

Topics in Data Ethics : Bias

•
DongJun Min

0. Bias

A Framework for Understanding Unintended Consequences of Machine Learning By Harini Suresh et al.

: Social science concept of bias

1. Historical Bias

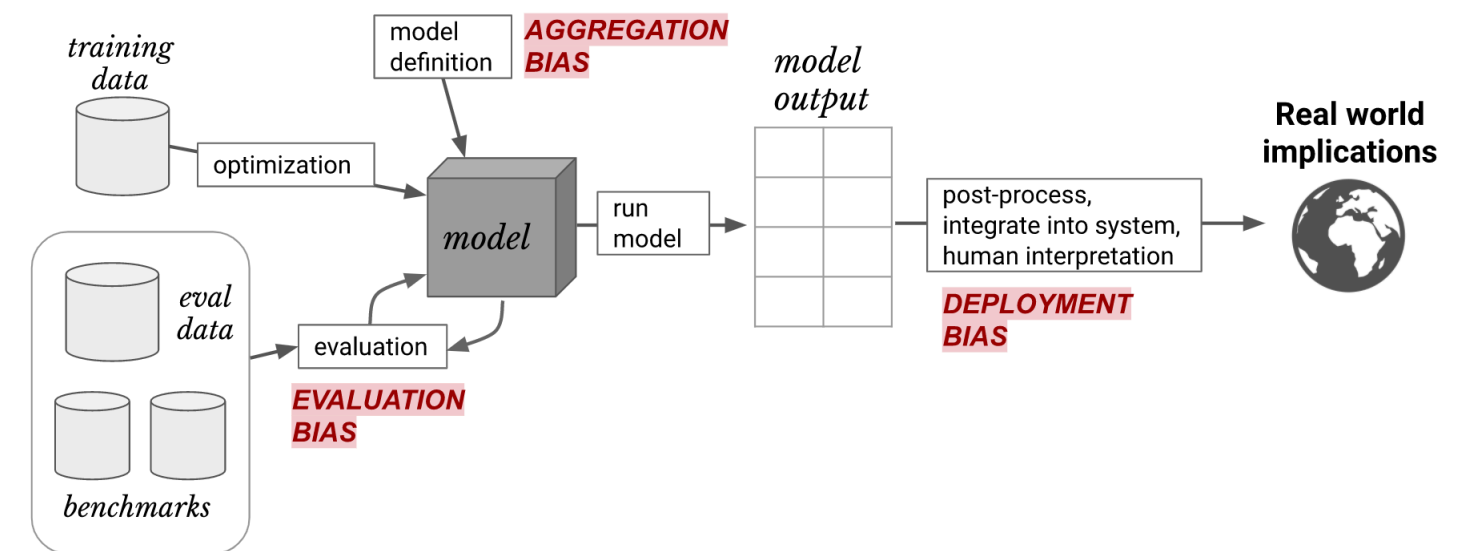
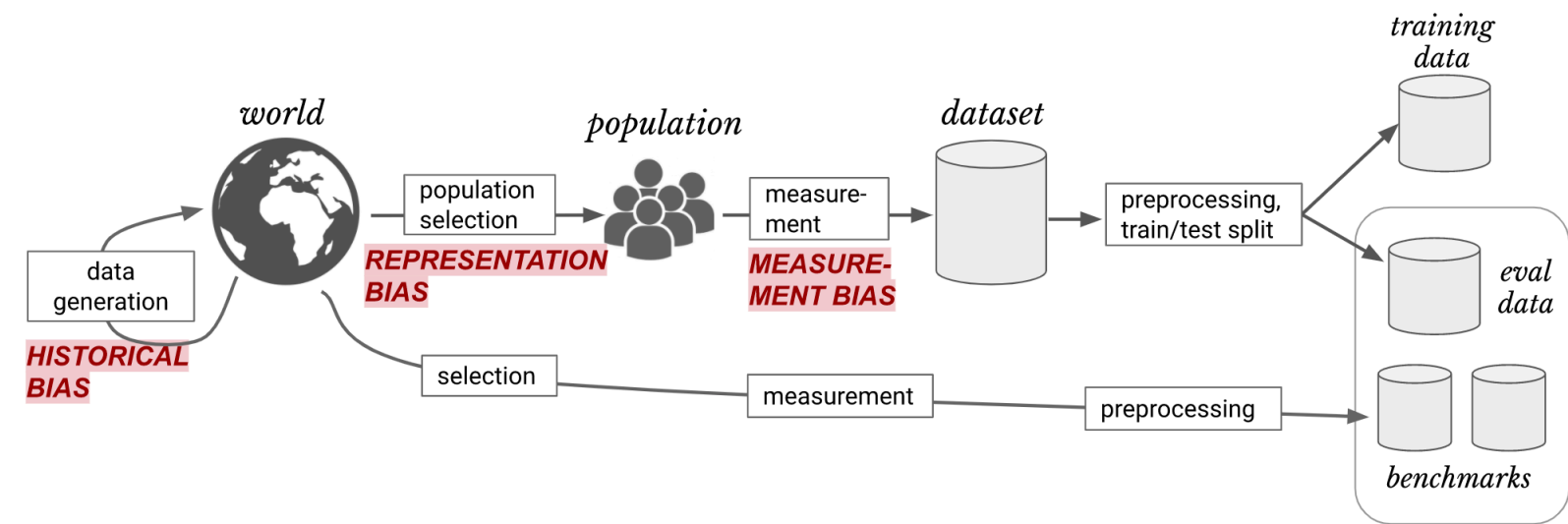
2. Representation Bias

3. Measurement Bias

4. Aggregation Bias

5. Evaluation Bias

6. Deployment Bias



1. Historical bias

people are biased, processes are biased, and society is biased

Racial Bias, COMPAS software

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%

Two Petty Theft Arrests

VERNON PRATER

Prior Offenses
2 armed robberies, 1 attempted armed robbery

Subsequent Offenses
1 grand theft

LOW RISK

3

BRISHA BORDEN

Prior Offenses
4 juvenile misdemeanors

Subsequent Offenses
None

HIGH RISK

8

Two Drug Possession Arrests

DYLAN FUGETT

Prior Offense
1 attempted burglary

Subsequent Offenses
3 drug possessions

LOW RISK

3

BERNARD PARKER

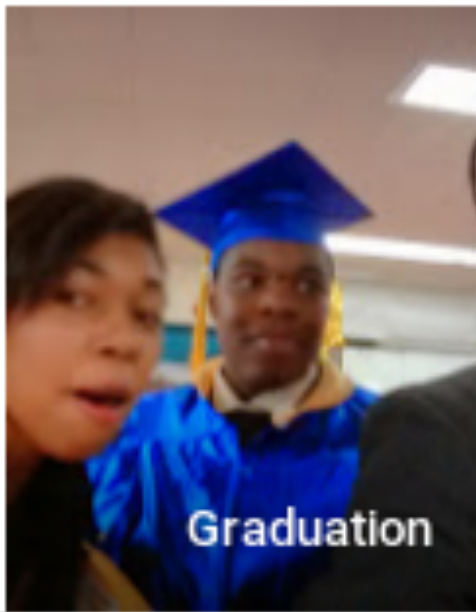
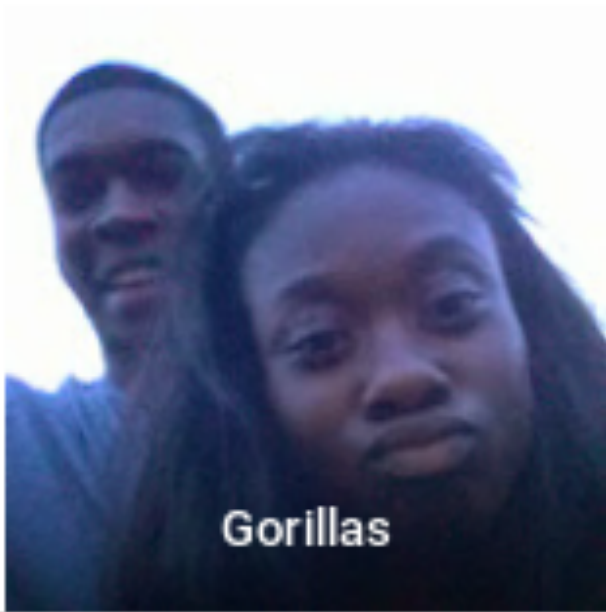
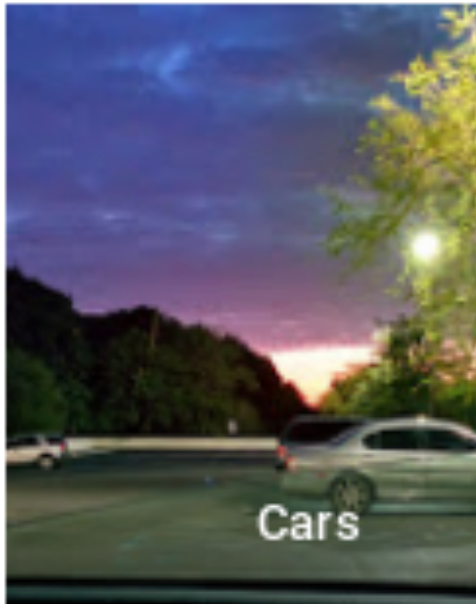
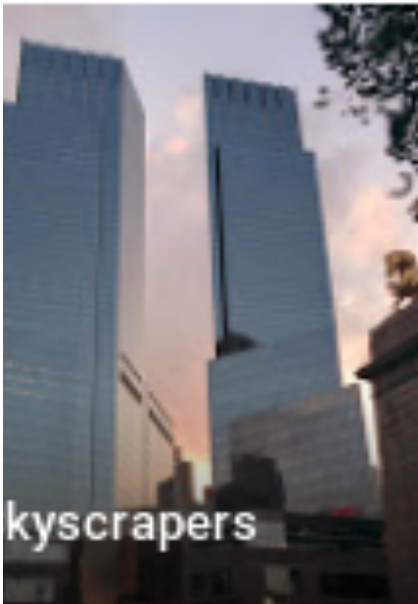
Prior Offense
1 resisting arrest without violence




Subsequent Offenses
None

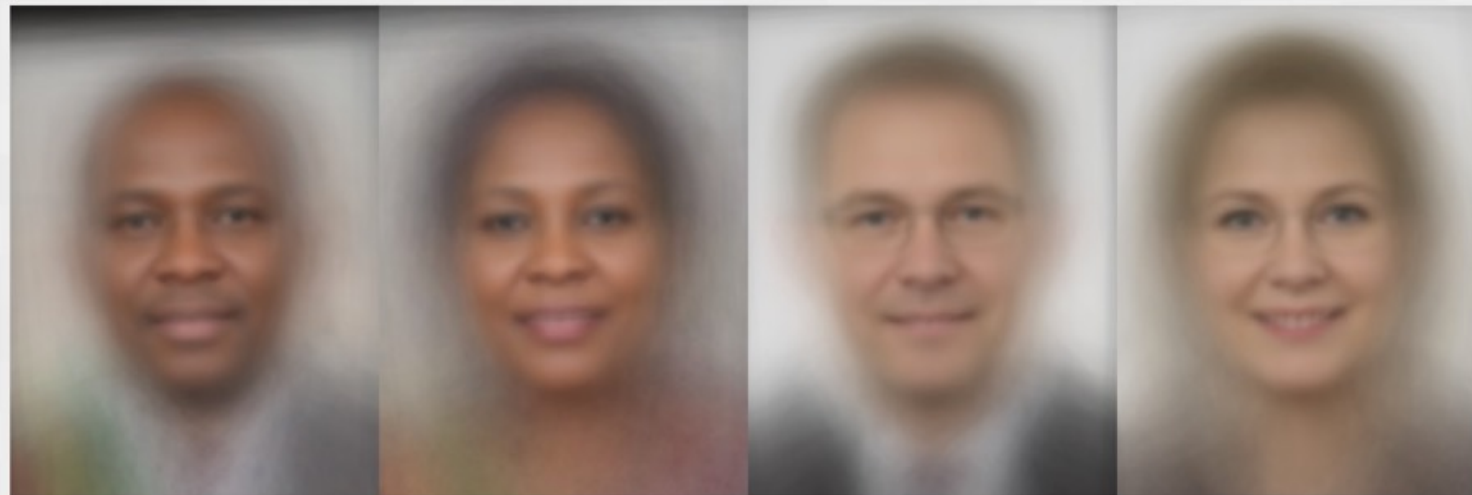
HIGH RISK

10

1. Historical bias

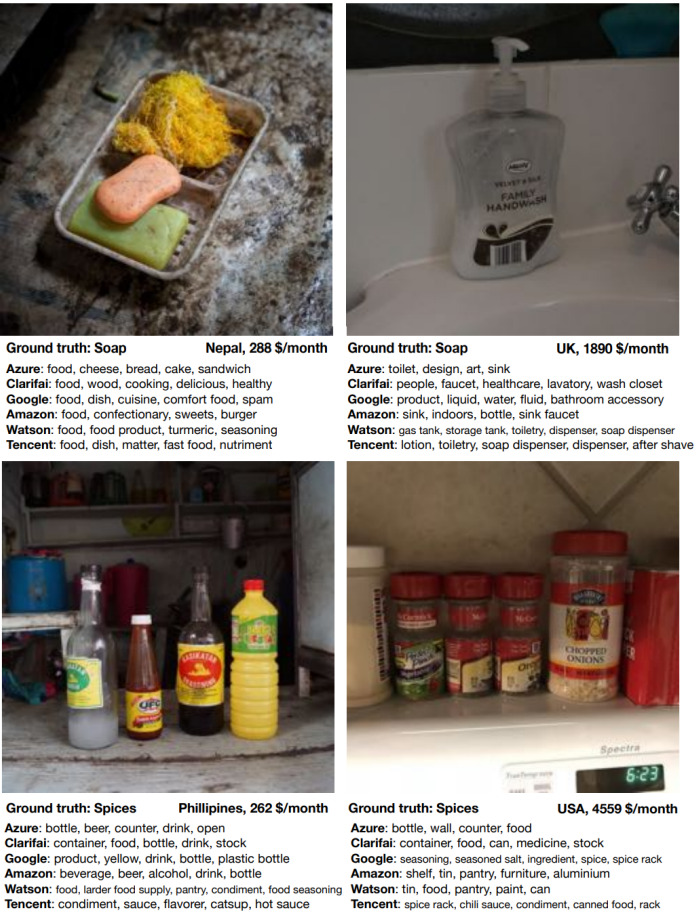
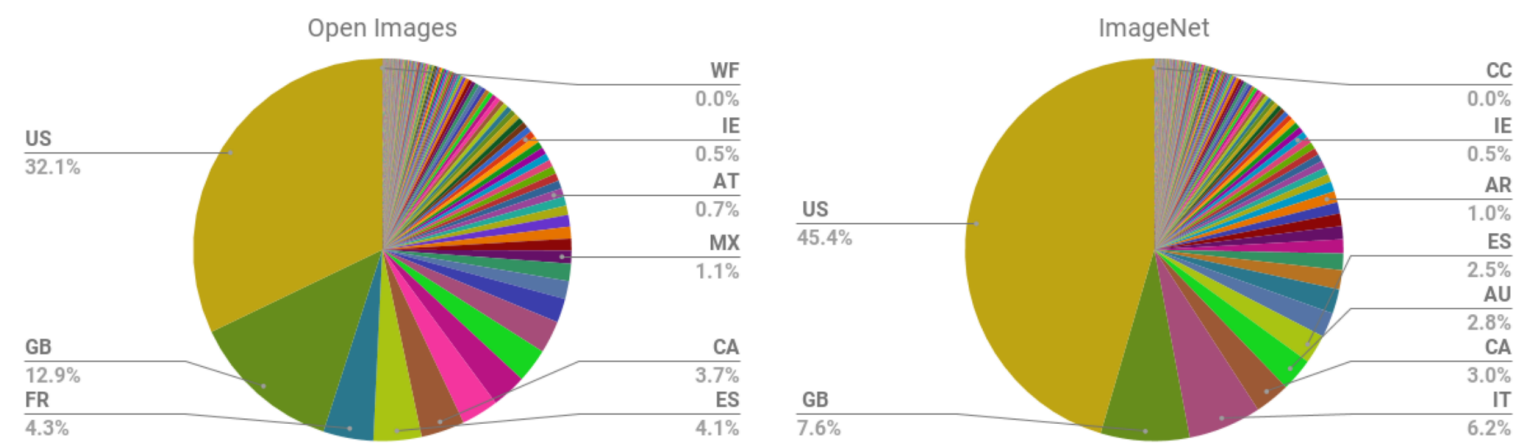


Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
 Microsoft	94.0% <div><div></div></div>	79.2% <div><div></div></div>	100% <div><div></div></div>	98.3% <div><div></div></div>	20.8% <div><div></div></div>
 FACE++	99.3% <div><div></div></div>	65.5% <div><div></div></div>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



1. Historical bias

- 1. No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World By Shreya Shankar et al.
- 2. Does Object Recognition Work for Everyone? By Terrance DeVries et al.



1. Historical bias

English Turkish Spanish Detect language ▾

↔

English Turkish Spanish ▾

Translate

She is a doctor.
He is a nurse.

✕

31/5000

O bir doktor.
O bir hemşire.

☆ 📄 🔊 ➦

English Turkish Spanish Turkish - detected ▾

↔

English Turkish Spanish ▾

Translate

O bir doktor.
O bir hemşire

✕

28/5000

He is a doctor.
She is a nurse ✓

☆ 📄 🔊 ➦

2. Measurement bias

What factors are most predictive of stroke?

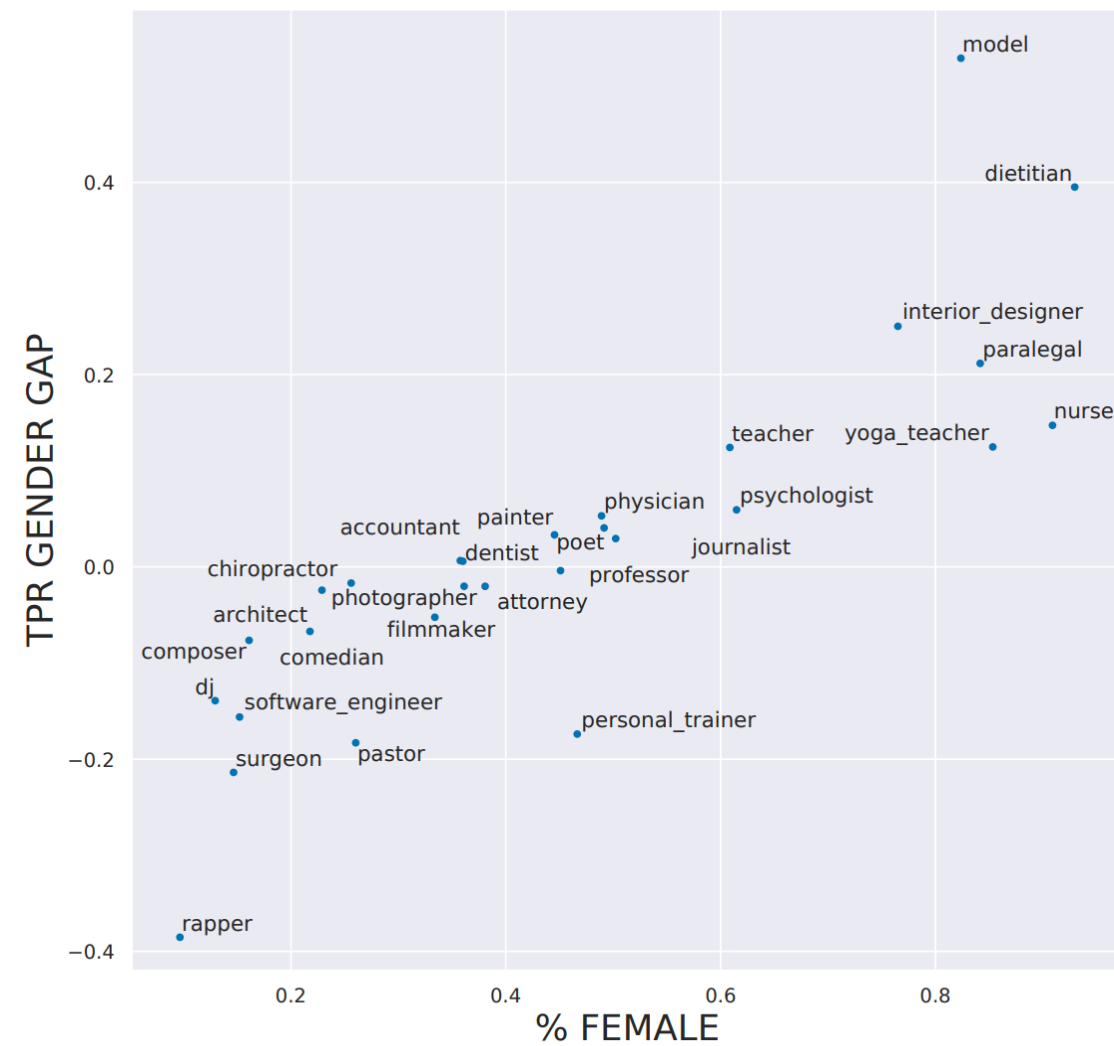
- >> Prior stroke
- >> Cardiovascular disease
- >> Accidental injury
- >> Benign breast lump
- >> Colonoscopy
- >> Sinusitis

3. **Aggregation bias**

Can occurs model that does not include these important variables and interactions.

4. Representation bias

Bias in Bios: A Case Study of Semantic Representation Bias in a High-Stakes Setting By Maria De-Arteaga et al.



5. Addressing different types of bias

Does Machine Learning Automate Moral Hazard and Error By Sendhil Mullainathan and Ziad Obermeyer

Consider these points about machine learning algorithms:

- >> Machine learning can create feedback loops
- >> Machine learning can amplify bias
- >> Algorithms & humans are used differently
- >> Technology is power

5. Addressing different types of bias

- >> People are more likely to assume algorithms are objective or error-free (even if they're given the option of a human override).
- >> Algorithms are more likely to be implemented with no appeals process in place.
- >> Algorithms are often used at scale.
- >> Algorithmic systems are cheap.