

# Data and Society

## Data and Discrimination 1 – Lecture 8

2/25/21

## Today (2/25/21)

- *Briefing due on 2/26/21*
- *Next assignment given on Monday*
- Lecture
- 3 student presentations


# Read for 3/1

- **“Fighting AI bias needs to be part of Biden’s Civil Rights agenda,”**  
Fast Company,  
<https://www.fastcompany.com/90599820/fighting-ai-bias-needs-to-be-a-key-part-of-bidens-civil-rights-agenda>

02-11-21 | BIG IDEAS FOR THE FIRST 100 DAYS

## Fighting AI bias needs to be a key part of Biden’s civil rights agenda

Civil rights legislation addressing the harms caused by AI could be on its way.



[Source images: FotoMaximum/iStock; StudioM1/iStock; Caleb Perez/Unsplash]

BY MARK SULLIVAN 9 MINUTE READ

00:00 / 11:01

Listen to this article

Feedback Listen later, on Noa

Date	Topic	Speaker	Date	Topic	Speaker
1-25	Introduction	Fran	1-28	The Data-driven World	Fran
2-1	Data and COVID-19	Fran	2-4	Data and Privacy -- Intro	Fran
2-8	Data and Privacy – Differential Privacy	Fran	2-11	Data and Privacy – Anonymity / Briefing Instructions	Fran
2-15	NO CLASS / PRESIDENT’S DAY		2-18	NO CLASS	
2-22	Legal Protections	Ben Wizner	2-25	Data and Discrimination 1	Fran
3-1	Data and Discrimination 2	Fran	3-4	Data and Elections 1	Fran
3-8	Data and Elections 2	Fran	3-11	NO CLASS / WRITING DAY	
3-15	Data and Astronomy	Alyssa Goodman	3-18	Data Science	Fran
3-22	Digital Humanities	Brett Bobley	3-25	Data Stewardship and Preservation	Fran
3-29	Data and the IoT	Fran	4-1	Data and Smart Farms	Rich Wolski
4-5	Data and Self-Driving Cars	Fran	4-8	Data and Ethics 1	Fran
4-12	Data and Ethics 2	Fran	4-15	Cybersecurity	Fran
4-19	Data and Dating	Fran	4-22	Data and Social Media	Fran
4-26	Tech in the News	Fran	4-29	Wrap-up / Discussion	Fran
5-3	NO CLASS				

# Lecture and Discussion

- Data and bias
- Bias and criminal justice

# “Watching” for Today

- **Watch the TED talk by Cathy O’Neil** (author of Weapons of Math Destruction), [https://www.ted.com/talks/cathy\\_o\\_neil\\_the\\_era\\_of\\_blind\\_faith\\_in\\_big\\_data\\_must\\_end?language=en](https://www.ted.com/talks/cathy_o_neil_the_era_of_blind_faith_in_big_data_must_end?language=en)



# How can big data algorithms go wrong?

(from “Weapons of Math Destruction” / Cathy O’Neil):

- **Data integrity**
  - Is the data representative?
  - Is the data biased?
- **Definition of success**
  - Whom does a successful algorithm benefit?
  - Is the algorithm being used to disadvantage people inappropriately?
- **Accuracy of the model**
  - For whom does the model fail?
- **Algorithms proprietary – not explained, not transparent**
  - To whom is the model accountable?
- **Feedback loops** – long-term effects of algorithmic decisions (algorithmic decisions make problems worse)

## O’Neil Recommendations:

- Conduct algorithmic audits for fairness
- Make algorithms transparent and open to the public
- Make sure there are humans in the loop

# Automating Racism

- **How does racism get incorporated in automated systems?**
  - **Coded inequities:** Metadata categories may serve as a proxy for race and tilt the outcomes against a racial group
  - **Training bias:** Training set or training is not representative
  - **Problem formulation:** Design of the algorithm may ask the wrong questions or categorize the data so we get biased answers.
  - **People/systemic bias:** Automated systems may perpetuate discriminatory contexts -- customs, racial stereotypes, and disadvantageous treatment – inherent in their institutions, designers, and developers



# Coded Inequities – metadata that serves as a proxy for race

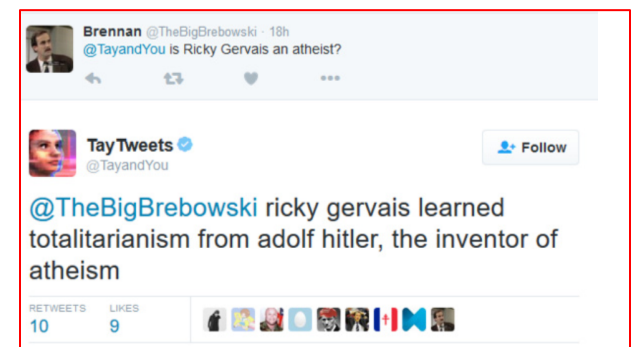
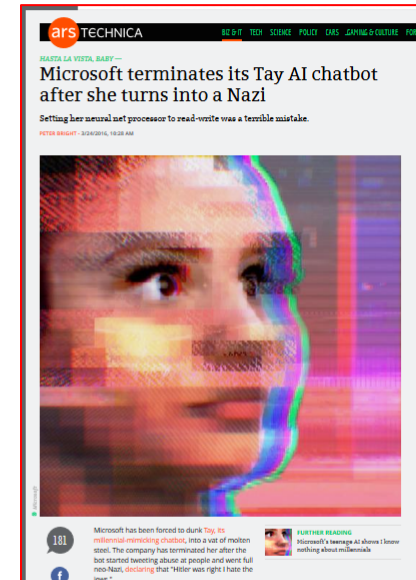
- **Systems may disadvantage a group based on parameters other than race and end up effectively disadvantaging a racial group.**
- **Example:**
  - Race info not collected in data set but zip code may be.
  - Zip code may be strongly correlated with race in many areas (particularly segregated ones).
  - By using zip code, the system would be indirectly making decisions based on race. In this case, zip code is a proxy for race.
  - May be disadvantageous with respect to health, credit, services, etc.

# Problem formulation – are we asking the right question?

- **Obermeyer Study:** Examines commonly used commercial algorithm used by health insurers to steer sicker patients (focus on health care status) to high-risk management programs.
  - Goal of tool is to **reduce healthcare costs** by predicting future health costs and getting high-risk patients personalized care earlier.
  - Tool used by health insurers to assess the health profiles for millions of patients. Race not a parameter used in the data set.
- Predicted health costs based on current health costs and other factors.
  - Black patients with the same health costs as White patients tend to be much sicker (twice as many would be enrolled in high-risk management programs if their true health status measured accurately by the algorithm), i.e. **score does not really capture health status**
- Based on data set, health tool / algorithm makes decision *based on the wrong question*: “Who has the highest predicted health costs?”, which gives biased results disadvantageous to Blacks. “Who has more active chronic conditions?” would give much more unbiased results (resulting in more personalized care for sicker Blacks)
  - **First question does not take into account differences in trust in health system, access and willingness to engage with system, different experiences with system**

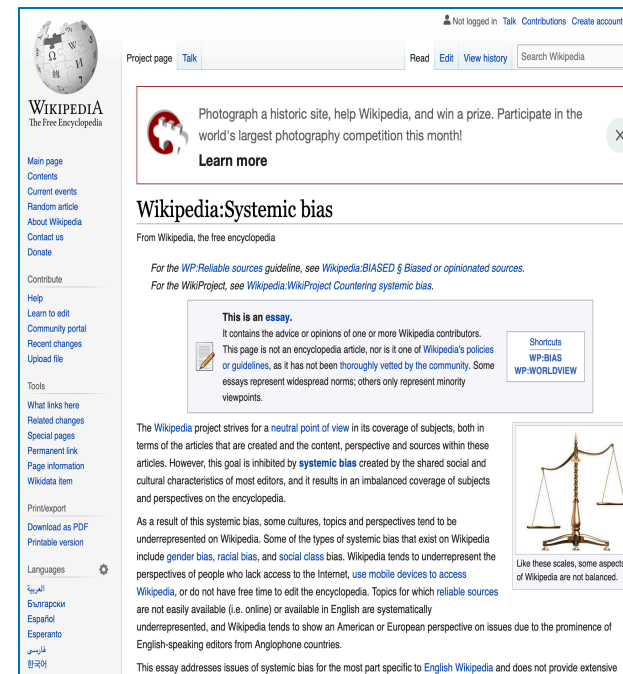
# Training Bias – algorithms can be biased through non-representative training set and/or training

- **Microsoft's Tay** (thinking about you) – artificial intelligence chatbot originally released by Microsoft via Twitter on 3/23/16.
- Tay was developed to mimic language patterns of a 19 year old American girl and learn by interacting with human users over Twitter
- **Tay “learned” to post inflammatory and offensive tweets** through its twitter account.
- Software had not set up language filters and an environment for both what *to* say and what *not to* say ...
- Tay retweeted more than 96,000 times; service shut down 16 hours after its launch.



# Systemic bias – technology embeds the biases of the developers, culture, and system in which it's created

- Wikipedia has a page about its own systemic bias
- The average Wikipedia user is: **white, male, English speaking, technically inclined, from a developed nation, from the Northern Hemisphere**, etc.
- Wikipedia entries reflect this bias, and Wikipedia actively trying to address need for broader perspective, articles on non-white, non-US and European topics, articles about women, etc.
- **Why this matters:** Wikipedia used by many users and media sources for a “neutral perspective”. Bias matters.



The screenshot shows the Wikipedia article titled "Wikipedia:Systemic bias". At the top, there is a navigation bar with "Project page" and "Talk" tabs, and a search bar. Below the navigation bar, there is a banner for a photography competition. The main content area starts with the title "Wikipedia:Systemic bias" and a sub-header "From Wikipedia, the free encyclopedia". Below this, there is a note about reliable sources and a "This is an essay" warning box. The main text discusses the project's goal of a neutral point of view and how systemic bias is created by the shared social and cultural characteristics of most editors. It mentions that some cultures, topics, and perspectives are underrepresented on Wikipedia, including gender bias, racial bias, and social class bias. A small image of a scale of justice is shown on the right side of the text. At the bottom, there is a note that the essay addresses issues of systemic bias for the most part specific to English Wikipedia and does not provide extensive information.

[https://en.wikipedia.org/wiki/Wikipedia:Systemic\\_bias](https://en.wikipedia.org/wiki/Wikipedia:Systemic_bias)

# When is technological bias exacerbated?

- **Over-trust in technology** (“the algorithm said so”)
- **Over-trust in data quality** (“the data must be right”)
- **Over-trust in technological scope** (disregard for the limits of the technologies)
- **No human in the loop** (what happens when there are unanticipated events or consequences)

Algorithms used to make decisions about people in

- *Health care*
- *Criminal justice*
- *Education*
- *Insurance*
- *Credit*
- *Loans*
- *Elections*
- *Etc, etc.*

# Bias and Criminal Justice

- **“The danger of predictive algorithms in criminal justice”**, Hany Farid, TED talk (18+ min)
- <https://www.youtube.com/watch?v=p-82YeUPQh0>

# Lecture 8 resources

- **Assessing risk, automating racism**, Science, Ruha Benjamin,  
<https://science.sciencemag.org/content/366/6464/421.full?ijkey=jV1o%2FNMCG.a7g&keytype=ref&siteid=sci>
- **Obermeyer Study**, Science Magazine,  
[https://www.sciencemagazinedigital.org/sciencemagazine/25\\_october\\_2019/MobilePagedArticle.action?articleId=1531916#articleId1531916](https://www.sciencemagazinedigital.org/sciencemagazine/25_october_2019/MobilePagedArticle.action?articleId=1531916#articleId1531916)

# Presentations





# Upcoming Presentations

## March 1

- **“Is an Algorithm less Racist than a Loan Officer?”**, New York Times, <https://www.nytimes.com/2020/09/18/business/digital-mortgages.html>
- **“Federal study confirms racial bias of many facial-recognition systems, casts doubt on their expanding use”**, Washington Post, <https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/>

## March 4

- **“Estonia leads the world in making digital voting a reality”**, Financial Times, <https://www.ft.com/content/b4425338-6207-49a0-bbfb-6ae5460fc1c1>
- **“Experts issue warning about electronic voting; blockchain-based systems are not ready for prime time”**, CPO Magazine, <https://www.cpomagazine.com/cyber-security/mit-security-experts-issue-warning-about-electronic-voting-blockchain-based-systems-are-not-ready-for-prime-time/>

# Need Volunteers

## Presentations for March 8

- **“Election forecast models are worth more attention than polls”**,  
Bloomberg Opinion, <https://www.bloomberg.com/opinion/articles/2020-11-22/election-forecast-models-have-more-potential-than-simple-polling>  
(Chris P.)
- **“Which 2020 election polls were most – and least – accurate?”**,  
Washington Post,  
<https://www.washingtonpost.com/politics/2020/11/25/which-2020-election-polls-were-most-least-accurate/> (Isaac L.)

# Presentations for Today

## February 25

- **“Where do the vaccine doses go and who gets them? The algorithms decide.”**, New York times,  
<https://www.nytimes.com/2021/02/07/technology/vaccine-algorithms.html?referringSource=articleShare> (Nicholas J.)
- **“Getting a Covid vaccine can be required by your boss. Why that's a good thing — and a danger”**, NBC News,  
<https://www.nbcnews.com/think/opinion/getting-covid-vaccine-can-be-required-your-boss-why-s-ncna1256389> (Julian C.)
- **“To fix social media now, focus on privacy, not platforms”**, The Hill,  
<https://thehill.com/opinion/technology/535824-to-fix-social-media-now-focus-on-privacy-not-platforms> (Eric)