Ethical Considerations in Data

UC Love Data Week
February 9, 2021

Introduction

- Who we are
  - Leigh Phan - Data Science Facilitator, UCLA Library Data Science Center
  - Stephanie Labou - Data Science Librarian, UC San Diego Library
  - Ibraheem Ali - Science Data Librarian, UCLA, Biomedical Library
  - Erin Foster - RDM Program Service Lead, UC Berkeley
- **We're not the experts** or have all the solutions, but are presenting concepts to consider in your own work
- Topics we'll cover

Workshop Overview

Introductions

Examples

- Data privacy
  - Issues of openness vs privacy
- Algorithmic bias
  - Considering bias in computational models
- Engaging communities in research
  - Centering communities & participants

Breakout Room Discussions

Reconvene & share discussion topics

Feedback survey

Guiding question:
Who is impacted by your research?

Data privacy

Why do we have to think about privacy?

BUSINESS

100 Million People In The U.S. Affected By Capital One Data Breach

July 30, 2019 4:23 PM ET
Heard on All Things Consid

Giant Equifax data breach: 143 million people could be affected

by Sara Ashley O'Brien @saraashleyo
September 8, 2017: 9:

Verizon data of 6 million users leaked online

by Selena Larson @selenalarson
July 12, 2017: 4:14 PM ET

Healthcare Data Leaks on GitHub: Credentials, Corporate Data and the PHI of 150,000+ Patients Exposed

Home | Cloud Computing in Healthcare | Healthcare Data Leaks on GitHub: Credentials, Corporate Data and the PHI of 150,000+ Patients Exposed

Marriott discloses new data breach impacting 5.2 million guests

The hotel group is sending emails to guests affected by the breach.

Carrie Mihalcik March 31, 2020 9:12 a.m. PT
LISTEN · 01:13

Why do we have to think about privacy?

Opinion

The Devastating Consequences of Being Poor in the Digital Age

When someone who is living paycheck to paycheck falls victim to an online fraud or a breach, the cascade of repercussions can be devastating.

By Mary Madden

Ms. Madden is a technology researcher.

April 25, 2019

2015 U.S. Transgender Survey (USTS) (ICPSR 37229)

Version Date: May 22, 2019 Cite this study | Share this page

Principal Investigator(s):

Sandy E. James, National Center for Transgender Equality (U.S.); Jody Herman, National C
Equality (U.S.); Lisa Mottet, National Center for Transgender Equality (U.S.); Ma'ayan Anafi,

Qual Health Res. 2017 May; 27(6): 893–908. PMCID: PMC5865471
Published online 2016 Sep 19. doi: 10.1177/1049732316669338 PMID: 27651071

Ethical and Safety Issues in Doing Sex Work Research: Reflections From a Field-Based Ethnographic Study in Kolkata, India

Sunny Sinha[1]

Data privacy - in general

We all know not to include information like full names or addresses in public/broadly accessible datasets

We may even know to “anonymize” data to some extent, removing or aggregating other potentially identifying variables

Range of privacy levels depend on the kind/type of data, from health data to informal online surveys

3 types of privacy considerations

Personally identifiable data

Linkable data

Statistical approaches to privacy

→ Ongoing tension of openness vs. privacy

Personally identifiable information - broader than you think!

Common PII definition (NIST): “PII is any information about an individual maintained by an agency, including (1) any information that can be used to **distinguish or trace an individual‘s identity**, such as name, social security number, date and place of birth, mother‘s maiden name, or biometric records; and (2) **any other information that is linked or linkable to an individual**, such as medical, educational, financial, and employment information.”

For example: full names, addresses, voice signatures, photographic images, and even IP addresses

There are also items that are not themselves PII, but when combined, could identify someone

https://csrc.nist.gov/glossary/term/PII#:~:text=personally%20identifiable%20information-,Personally%20Identifiable%20Information,SP%20800%2D79%2D2%20EGovAct

Can data ever be truly anonymized?

Article | Open Access | Published: 23 July 2019

Estimating the success of re-identifications in incomplete datasets using generative models

Luc Rocher, Julien M. Hendrickx & Yves-Alexandre de Montjoye

Nature Communications **10**, Article number: 3069 (2019) | Cite this article

117k Accesses | **78** Citations | **2604** Altmetric | Metrics

heavily incomplete dataset. On 210 populations, our method obtains AUC scores for predicting individual uniqueness ranging from 0.84 to 0.97, with low false-discovery rate. Using our model, we find that 99.98% of Americans would be correctly re-identified in any dataset using 15 demographic attributes. Our results suggest that even heavily sampled anonymized datasets are unlikely to satisfy the modern standards for anonymization set forth by GDPR and seriously challenge the technical and legal adequacy of the de-identification release-and-forget model.

https://www.nature.com/articles/s41467-019-10933-3/

Privacy vs. accuracy

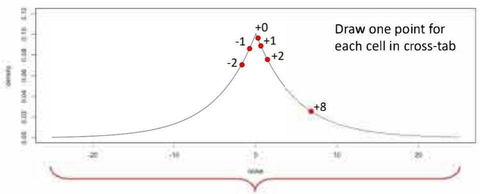
Differential privacy: statistical approach to privacy for small n, identifiable groups

Construct cross-tabs from “true” data

	School Attendance		
	Never	Attending	Past
Male	3	12	33
Female	4	17	31

Population = 100

Draw noise from Laplace distribution



spread is determined by ε

Add noise to cross-tab

	School Attendance		
	Never	Attending	Past
Male	3 - 1 = **2**	12 + 0 = **12**	33 + 1 = **34**
Female	4 + 8 = **12**	17 + 2 = **19**	31 - 2 = **29**

Sum = 108

Privacy in your own research - guiding questions

- What variables in your data may need to remain private/confidential?
  - Does meaningful analysis *require* full variables, or are some not necessary?

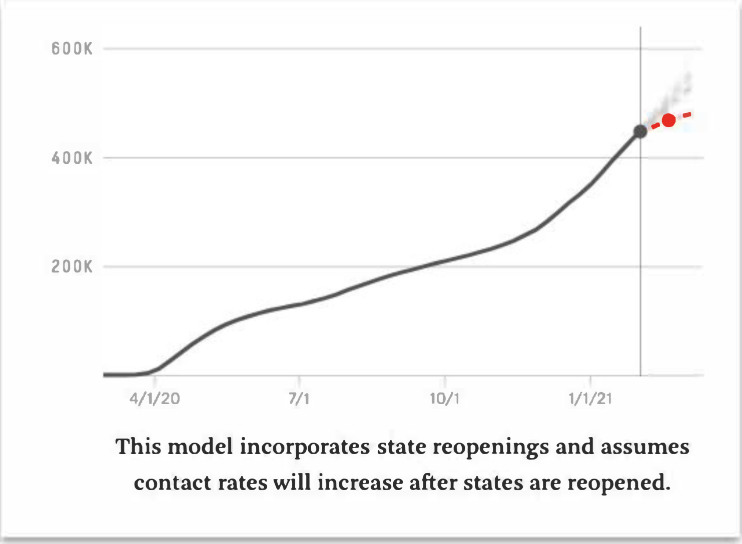
- What methodologies may be appropriate, based on your domain and research question?
  - Removing, aggregating, statistical approaches

- Openness vs. privacy scale in your discipline

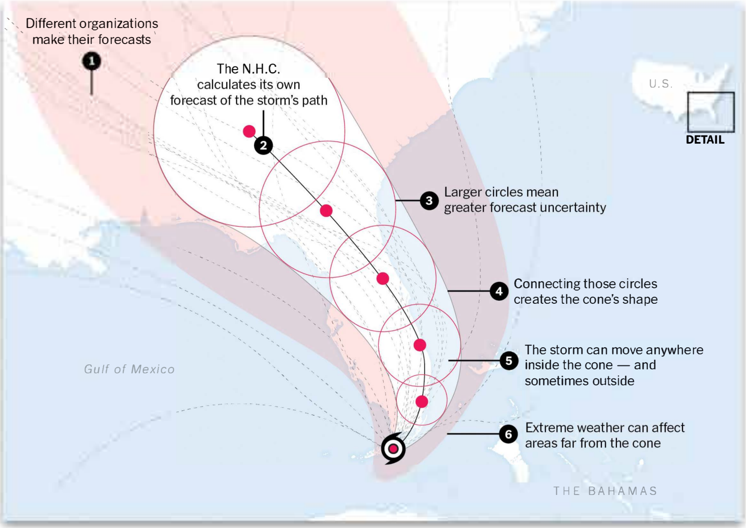
Algorithmic bias

Models are used to create predictions of natural phenomena

Infection Rate Prediction Models (COVID-19)

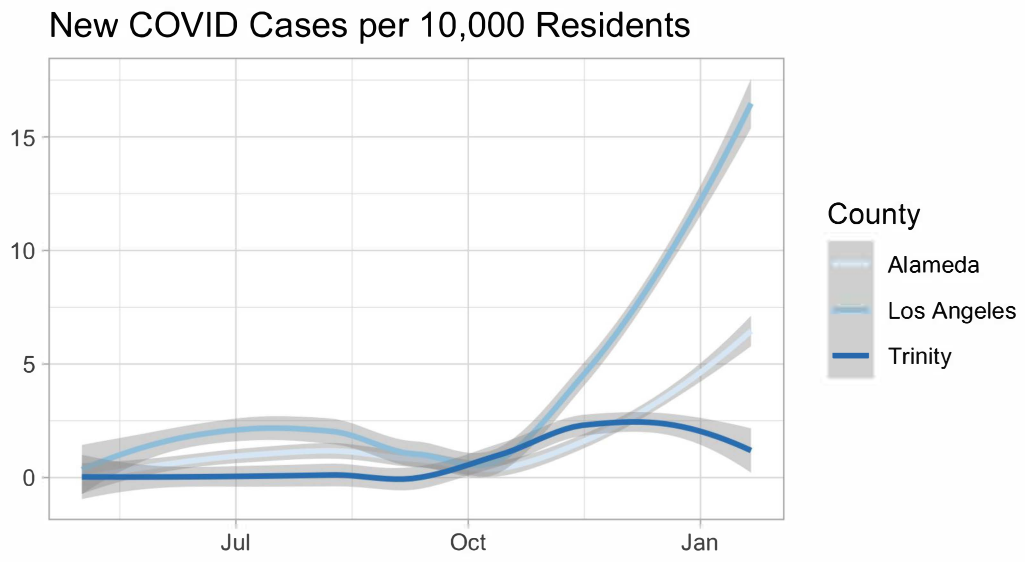


Hurricane Trajectory Models



COVID-19 Models, FiveThirtyEight, 2021
NYT Learning Network, Sep 11, 2019

Models rely on specific assumptions to create predictions



“A model’s blind spots reflect the judgments and priorities of its creators.”
- Cathy O’Neill

Incorrect assumptions lead to real world consequences

We will review the following examples:

- Bias in facial recognition software
- Algorithmic bias in health insurance financial allocation
- Predicting criminality with the COMPAS recidivism algorithm

COMPAS Recidivism Algorithm

COMPAS: Correctional Offender Management Profiling for Alternative Sanctions

- Software used to assess a defendant's risks of re-offending (within two years)
- ProPublica obtained the risk scores assigned to more than 7,000 people arrested in Broward County, Florida, in 2013 and 2014 and checked to see how many were charged with new crimes over the next two years
- High percentage of false positives and skewed toward Black defendants

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

Prediction Fails Differently for Black Defendants

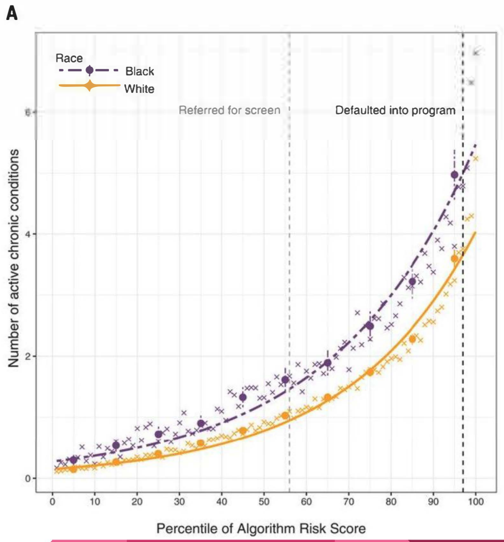
	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

Angwin et al., 2016. ProPublica.

Racial bias in a commercial risk-prediction tool for allocating health insurance support

- Goal: Provide additional support for people with significant healthcare needs
  - Give individual patients a ‘risk score’
  - Individuals with high risk score qualify for the program
  - Moderate risk scores are referred to a primary care physician
- Reality:
  - Due to barriers in healthcare access, Black patients had more health care issues than white patients at the same risk score.
  - Algorithm predicts healthcare costs instead of illness, leading to a significant disparity in the populations
  - Correcting for the bias increased the percentage of black people getting help from 17.7% to 46.5%



Obemeyer et al 2019. Science.

Facial Recognition - Gender Shades

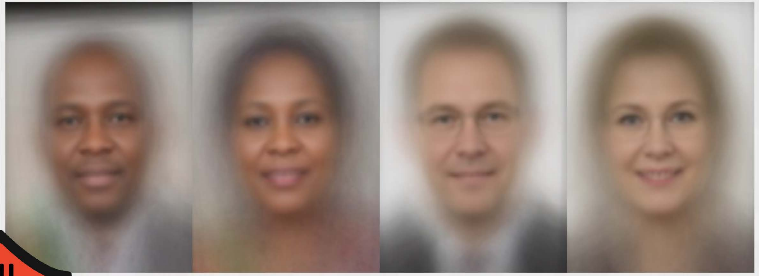
Evaluated accuracy of AI classification products

- Constructed dataset from 1,270 images of public officials from 3 African countries and 3 European countries
- Evaluated Microsoft, IBM, Face++

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

Results

- Overall high accuracy
- All companies performed better on male than female subjects and on lighter subjects as a whole
- Analysis by intersectional groups revealed all companies performed worst on darker-skinned female subjects



2019 Re-evaluation
Algorithmic Justice League

Buolamwini, J., Gebru, T. (2018). Proceedings of Machine Learning Research.

Facial Recognition - FairFace

Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation

- Constructed a novel face image dataset, containing 108,501 images from Yahoo! Flickr CC-100M dataset
- Defined 7 race groups: White, Black, Indian, East Asian, Southeast Asian, Middle East, and Latino

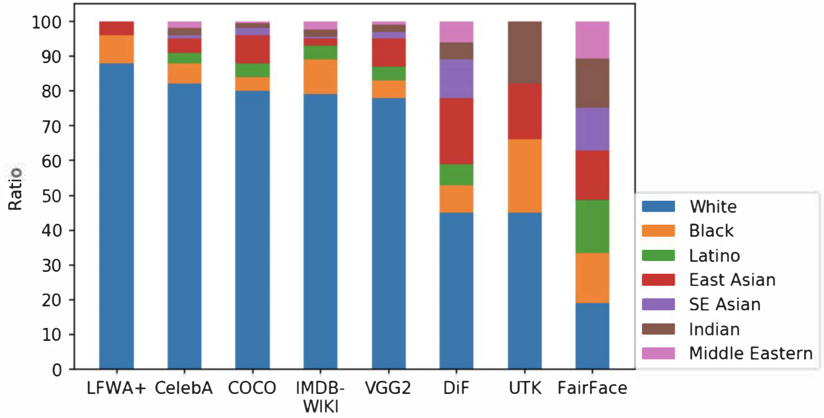


Figure 1: Racial compositions in face datasets.

Karkkainen, K., & Joo, J. (2021).

Facial Recognition - Amazon Rekognition

- Has claimed to be able to detect emotions on faces of subjects (Amazon Web Services, 2021).
- In 2018, mistakenly matched 28 Congress members to criminal mugshots; disproportionately misidentified members of color (Snow, 2018).
- Has been sold to law enforcement agencies and U.S. Immigration and Customs Enforcement (ICE) (Shaban, 2018).
- June 2020: Announced moratorium on selling Rekognition to law enforcement (Allyn, 2020).

During June 2020, IBM also announced it would sunset its facial recognition products and no longer provide to police departments (Allyn, 2020).

Returning to our guiding question:

Who is impacted by your research?

- What is the right question? Accuracy or privacy?
  - Those impacted by both COMPAS and health insurance risk assessment were marginalized, predominantly Black communities
- Consent, awareness, engagement

Engaging communities in research

Center communities in research

How to include communities in research that impacts them?

Will focus on three areas:

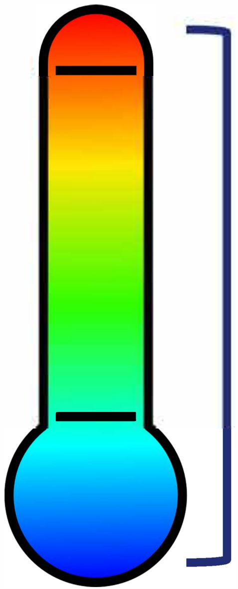
- Reporting back to study participants
- Leveraging community expertise
- Enabling broader social change

UC Santa Cruz. The Merrill Moat - "Build Trust Not Walls."
Artists: Ana Camarena & Ademir Cedillos. Year: 2007, Mural #24.

Reporting back: For Healthy Kids! Project

A 10+ year study investigating pesticide exposure in farmworkers and nonfarmworkers in an agricultural community.

- Used a community based participatory research approach to engage with the community.
- Asked community about best way to communicate back results:
  - Community promoters who previously collected samples/questionnaires from study participants later disseminated results.
  - Identified a graphic as most effect method to communicate the exposure level & risk.



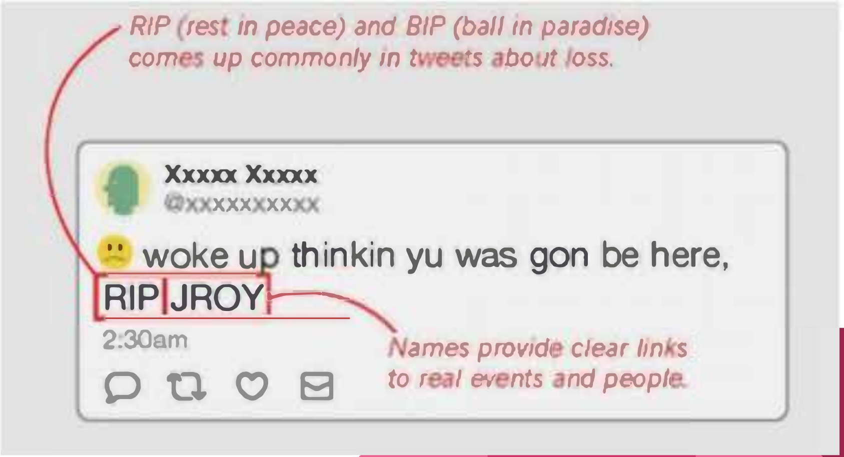
Settled on thermometer graphic as best visual way to communicate pesticide exposure data.

Thompson, Beti PhD; Carosso, Elizabeth BA; Griffith, William PhD; Workman, Tomomi BS; Hohl, Sarah MPH; Faustman, Elaine PhD.Journal of Occupational and Environmental Medicine: October 2017 - Volume 59 - Issue 10 - p 982-987 doi: 10.1097/JOM.0000000000001107

Context matters: SAFE Lab

Research focusing on social media communication amongst gang-involved/affiliated youth.

- Worked with domain experts to create community-informed algorithms.
- Follow a code of ethics, including:
    - Transparency at all levels
    - Explore new ways to document consent
    - Not sharing data with groups that use “punitive and criminalizing methodologies”

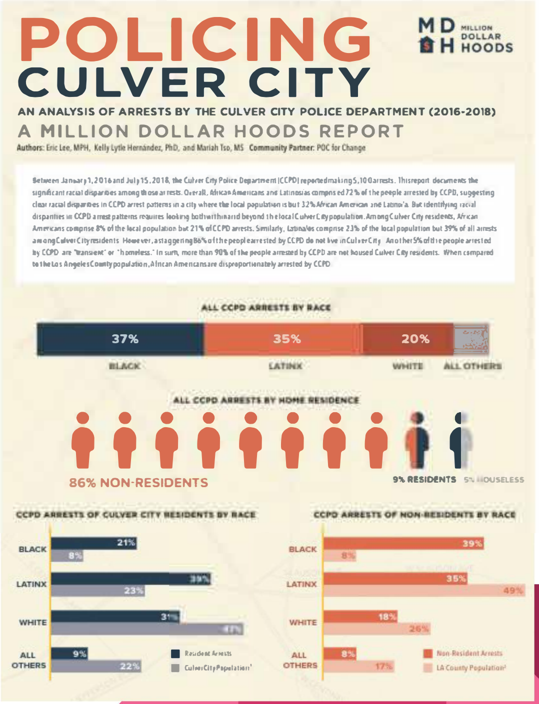


Frey, W. R., Patton, D. U., Gaskell, M. B., & McGregor, K. A. (2018). Artificial intelligence and inclusion: Formerly gang-involved youth as domain experts for analyzing unstructured twitter data. *Social Science Computer Review*, 0894439318788314.

Social change: Million Dollar Hoods Project

A community-based research initiative focusing on the human and fiscal costs of mass incarceration.

- Work with local advocates and activists to guide research.
- Sponsor the Big Data for Justice Summer Institute that trains students & community members.
- Inform policy reform undertaken by the Los Angeles Social Equity Program.



Lee, E., Lytle Hernández, K., and Tso M. (2020). "Policing Transitional-Aged Youth in Culver City: An Analysis of Arrests by the Culver City Police Department 2016-2018". Los Angeles, CA. The Million Dollar Hoods Project.

Final thoughts & guiding questions

What other topics arise when engaging communities and data?

- Data governance, data sharing, much more!

Continue the conversation - how do we take ownership and push this forward in our research?

Photo by LOGAN WEAVER on Unsplash

Breakout rooms

Reporting out

Additional resources

References & further reading

- Allyn, B. (2020, June 11). Amazon Announces 1-Year Moratorium On Police Use Of Its Facial-Recognition Technology. NPR. https://www.npr.org/2020/06/11/875311072/amazon-announces-1-year-moratorium-on-police-use-of-its-facial-recognition-techn
- Allyn, B. (2020, June 9). IBM Abandons Facial Recognition Products, Condemns Racially Biased Surveillance. https://www.npr.org/2020/06/09/873298837/ibm-abandons-facial-recognition-products-condemns-racially-biased-surveillance
- Amazon Web Services, Inc. (2021). Emotion - Amazon Rekognition Documentation. https://Docs.Aws.Amazon.Com/Rekognition/Latest/Dg/API_Emotion.Html
- Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code* (1st ed.). Polity.
- Buolamwini, J., Gebru, T. (2018). "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of Machine Learning Research 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf
- Carroll, S.R., Rodriguez-Lonebear, D. and Martinez, A., 2019. Indigenous Data Governance: Strategies from United States Native Nations. *Data Science Journal*, 18(1), p.31. DOI: http://doi.org/10.5334/dsj-2019-031
- D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. The MIT Press.
- Karkkainen, K., & Joo, J. (2021). FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1548-1558).

References & further reading (cont'd)

- Kearns, M., & Roth, A. (2019). *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford University Press.
- Lar-Son, K. (2021, February 2). *Data as Relation: Indigenous Data Sovereignty and Ethic of Care* [Video]. YouTube. https://www.youtube.com/watch?v=QGYse9iDPWI
- Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism* (Illustrated ed.). NYU Press.
- Obermeyer et al. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. Doi: https://doi.org/10.1126/science.aax2342
- O'Neil, C. (2017). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Reprint ed.). Crown.
- ProPublica. (2016). How We Analyzed the COMPAS Recidivism Algorithm. ProPublica. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
- Shaban, H. (2018, June 22). Amazon employees demand company cut ties with ICE. Washington Post. https://www.washingtonpost.com/news/the-switch/wp/2018/06/22/amazon-employees-demand-company-cut-ties-with-ice/
- Snow, J. (2018, August 3). Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots. American Civil Liberties Union. https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

Leading Organizations

- Algorithmic Justice League
- Black in AI
- GenderShades.org
- UC San Diego Institute for Practical Ethics
- UC Berkeley Algorithmic Fairness and Opacity Group (AFOG)
- UCLA C2i2 - Center for Critical Internet Inquiry

Thank you for joining the conversation!

Please take our feedback survey!