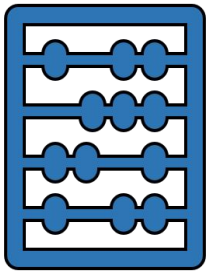
Bias in Machine Learning Algorithms

don’t let it happen to you

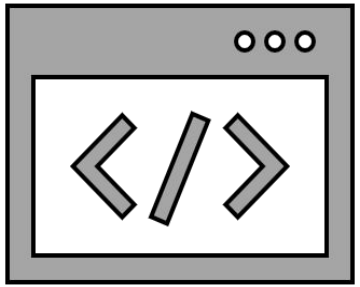
Overview



ML intro

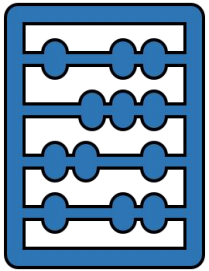
Sources of bias

Removing bias



Avoiding bias

What is Machine Learning (ML)?



ML refers to computer algorithms that improve **automatically** through experience.

Supervised / Unsupervised

Classification

Predictive

examples:

newsfeeds; recommendations; self-driving vehicles; facial recognition, predictive financial analytics, route planning,

What bias might we encounter?

Established:

historical bias that exists within the data

Technical:

software and/or hardware limitations

Emergent:

generate and perpetuate bias

How do we remove the bias?

Step 1: Detect

Inspect the **results**

Inspect the **data**

Consider proxies

“For whom does this data fail?” - Cathy O’Neil (Algorithm Auditor)

Step 2: Adjust

Change your **success criteria**

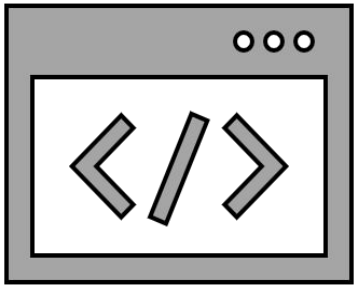
Change your **seed data**

Don’t forget to consider proxies!

Include transparency

Test: See Step 1

How can we avoid bias?



Be aware of the potential sources of bias.

Consider carefully our success criteria

Include transparency in results

Find for whom the algorithm fails

Test. Adjust. Repeat

Machine Learning algorithms will not remove human bias, only reflect and perpetuate it

Thank you