

Introduction to Data Science

Data Ethics - Algorithmic Bias

Joanna Bieri DATA101

Algorithmic Bias

ASK

What do we mean
by algorithmic bias?

Algorithmic Bias

ANSWER

- An algorithm makes decisions that are unfair to a group.
- Usually nobody programmed it to be unfair. It learned the bias from the data it was trained on.
- The data comes from our society, which has a history of bias.

The Hathaway Effect

EXAMPLE



- Berkshire Hathaway (Warren Buffett's company) stock went UP on days with big Anne Hathaway news.
- Bride Wars opens 1/5/2009: up 2.61%. Anne named Oscars co-host 11/29/2010: up 0.25%.
- The idea (Dan Mirvish, HuffPost 2011): people search "Hathaway", trading bots see the buzz, and they can't tell Anne from Berkshire.
- This is observational. It might not be a real effect!

The Hathaway Effect

DISCUSS

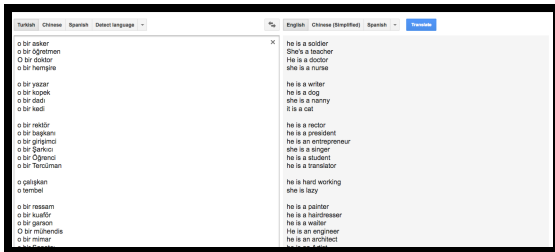
Do you think this is a real effect?

This one is funny.

How could the same mistake go really wrong?

Google Translate

EXAMPLE



- Turkish has one pronoun, “o”, for he AND she. To translate into English the algorithm has to pick one.
- It picked based on the text it learned from, so jobs and traits came out with the usual gender stereotypes.

Google Translate

DISCUSS

Where did it learn to pick he or she?

Is the algorithm biased, or the data?

Amazon's Hiring Algorithm

EXAMPLE

- Amazon built an AI tool to score job candidates from one to five stars, like shoppers rate products. (Reuters, 2018)
- It was trained on resumes from people Amazon hired in the past.
- It taught itself that male candidates were preferable for software and technical jobs.
- It also recommended unqualified candidates for all kinds of jobs. Amazon scrapped it.

Hiring Algorithms

ASK

Amazon's tool was never told
to prefer men.

How did it learn to?

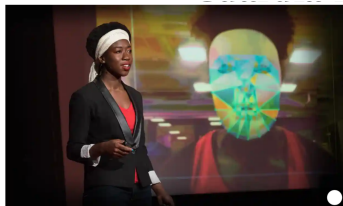
Hiring Algorithms

ANSWER

- It was trained on past hiring decisions that were biased.
- It picks up on any feature in the data that goes along with who got hired before.
- It was trained to minimize its cost function. Copying the old hiring decisions does that, so it copies the bias too.

Facial Recognition

EXAMPLE



Interview
'A white mask worked better': why algorithms are not colour blind

Ian Tucker
When Joy Buolamwini found that a robot recognised her face better when she wore a white mask, she knew a problem needed fixing

Sun 28 May 2017 13:27 BST

Joy Buolamwini is a graduate researcher at the MIT Media Lab and founder of the Algorithmic Justice League - an organisation that aims to challenge the biases in decision-making software. She grew up in Mississippi, gained a Rhodes scholarship, and she is also a Fulbright fellow, an Astronaut scholar and a Google Anita Borg scholar. Earlier this year she won a \$50,000 scholarship funded by the makers of the film *Hidden Figures* for her work fighting coded discrimination.

- Joy Buolamwini, a graduate researcher at the MIT Media Lab, founded the Algorithmic Justice League.
- Facial recognition was trained on mostly white faces, so it did a bad job on non-white faces.
- When she put on a plain white mask, the algorithm tracked the mask better than her own face! (The Guardian, 2017)

Facial Recognition

DISCUSS

Who was missing from the training data?

Where is facial recognition used today?

Who gets hurt when it is wrong?

Criminal Sentencing

EXAMPLE

- Courts use risk scores that predict how likely someone is to commit another crime. The scores affect bail and sentencing.
- ProPublica ("Machine Bias", 2016) checked the scores for more than 7,000 people arrested in Broward County, Florida.
- Only 20% of the people predicted to commit violent crimes actually did. Overall accuracy was 61%.
- Black defendants were falsely flagged as future criminals at almost twice the rate of white defendants. White defendants were mislabeled as low risk more often.

Criminal Sentencing

PREDICT

group	did NOT reoffend	flagged high risk anyway
A	100	20
B	100	40

A made-up risk score. Among people who did not reoffend, how often was each group falsely flagged as high risk?

Criminal Sentencing

ANSWER

group	did NOT reoffend	flagged high risk anyway
A	100	20
B	100	40

- Group A: 20 out of 100, so 20%. Group B: 40 out of 100, so 40%.
- Group B is falsely flagged twice as often. This is the kind of gap ProPublica found.
- Each false flag is a real person, and decisions about their liberty and freedom are being made with these scores!

Data Ethics and Algorithms

ASK

Many people believe that taking humans out of the decision makes it MORE fair.

What is wrong with that idea?

Data Ethics and Algorithms

ANSWER

- It is trained on data from our society, so it learns our bias.
- Who is in the data, and who collected it, shapes what it learns.
- An algorithm can repeat a biased decision much faster and for many more people. Since a computer made the decision, people often assume it is fair.

Data Ethics in your Work

DISCUSS

- At some point you will learn tools that can be used unethically.
- You might be tempted to use them anyway: business goals, the pursuit of knowledge, or because your boss told you to.
- Want to do good with data? Data Science for Social Good, DataKind.

In Class Question 1

DISCUSS

What is algorithmic bias?

Give a few examples!

In Class Question 2

DISCUSS

How does algorithmic bias happen?

What are some ways the bias enters
our mathematical algorithms?

In Class Question 3

DISCUSS

What are some ways that you can train yourself to make good decisions with data science and modeling?

Be specific here!

In Class Question 4

DISCUSS

How might you approach someone else
to point out bias or ethical issues in their work?

Questions?

Questions?