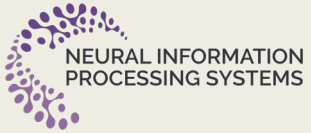
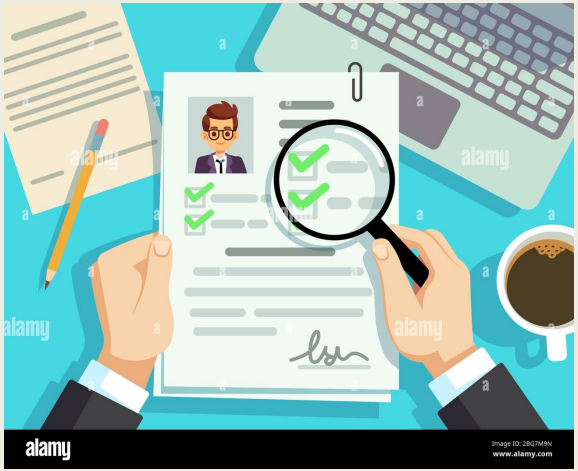
On Learning Fairness and Accuracy on Multiple Subgroups

Changjian Shui, Gezheng Xu, Qi Chen, Jiaqi Li,
Charles X. Ling, Tal Arbel, Boyu Wang, Christian Gagné



AI in sociotechnical system



Candidate evaluations for job positions

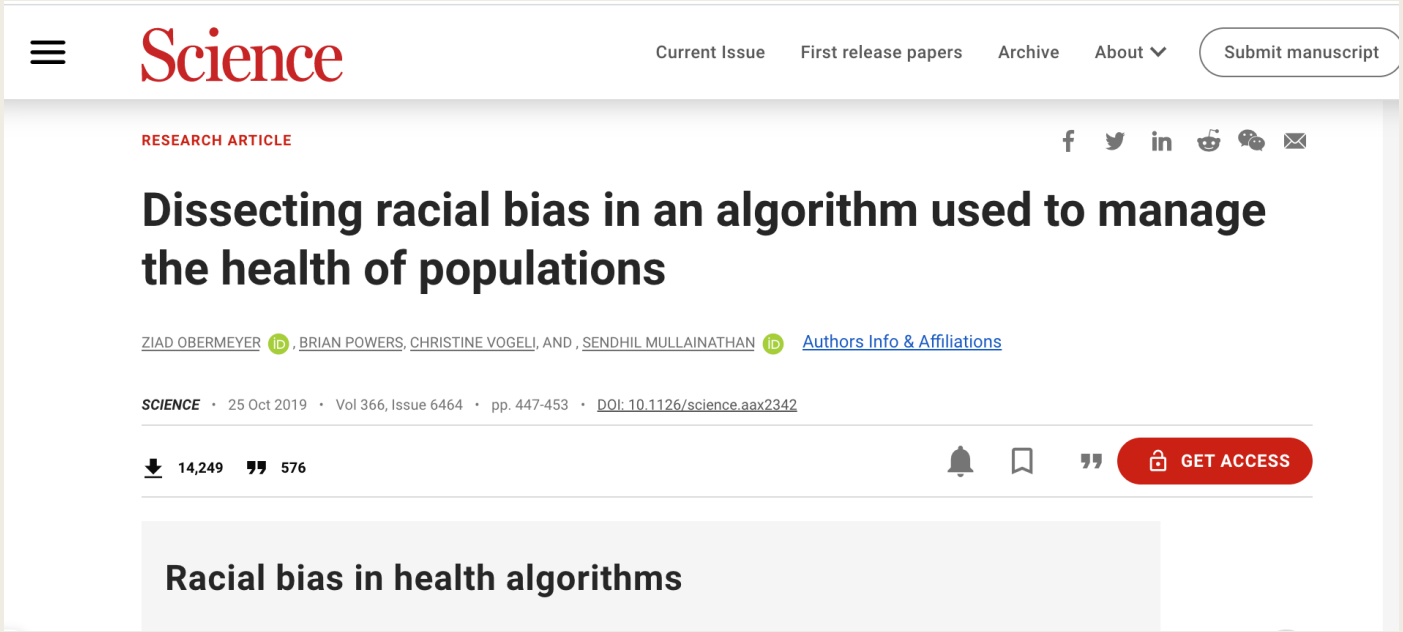
Health risk assessment

Driven by AI algorithms

Source: https://www.vecteezy.com/vector-art/2207767-health-risk-assessment-blue-gradient-concept-icon

Algorithmic Discrimination

Medical AI



Obermeyer et al., 366 Science 447 (2019)

Group Fairness

No prediction disparities in different demographics.

- *Age, gender, race, hospital……*
- *No unified definitions.*

Trivial Fair Decision

Coin flipping can trivially achieve fair prediction.

- For any job application, the offer is random.

The prediction should be informative!

Source: https://en.wikipedia.org/wiki/Coin_flipping

Desiderata in fair learning

- Informative.
  - *Learning the utility of the data*
- Fair
  - *No prediction disparities*

Possibility to simultaneously achieve these two?

Depending on fairness notion.

Group sufficiency

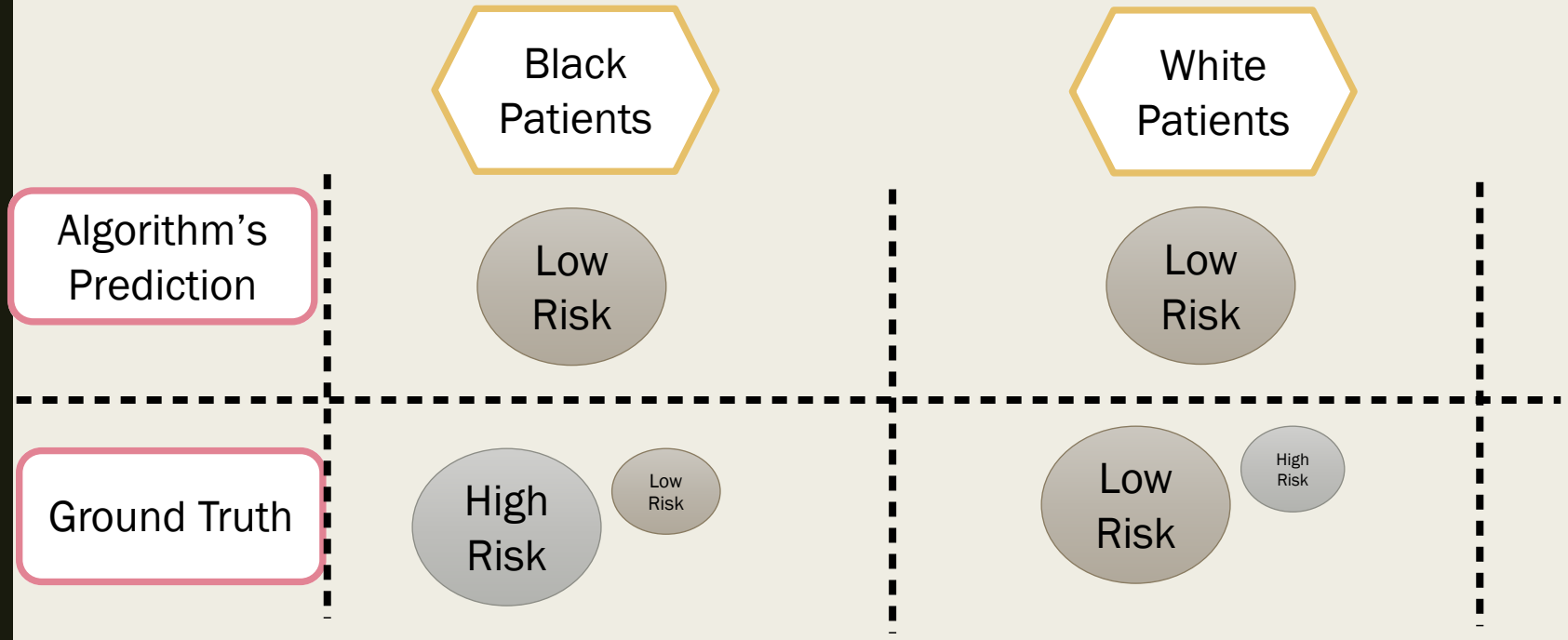
Example in Health AI

- AI algorithms predict the health-care score for each patient.
- Higher score -> Sicker

(need to transfer to ICU)

Obermeyer et al., 366 *Science* 447 (2019)

Calibration Bias (Example in Health)



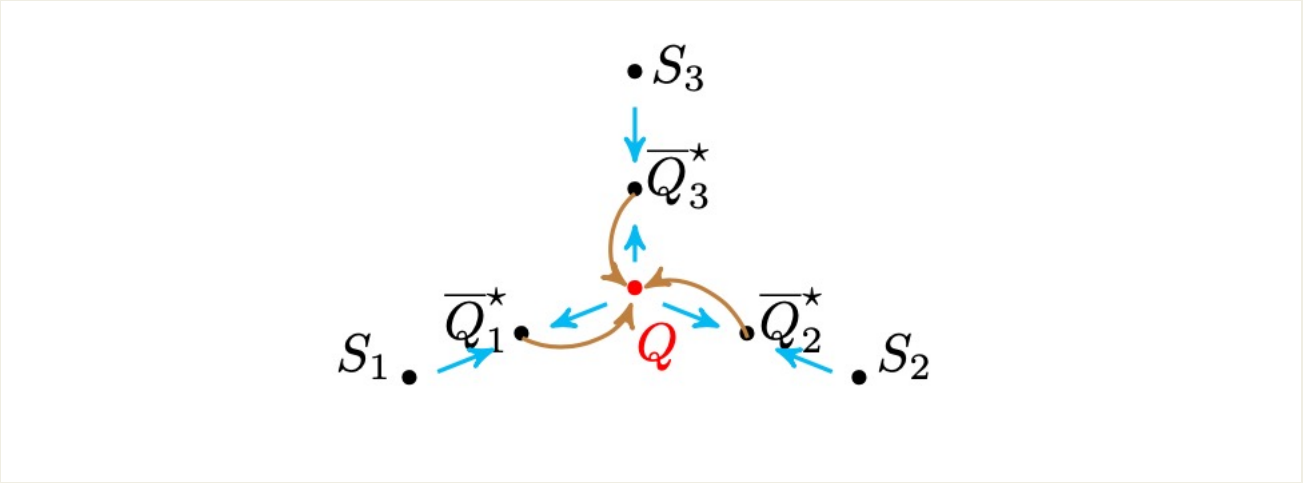
Severity of Black patients is under-estimated.

Obermeyer et al., 366 *Science* 447 (2019)

Formal definition

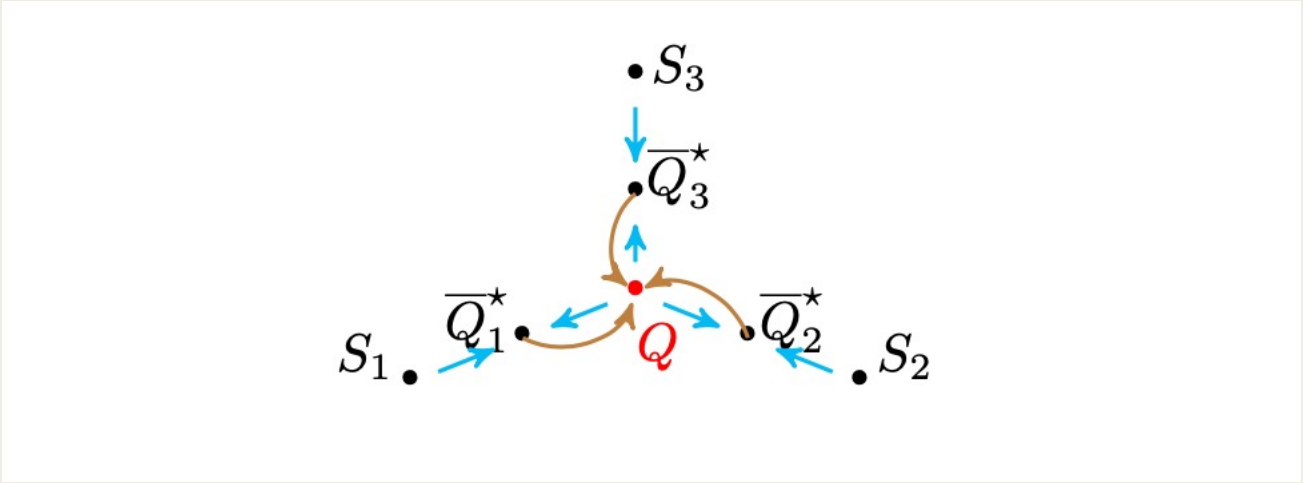
- Group sufficiency: $E[Y|f(X)] = E[Y|f(X), A]$
- Mitigate bias across multiple (or many) subgroups
- Limited data within each subgroup
- Learning data utility with comparable accuracy

Proposed algorithm (informal)



1. Q: fair and informative predictor.
2. S_1, S_2, S_3: different subgroup (e.g., data from different ages)
3. Q_1, Q_2, Q_3 : subgroup specific predictors

Step One

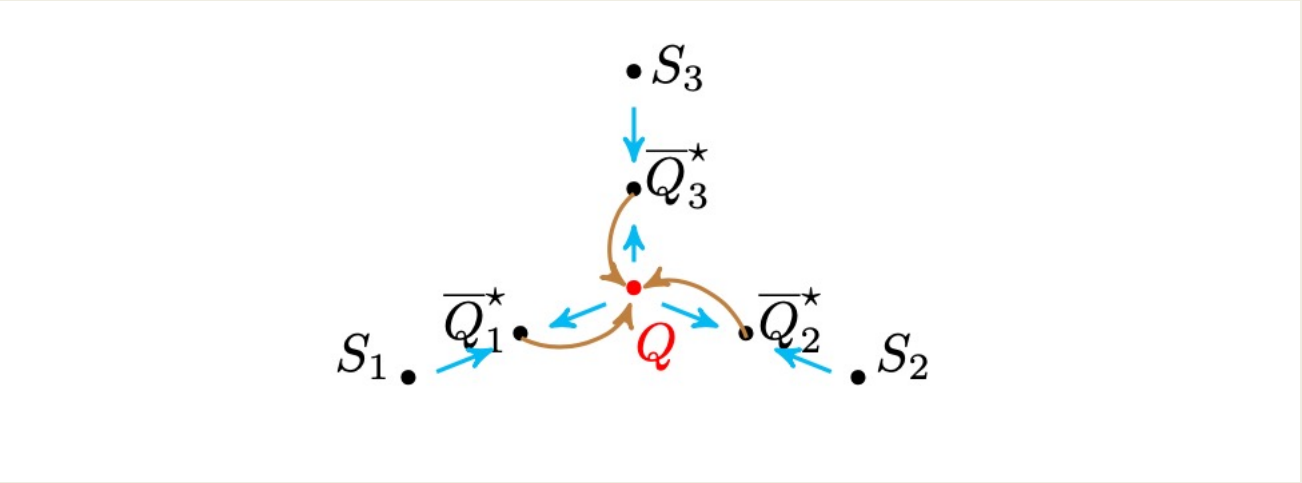


Fix Q (the fair and informative predictor.)

-> Learn Q_1, Q_2, Q_3 from S_1, S_2, S_3 and Q (shared prior information).

Return: Q_1, Q_2, Q_3

Step Two



Fix Q_1, Q_2 , Q_3 (the subgroup specific predictor.)

-> Learn Q (shared prior information) from Q_1, Q_2 , Q_3 to be closed as possible.

Return: Q

Theoretically prove a fair and informative predictor (see paper).

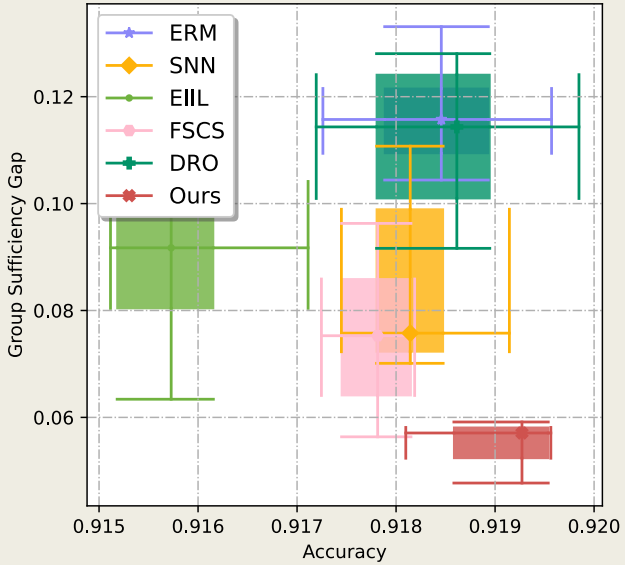
Application (Toxic comments recognition)

Toxic	Comment Text
0	I applaud your father. He was a good man! We need more like him.
0	As a Christian, I will not be patronizing any of those businesses.
0	What do Black and LGBT people have to do with bicycle licensing?
0	Government agencies track down foreign baddies and protect law-abiding white citizens. How many shows does that describe?
1	Maybe you should learn to write a coherent sentence so we can understand WTF your point is.

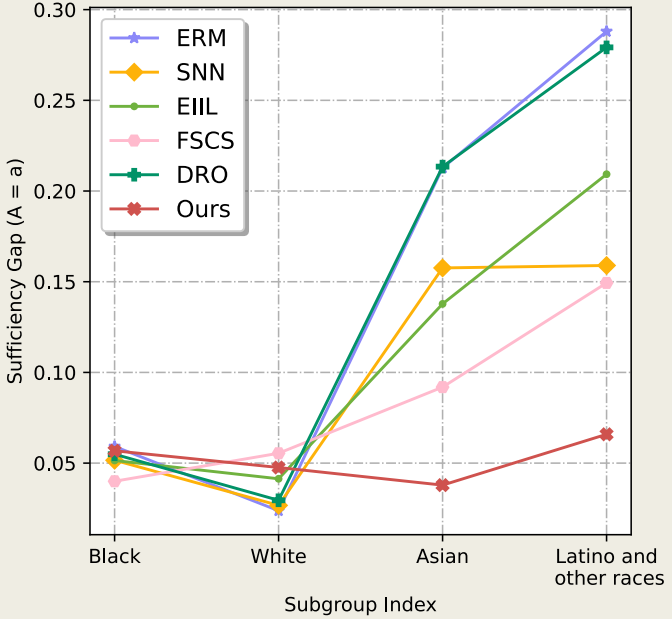
Source: https://arxiv.org/pdf/2012.07421.pdf

- Predict the comment being toxic or not.
- Machine Learning algorithms showed biased results on the different races.

Application (Toxic comments recognition)



Accuracy



Different demographics

Our Framework

higher accuracy
lower sufficiency gap

small sufficiency gap for each group

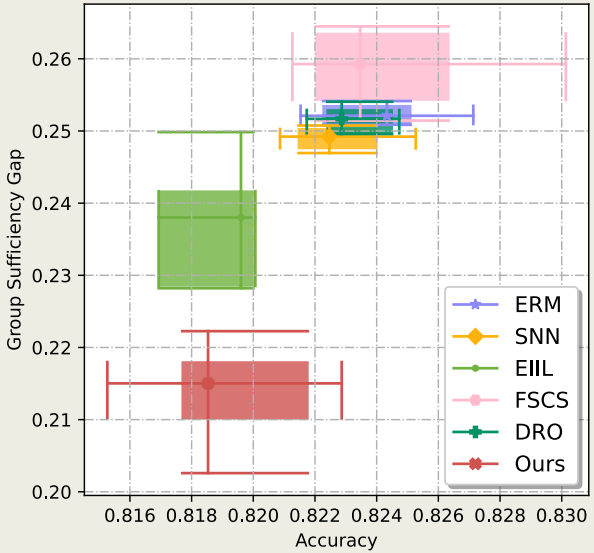
Application (Amazon reviews)

	Reviewer ID (d)	Review Text (x)	Stars (y)
Train	Reviewer 1	They are decent shoes. Material quality is good but the color fades very quickly. Not as black in person as shown.	5
		Super easy to put together. Very well built.	5
	Reviewer 2	This works well and was easy to install. The only thing I don't like is that it tilts forward a little bit and I can't figure out how to stop it.	4
		Perfect for the trail camera	5
		…	
	Reviewer 10,000	I am disappointed in the quality of these. They have significantly deteriorated in just a few uses. I am going to stick with using foil.	1
		Very sturdy especially at this price point. I have a memory foam mattress on it with nothing underneath and the slats perform well.	5

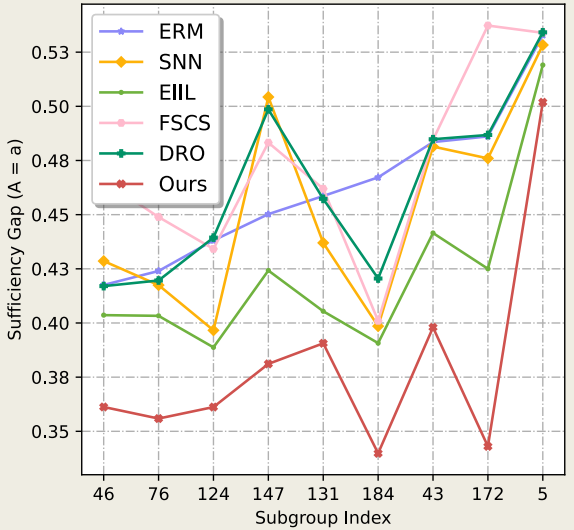
- Predict the *star* from the review.
- Machine Learning algorithms showed biased results on different clients.

Source: https://arxiv.org/pdf/2012.07421.pdf

Application (Amazon reviews)



Accuracy



Different clients

Our Framework

comparable accuracy
lower group sufficiency gap

small group sufficiency gap for each client

Conclusions

- A novel provable framework:
  - Mitigate group sufficiency bias;
  - Preserve the utility of data;

Thank you!