Zhang • Published in ArXiv 2016

We study, for the first time, automated inference on criminality based solely on still face images. Via supervised machine learning, we build four classif
using facial images of 1856 real persons controlled fo
nearly half of whom were convicted criminals, for disc
criminals. All four classifiers perform consistently wel
automated face... CONTINUE READING

VIEW PDF SAVE TO LIBRARY CREATE ALER

Is this a good idea?

Could this go wrong in some way?

https://www.semanticscholar.org/paper/Automated-In
1cd357b675a659413e8abf2eafad2a463272a85f

see also: https://callingbullshit.org/case_studies/case_study_criminal_machine_learning.html

“Unlike a human examiner/judge, a computer vision algorithm or classifier has absolutely no subjective baggages [sic], having no emotions, no biases whatsoever due to past experience, race, religion, political doctrine, gender, age, etc., no mental fatigue, no preconditioning of a bad sleep or meal. The automated inference on criminality eliminates the variable of meta-accuracy (the competence of the human judge/examiner) all together.”

(Wu & Zhang, 2016)

Do you agree?

The Data

Is this data set unbiased? In what ways could it be biased?

- 1856 Chinese men
- ages between 18 and 55
- no facial hair, scars or other markings
- non-criminals: photos acquired from the internet using web crawler (e.g. from company websites)
- criminals: ID photos published as wanted subjects, provided by police department
- algorithm classifies data set with 90% accuracy

One possible bias: attractive defendants are less likely convicted, or with less severe sentences, than unattractive defendants.

e.g. Gunnell + Ceci, 2010

Examples

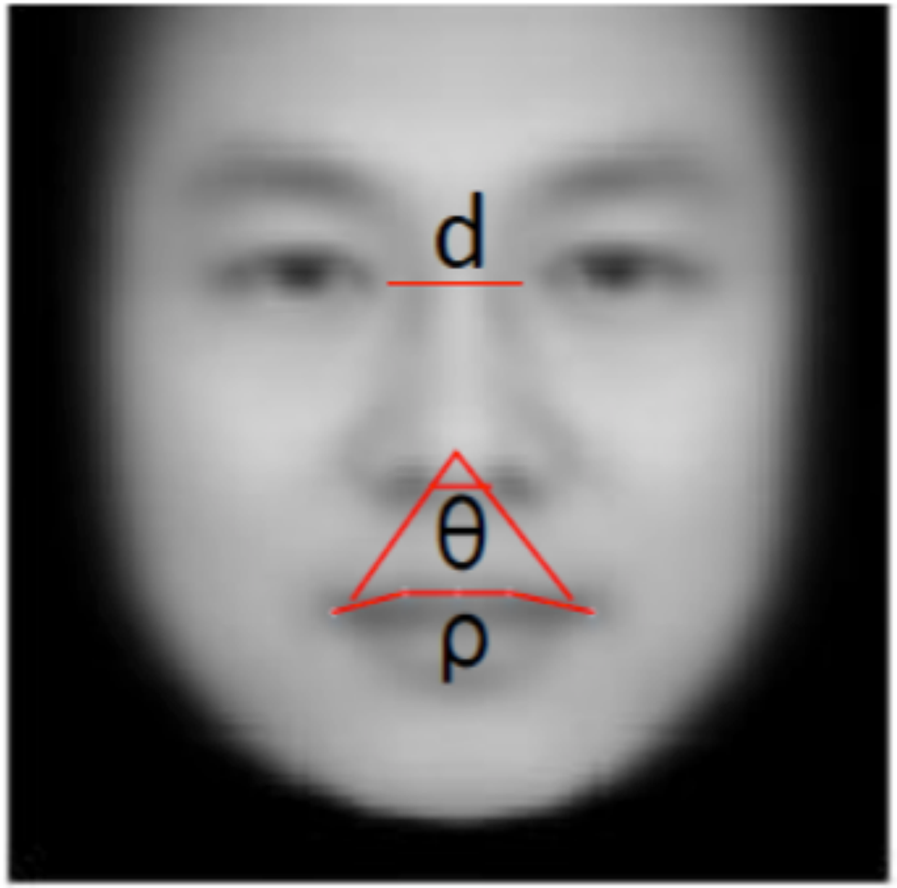
(a) Three samples in criminal ID photo set S_c.

criminal

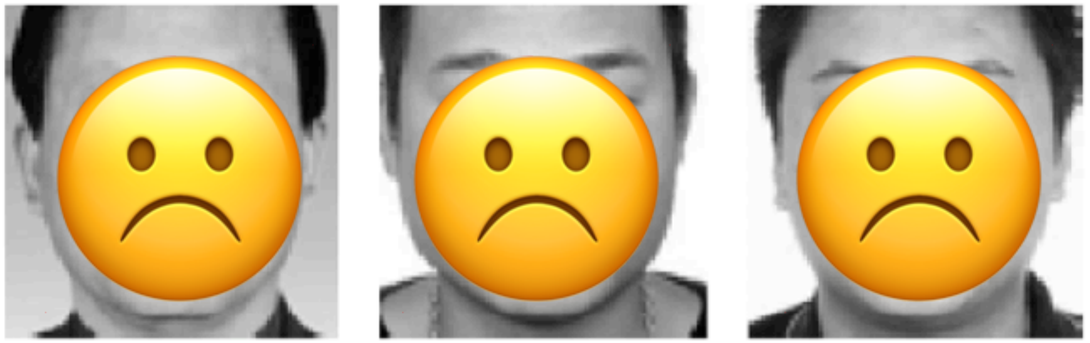
(b) Three samples in non-criminal ID photo set S_n

non-criminal

What features in an image are discriminative?



Examples



(a) Three samples in criminal ID photo set S_c.

(b) Three samples in non-criminal ID photo set S_n

This is a bad interpretation of a bad training data set!

Machine learning is only as good as its **training data set.**

In most cases, you can identify a **faulty application** of an algorithm solely by looking at the **training data** used and the **interpretation** of the results.

What makes the application of machine learning ethical?

Why is this important at all?

What can we do to ensure that we are applying machine learning ethically and responsibly?

Does data science need special ethical consideration?

Mostly, the answer is “no”.

Slide from: Laura Norén

DATAPRACTICES.ORG

SUBSCRIBE

Data Values & Principles

These values and principles, taken together, describe the most effective, ethical, and modern approach to data teamwork.

READ AND SIGN

Data Practices Courseware

Consume and collaborate on free and open courses designed to help everyone from the novice to the expert data practitioner.

EXPLORE COURSES

Checklists help eliminate basic mistakes

https://etherpad.wikimedia.org/p/ata-ml-ethics

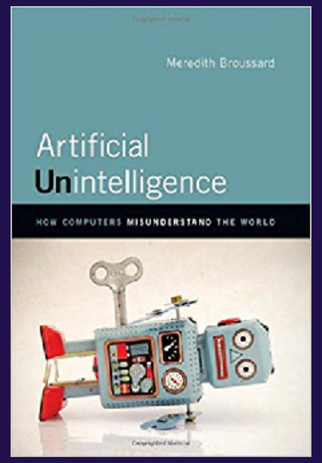
- ❑ Have we listed how this technology can be attacked or abused?
- ❑ Have we tested our training data to ensure it is fair and representative?
- ❑ Have we studied and understood possible sources of bias in our data?
- ❑ Does our team reflect diversity of opinions, backgrounds, and kinds of thought?
- ❑ What kind of user consent do we need to collect to use the data?
- ❑ Do we have a mechanism for gathering consent from users?
- ❑ Have we explained clearly what users are consenting to?
- ❑ Do we have a mechanism for redress if people are harmed by the results?
- ❑ Can we shut down this software in production if it is behaving badly?
- ❑ Have we tested for fairness with respect to different user groups?
- ❑ Have we tested for disparate error rates among different user groups?
- ❑ Do we test and monitor for model drift to ensure our software remains fair over time?
- ❑ Do we have a plan to protect and secure user data?

Mason, Loukides, Patil: Ethics and Data Science (2018)

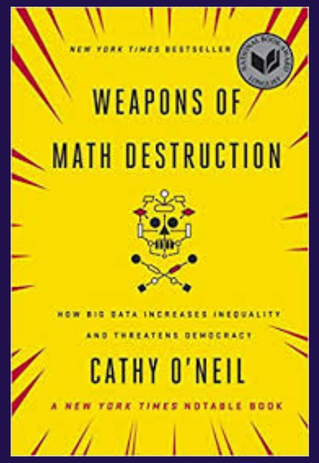
What can we do?

- keep up with the developments in data science and ethics
- think about the ethical implications of your project in advance
- make sure your team represents many different backgrounds and experiences
- take your university’s training in human subject research
- use check lists
- be open to feedback and criticism

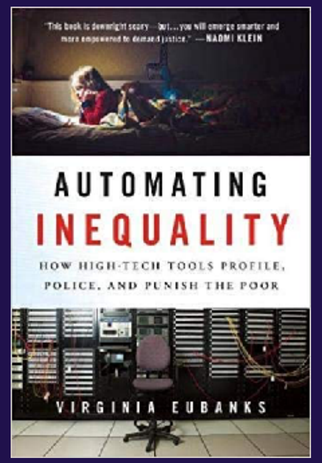
Resources



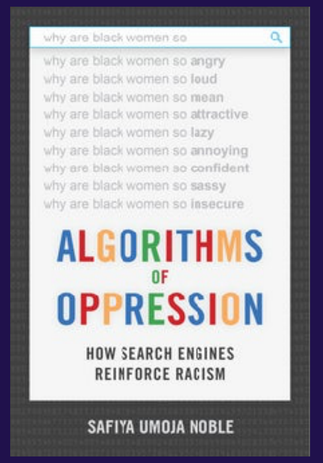
Meredith Broussard



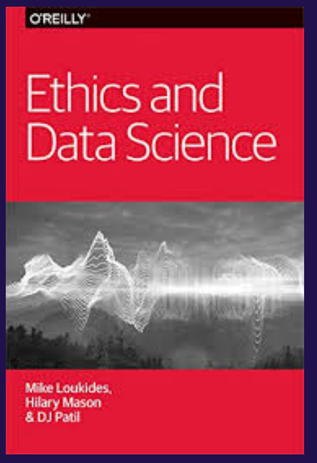
Cathy O'Neil



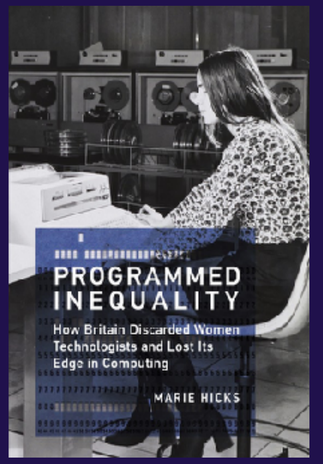
Virginia Eubanks



Safiya Umoja Noble



M Loudikes, H Mason, DJ Patil



Marie Hicks

More resources

- https://callingbullshit.org
- https://datapractices.org
- https://datasociety.net
- https://fatconference.org
- https://ainowinstitute.org
- https://www.datafordemocracy.org