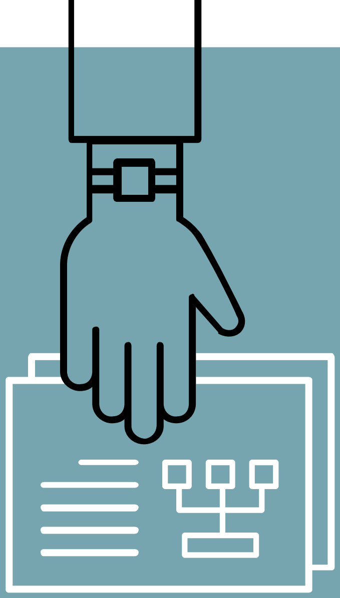
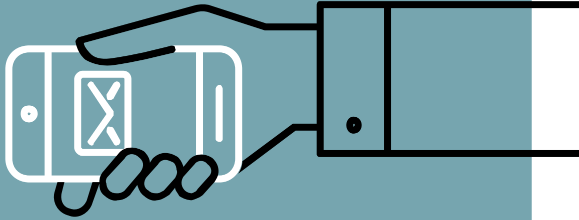
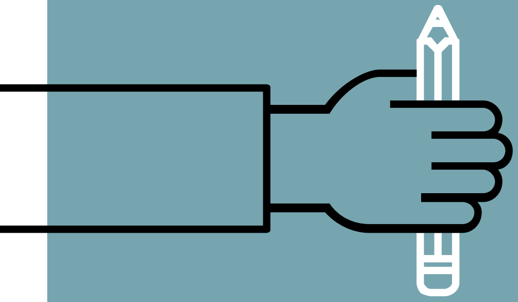
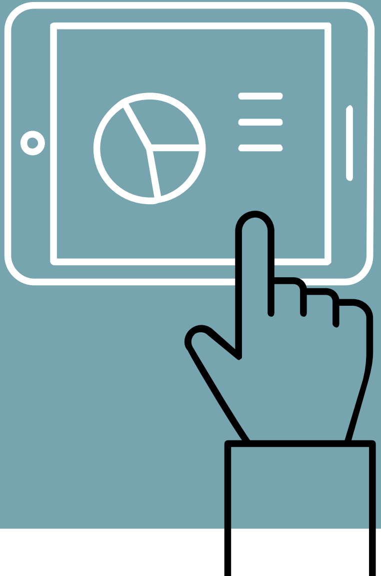



Algorithmic Fairness in Practice

ENAR Tutorial
March 15, 2021
Heather Mattie

Dr. Heather Mattie

Lecturer on Biostatistics

Co-Director | Health Data Science Master's Program

Harvard T.H. Chan School of Public Health

Email: hemattie@hsph.harvard.edu

Health Data Science

> Health Data Science > Program

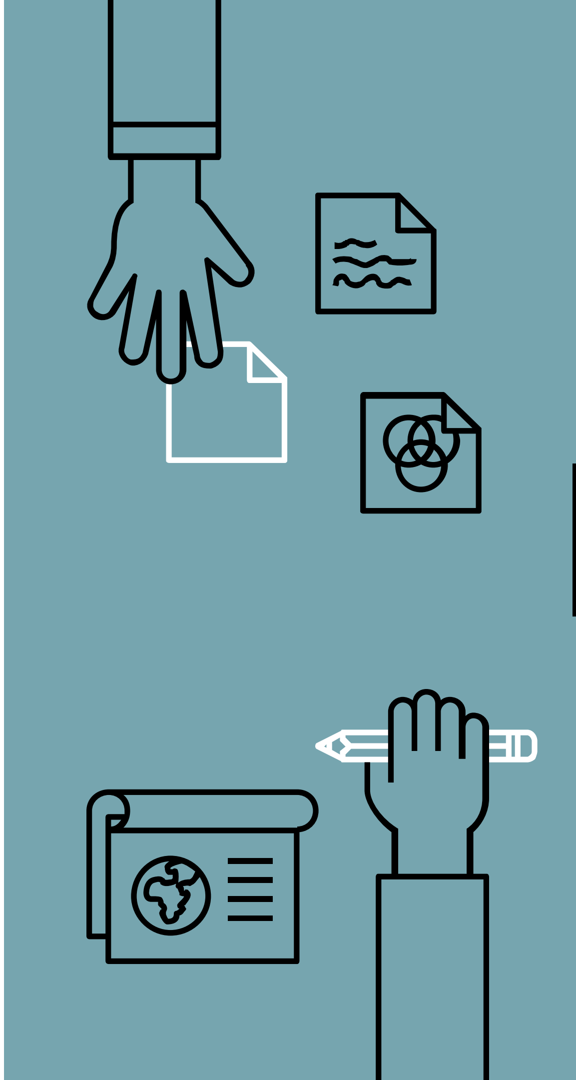
HEALTH DATA SCIENCE

Search this section

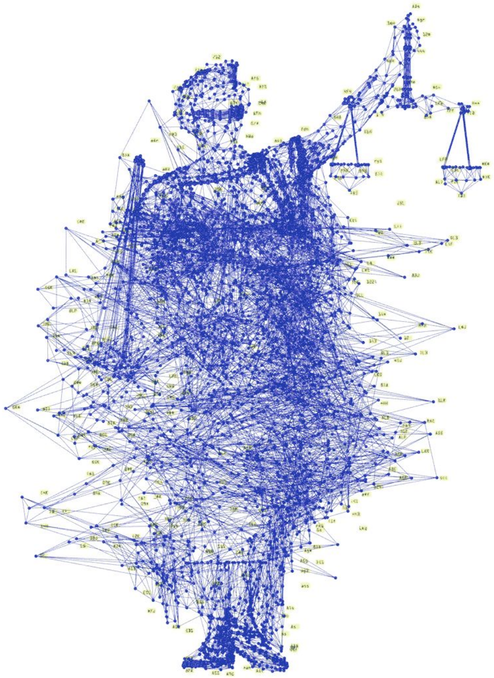
- **Program**
  - Student Handbook
- Program Faculty
- Prospective Students
- FAQs
- How to Apply
- Contact
- Resources

Program

https://www.hsph.harvard.edu/health-data-science/



What is *algorithmic fairness/bias?*



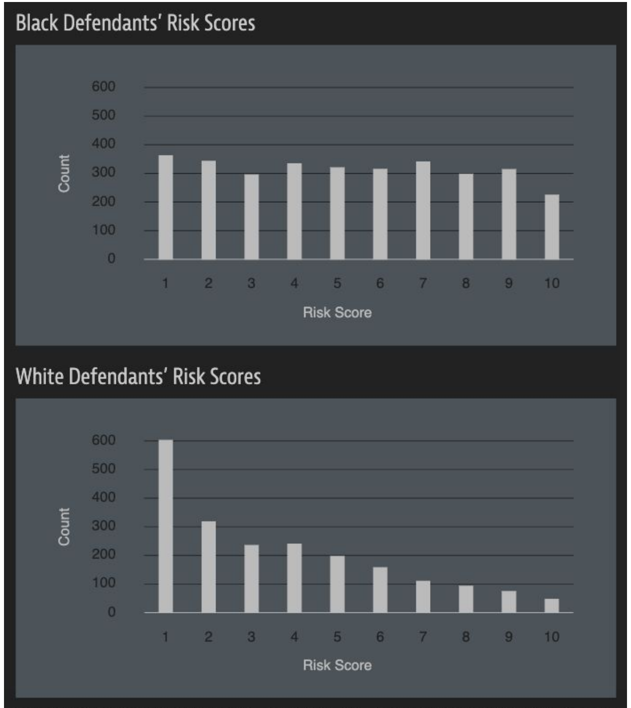
“Algorithmic bias describes **systematic and repeatable errors** in a computer system that create **unfair outcomes**, such as privileging one arbitrary group of users over others” - [Wikipedia]

Machine Bias

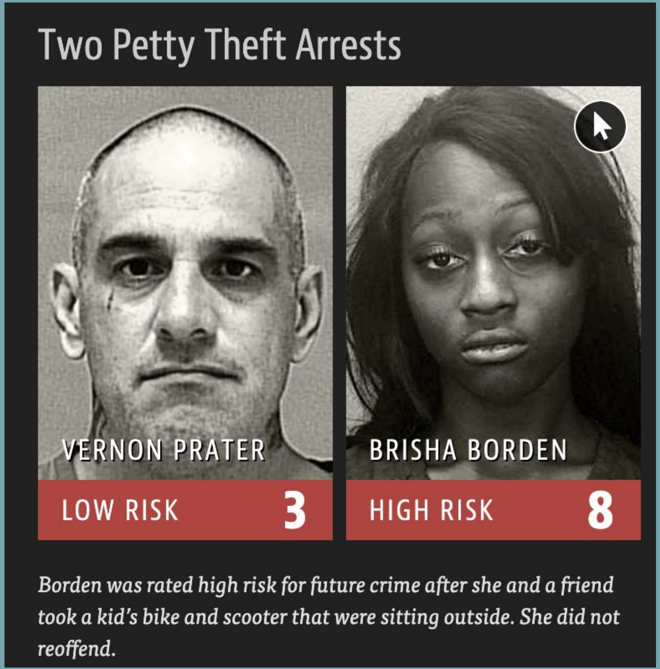
There's software used across the country to predict future criminals. And it's biased against blacks.

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

May 23, 2016



These charts show that scores for white defendants were skewed toward lower-risk categories. Scores for black defendants were not. (Source: ProPublica analysis of data from Broward County, Fla.)



Source:https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

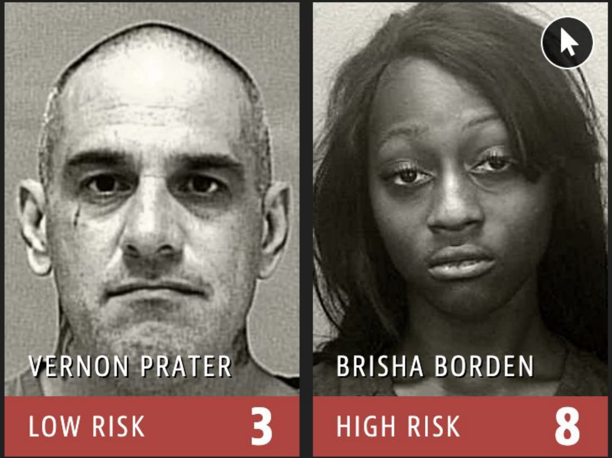
Machine Bias

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

Two Petty Theft Arrests



Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

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%

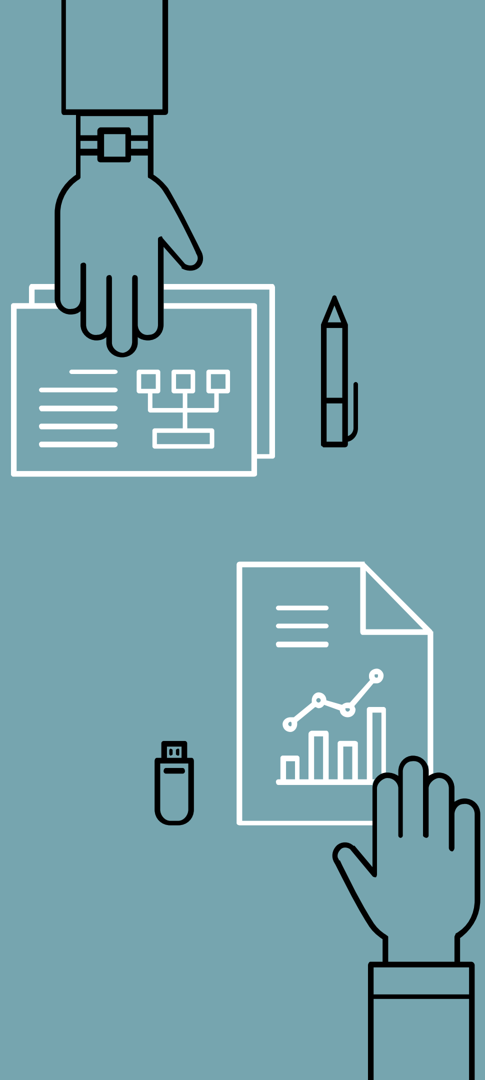
Source: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Perpetuating Gender-Based Employment Discrimination

Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs

Amazon scraps secret AI recruiting tool that showed bias against women



Sources:

https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study

TECHNOLOGY

How Algorithms Can Bring Down Minorities' Credit Scores

Analyzing people's social connections may lead to a new way of discriminating against them.

> A German company called Kreditech, for instance, asks loan applicants to share information from their social-media networks, which they can comb for details about their friends. Being connected to someone who's already paid back a loan with the company is "usually a good indicator," the company's chief financial officer told *Financial Times*.

> In India and Russia, FICO—the company behind the popular FICO credit scores—is partnering with startups like Lenddo to capture information about users from their cellphones. Lenddo uses locations reported by applicants' phones to figure out whether they really live and work where they say they do, and then analyzes an applicant's network to figure out "if they are in touch with other good borrowers—or with people with long histories of fooling lenders," *Bloomberg* reports.

Source:
https://www.theatlantic.com/technology/archive/2016/12/how-algorithms-can-bring-down-minorities-credit-scores/509333/

What is algorithmic bias in *healthcare*?

"Algorithmic bias occurs when the application of an algorithm compounds existing inequities in socioeconomic status, race, ethnic background, religion, gender, disability or sexual orientation to amplify them and adversely impact equity in health systems."

[Artificial intelligence and algorithmic bias: implications for health systems]

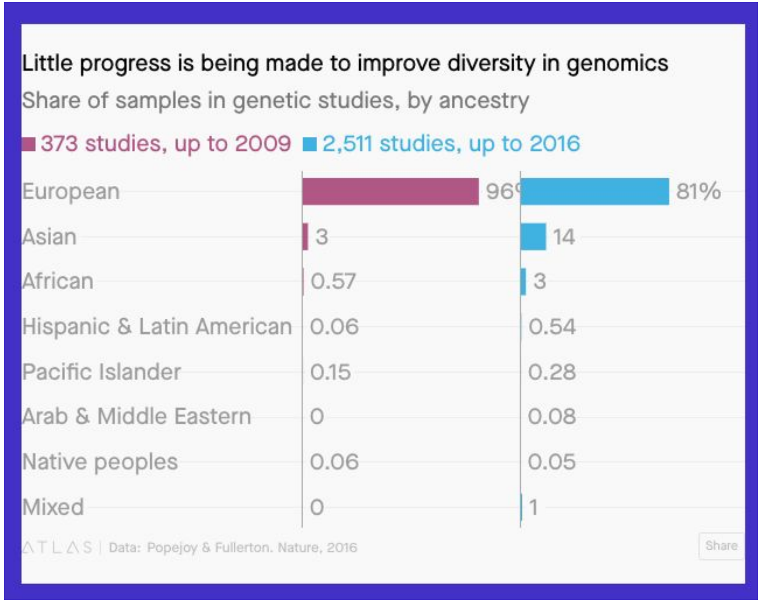
How algorithms can create inequality in health care, and how to fix it

by Matt Wood, University of Chicago

analyze data in the electronic medical records. One of the projects they're working on is to help decrease the length of stay for patients, because it's in everyone's best interest to have patients go home as soon as they're ready to leave. The thought was if we can identify patients who are most likely to be discharged early, we can assign a case manager to make sure there are no further blockages or barriers that could prevent them from leaving the hospital in a timely manner.

The data analytics group initially developed the algorithms based on clinical data, and then they found that adding the zip code where the patient lives improved the accuracy of the model identifying those people who would have shorter lengths of stay. The problem is when you add a zip code, if you live in a poor neighborhood or a predominantly African-American neighborhood, you were more likely to have the longer length of stay. So, the algorithm would have led to the paradoxical result of the hospital providing additional case management resources to a predominantly white, more educated, more affluent population to get them out of the hospital earlier, instead of to a more socially at-risk population who really should be the ones that receive more help.

Source:https://medicalxpress.com/news/2018-12-algorithms-inequality-health.html



81% of participants in genome-mapping studies were of European descent.

There is some precedent for the government to step in to ensure diversity during the data-gathering phase of new health policies and medical treatments. In 1993, the US Congress compelled the National Institutes of Health to bring more diversity to the medical studies it funded. It's not clear Congress or the NIH can solve this problem alone; more than 20 years later, 81% of genomics research is still from those of European descent. And furthermore, a 2015 study found that 2% of the more than 10,000 NIH-funded cancer studies include enough minority groups to be statistically significant. The study points to multiple potential causes, including inadvertent incentives in the NIH's funding structure, but the simplest is a lack of diversity in the medical field itself, and the propensity for non-white researchers to be funded less often.

Source:https://qz.com/1367177/if-ai-is-going-to-be-the-worlds-doctor-it-needs-better-textbooks/

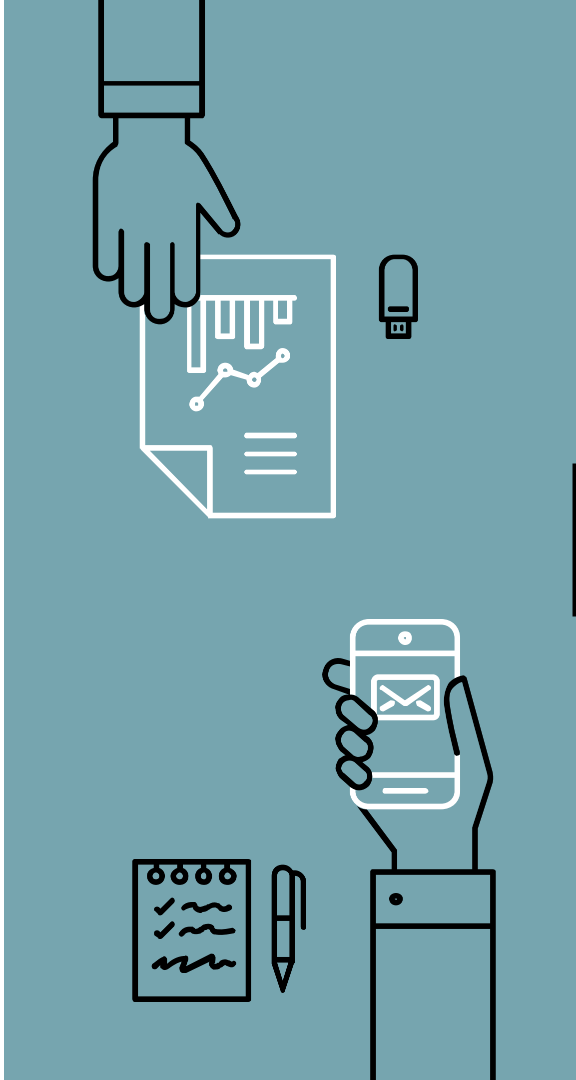
ECONOMICS

Dissecting racial bias in an algorithm used to manage the health of populations

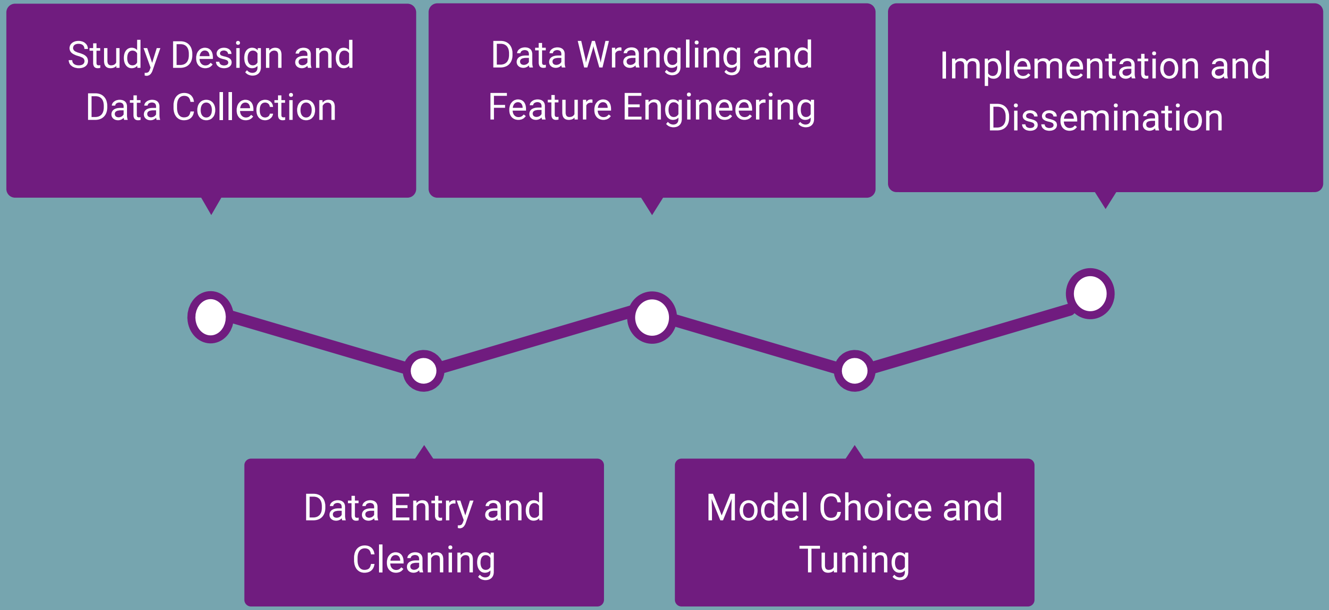
Ziad Obermeyer[1,2]*, Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5]*†

Health systems rely on commercial prediction algorithms to identify and help patients with complex health needs. We show that a widely used algorithm, typical of this industry-wide approach and affecting millions of patients, exhibits significant racial bias: At a given risk score, Black patients are considerably sicker than White patients, as evidenced by signs of uncontrolled illnesses. Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%. The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients. Thus, despite health care cost appearing to be an effective proxy for health by some measures of predictive accuracy, large racial biases arise. We suggest that the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts.

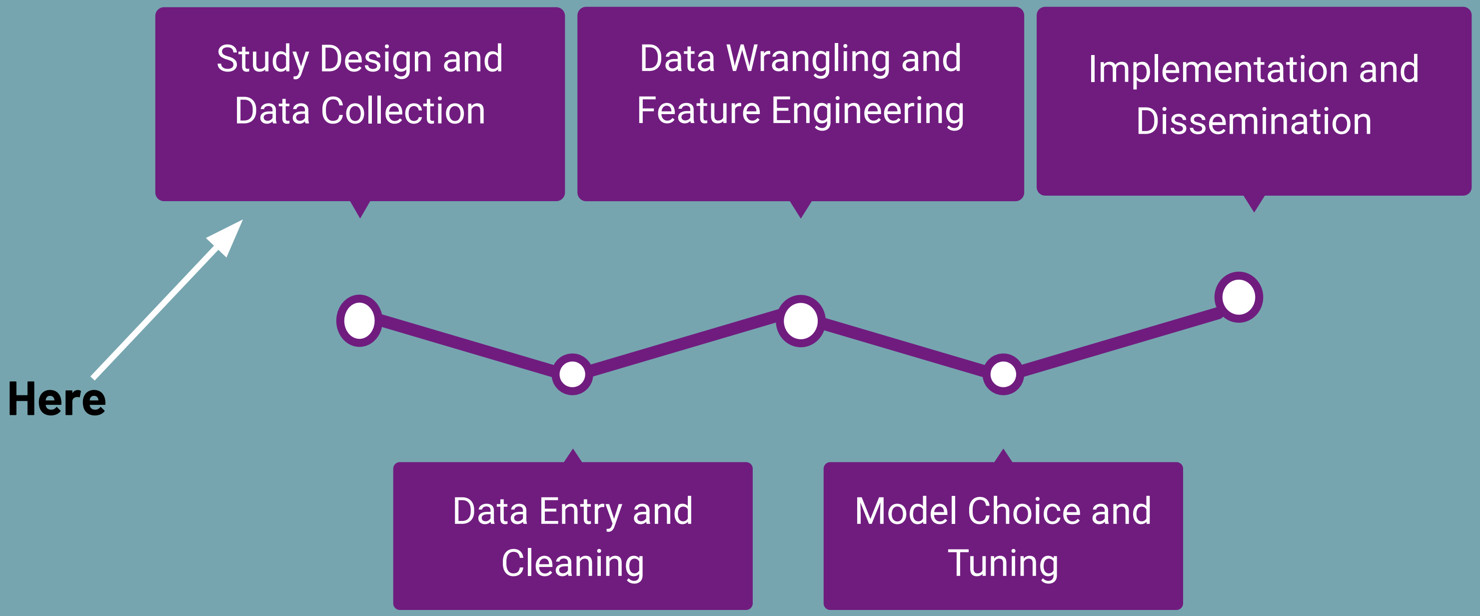
Source: https://science.sciencemag.org/content/366/6464/447/tab-pdf



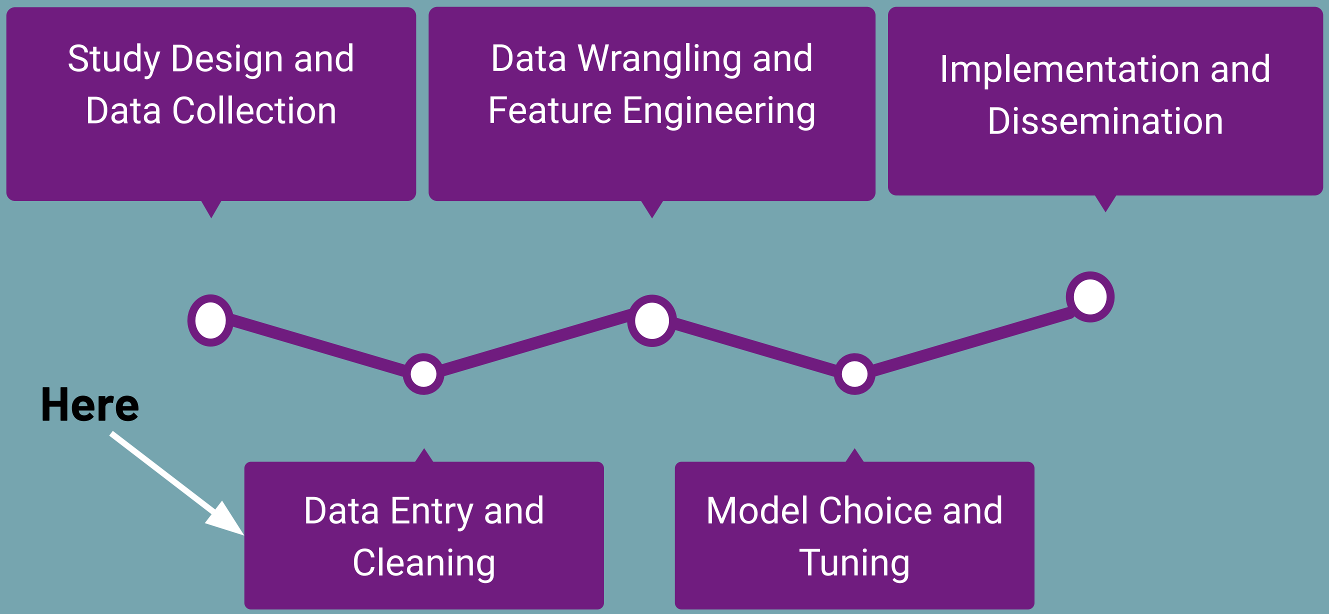
Where can bias be introduced in the process?



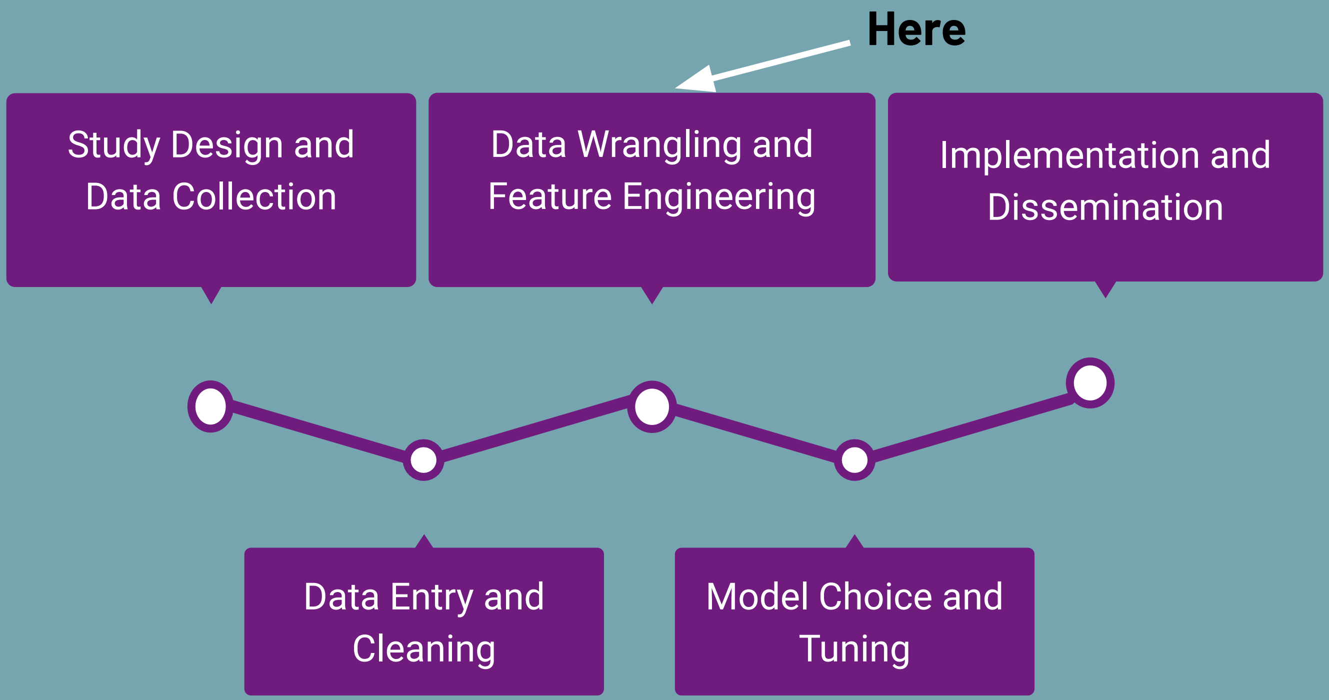
Where can bias be introduced in the process?



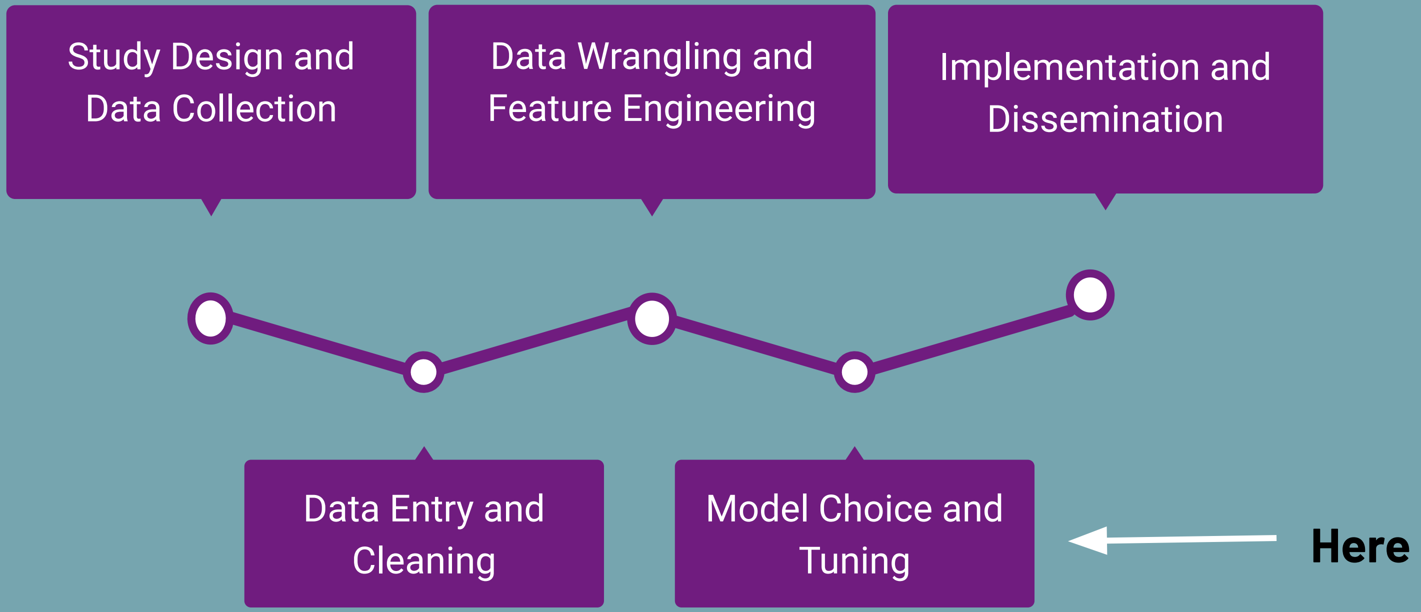
Where can bias be introduced in the process?



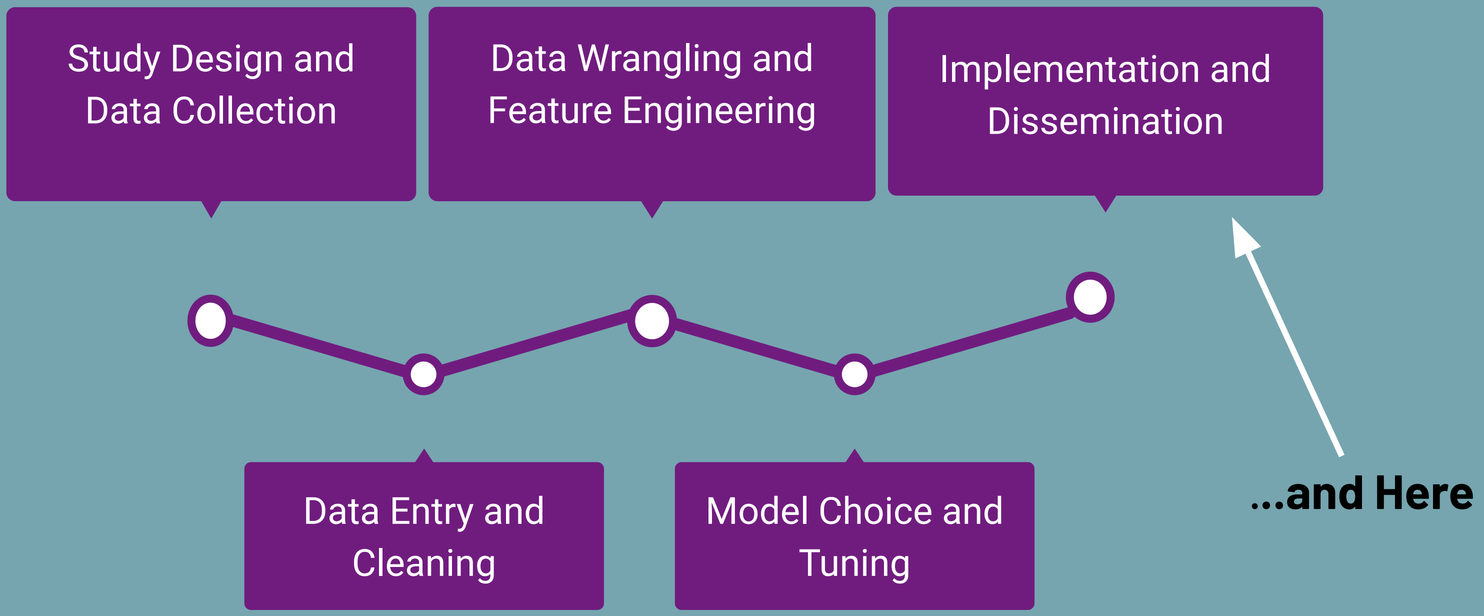
Where can bias be introduced in the process?



Where can bias be introduced in the process?



Where can bias be introduced in the process?



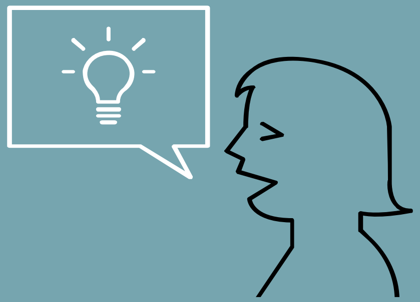
How do we fix this?

“… human beings cannot overcome all forms of bias. But slowing down and learning what those traps are—as well as how to recognize and challenge them—is critical.”

- Yaël Eisenstat

Former CIA officer, national security advisor to vice president Biden, integrity operations head at Facebook

Source: https://www.wired.com/story/the-real-reason-tech-struggles-with-algorithmic-bias/



Data Management

Data Management

- Samples should be **representative** of the population algorithm will be applied to

Data Management

- Samples should be **representative** of the population algorithm will be applied to
- Sufficient sampling of **underrepresented groups** - when possible
  - By definition, there is less information known about underrepresented groups. Oversampling these groups during data collection will help reduce further marginalization of these groups during model fitting and implementation

Data Management

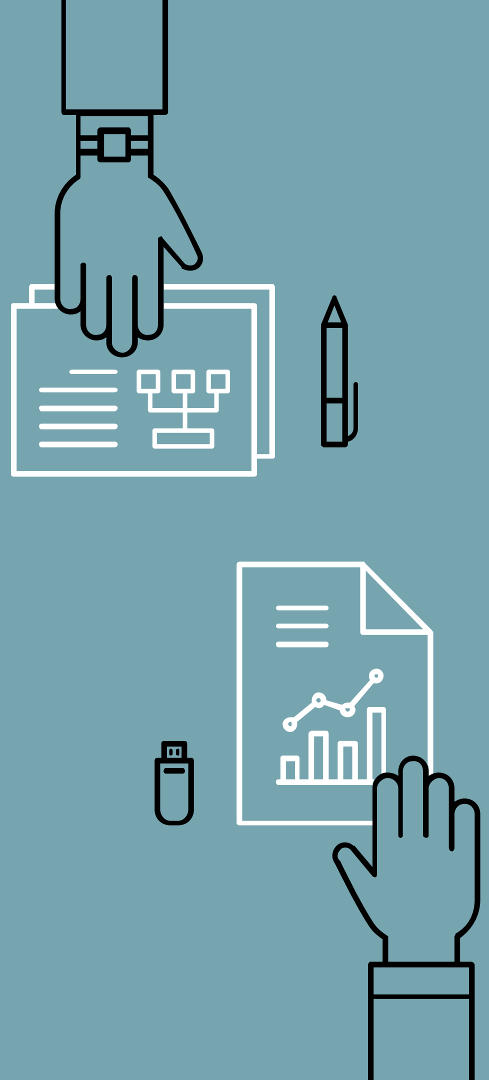
- Samples should be **representative** of the population algorithm will be applied to
- Sufficient sampling of **underrepresented groups** - when possible
  - By definition, there is less information known about underrepresented groups. Oversampling these groups during data collection will help reduce further marginalization of these groups during model fitting and implementation
- **Thorough documentation** of any cleaning or transformations done
  - Also known as data provenance

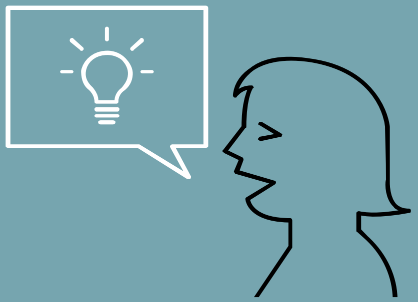
Data Management

- Samples should be **representative** of the population algorithm will be applied to
- Sufficient sampling of **underrepresented groups** – when possible
  - By definition, there is less information known about underrepresented groups. Oversampling these groups during data collection will help reduce further marginalization of these groups during model fitting and implementation
- **Thorough documentation** of any cleaning or transformations done
  - Also known as data provenance
- Ideally, teams **inspect** quality of data labeling, handling, etc. and don't leave it to one person

Data Management

- Samples should be **representative** of the population algorithm will be applied to
- Sufficient sampling of **underrepresented groups** - when possible
  - By definition, there is less information known about underrepresented groups. Oversampling these groups during data collection will help reduce further marginalization of these groups during model fitting and implementation
- **Thorough documentation** of any cleaning or transformations done
  - Also known as data provenance
- Ideally, teams **inspect** quality of data labeling, handling, etc. and don't leave it to one person
- **Feature engineering** should be discussed among a team in an effort to limit implicit bias





Choice of Algorithm and Performance Metrics

Algorithm Choice

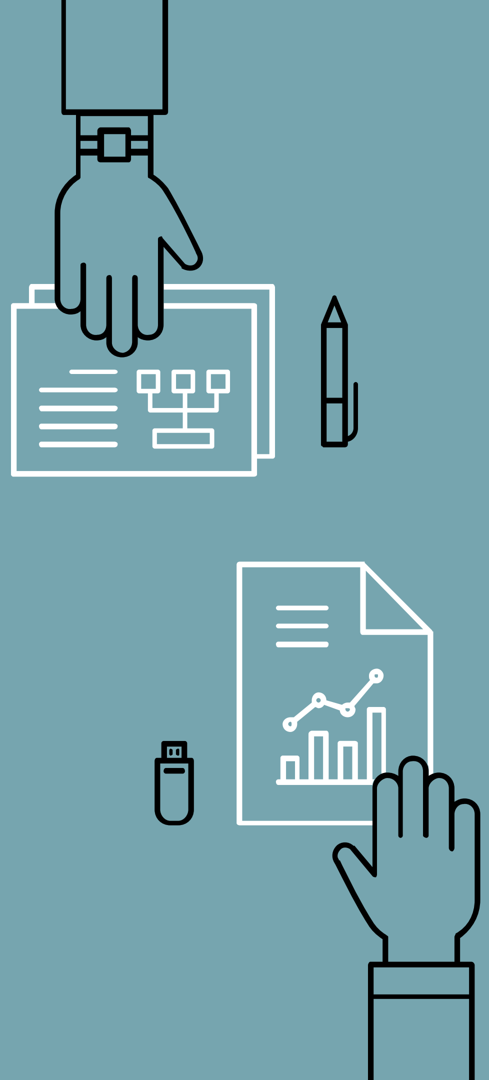
- Use appropriate methods
  - Type of outcome (continuous, categorical, binary, survival, etc.) and type of predictors determine which methods are appropriate for that type of data
  - Method assumptions need to be met. Examples:
    - Normality of the data
    - Independent observations
    - No missingness
    - All subgroups equally represented
  - If assumptions not met, correct for them

Algorithm Choice

- Use appropriate methods
  - Type of outcome (continuous, categorical, binary, survival, etc.) and type of predictors determine which methods are appropriate for that type of data
  - Method assumptions need to be met. Examples:
    - Normality of the data
    - Independent observations
    - No missingness
    - All subgroups equally represented
  - If assumptions not met, correct for them
- Model complexity
  - Too complex = not generalizable to other populations
  - Too simple = decrease in overall performance

Algorithm Choice

- Algorithm performance
  - Setting an appropriate baseline/benchmark



Algorithm Choice

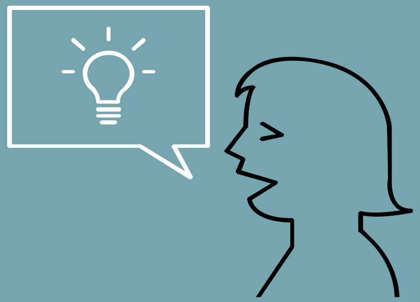
- Algorithm performance
  - Setting an appropriate baseline/benchmark performance to beat
  - Use more than 1 performance metric to measure model performance
    - Don’t just choose/report the metric that makes your model look good

Algorithm Choice

- Algorithm performance
  - Setting an appropriate baseline/benchmark performance to beat
  - Use more than 1 performance metric to measure model performance
    - Don’t just choose/report the metric that makes your model look good
  - Think about the **scale** of an incorrect prediction is
    - Your model may classify 95% of people correctly, but what about those 5%? How detrimental is the misclassification?

Algorithm Choice

- Algorithm performance
  - Setting an appropriate baseline/benchmark performance to beat
  - Use more than 1 performance metric to measure model performance
    - Don’t just choose/report the metric that makes your model look good
  - Think about the **scale** of an incorrect prediction is
    - Your model may classify 95% of people correctly, but what about those 5%? How detrimental is the misclassification?
  - A loss in overall performance is usually worth a gain in equity among groups

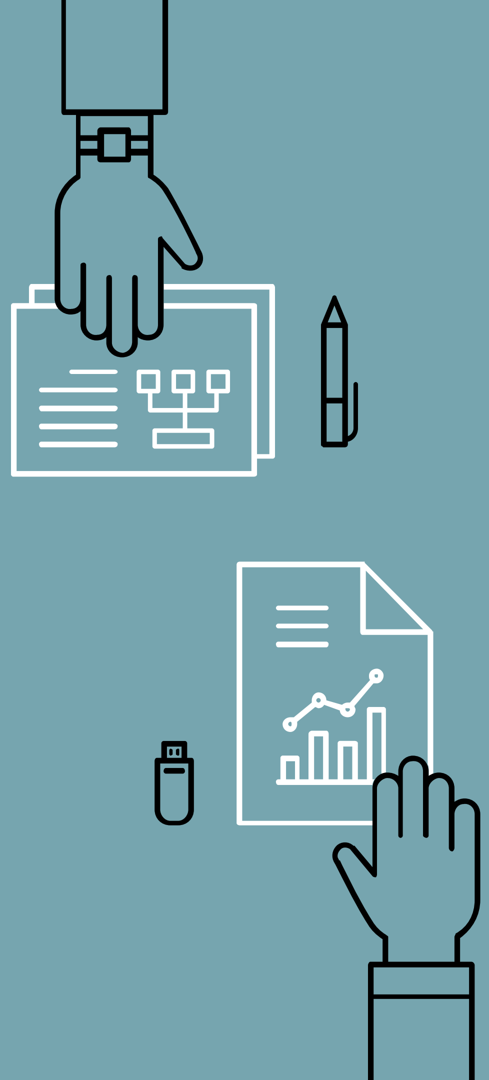


Transparency

Transparency

Reproducibility

- Open software
- Thorough documentation
  - Literate programming
  - Readme file with details and instructions
  - Intuitive file, folder and variable names
  - Data dictionary or codebook
- Open data
- Toy examples



Transparency

Reproducibility

- Open software
- Thorough documentation
  - Literate programming
  - Readme file with details and instructions
  - Intuitive file, folder and variable names
  - Data dictionary or codebook
- Open data
- Toy examples

Accountability

- Be upfront with the limitations of the model/framework/data etc.
- Report several performance measures
  - Not just the one or ones that make your model/method look good
- Report the generalizability of your model and if it should be applied to different populations

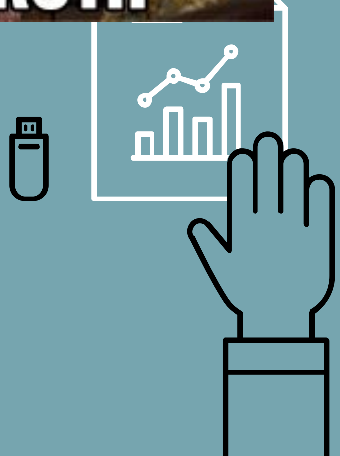
Transparency

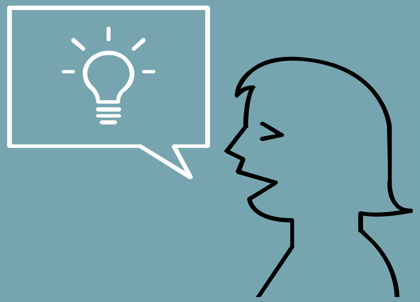
Reproducibility

- Open software
- Thorough documentation
  - Literate programming
  - Readme file with details and instructions
  - Intuitive file, folder and variable names
  - Data dictionary or codebook
- Open data
- Toy examples

Accountability

- Be upfront with the limitations of the model/framework/data etc.
- Report several performance measures
  - Not just the one or ones that make your model/method look good
- Report the generalizability of your model and if it should be applied to different populations

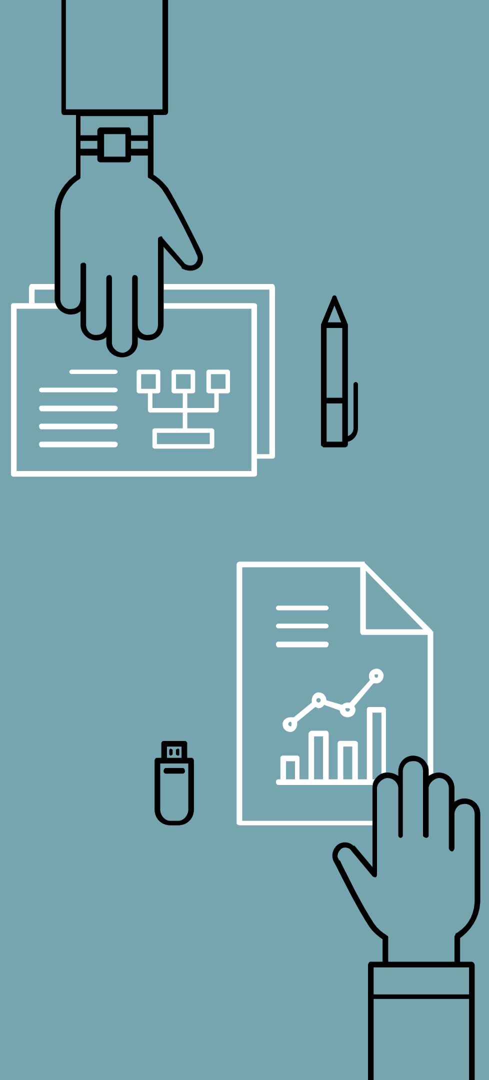




Diverse Data Science Teams

Diverse Teams

- Having diverse teams – in terms of **gender**, **ethnicity**, **training, background, etc.** – will increase the likelihood of unconscious biases being recognized and addressed
- Everyone views the world through a different lens



Conclusions

▷ Humans and data are biased

Conclusions

- Humans and data are biased
- Algorithmic bias is becoming more of a concern as AI is used more and more in healthcare

Conclusions

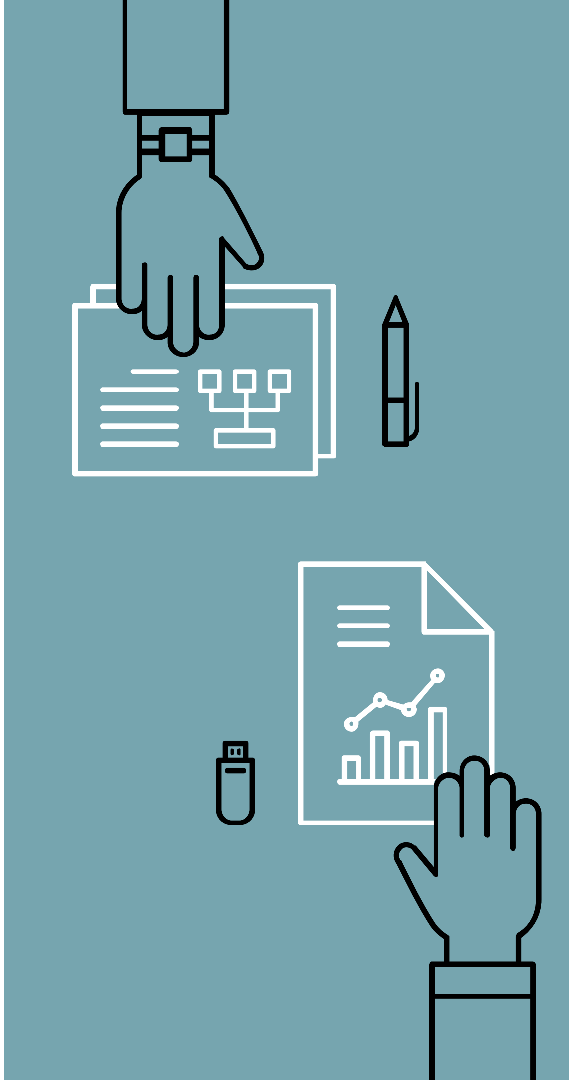
- Humans and data are biased
- Algorithmic bias is becoming more of a concern as AI is used more and more in healthcare
- An algorithm is only as good as the data training it and the people building it and implementing it

Conclusions

- Humans and data are biased
- Algorithmic bias is becoming more of a concern as AI is used more and more in healthcare
- An algorithm is only as good as the data training it and the people building it and implementing it
- While completely eliminating algorithmic bias is impossible (for now), there are many ways to mitigate and reduce bias throughout the entire analysis and dissemination process

Conclusions

- Humans and data are biased
- Algorithmic bias is becoming more of a concern as AI is used more and more in healthcare
- An algorithm is only as good as the data training it and the people building it and implementing it
- While completely eliminating algorithmic bias is impossible (for now), there are many ways to mitigate and reduce bias throughout the entire analysis and dissemination process
- Go be ethical health data scientists!



Resources

- Artificial intelligence and algorithmic bias: implications for health systems
- Algorithmic Fairness
  - A great review paper
- Tomorrow! Algorithmic Fairness Session