The motivation for a new model is to minimize algorithmic bias induced by COMPAS, which could inadvertently reinforce existent social, racial and ethnic bias and thereby achieve fairness [[1]](#footnote-0) and "mitigation of unwanted bias" . The stakeholders are 1) the ethnic minority groups, namely African-American, Hispanic, Asian, Native American and others, as otherwise they would be treated differently due to their ethnicity, which clearly violates the concept of “fairness”.For example, black people are likely to receive longer sentences than white people who committed the same crime.[[2]](#footnote-1) 2) the government. Financial income would be boosted through enforcing a fair algorithm. From the financial data, it is self-evident that the financial loss caused by incorrect prediction partly resulted from biased algorithm could be enormous when added up. 3) the society in general. Clearly,latent security threats could be diminished by implementing a better algorithm and furthermore, it could contribute to the stability of the society as unfair treatment resulted from racial disparity could trigger a myriad of problems.Biases exist because some racial groups are prone to receive longer sentence as mentioned before, which in turn impacts the result of the algorithm. Historical bias is often reflected in the data, which needs to be minimized. Also, higher observed arrest rate of some ethnicity might be a result of better policing in certain regions. In other words,if individuals in the training data have not already have equal opportunity, the data and the algorithm would be inevitably skewed. The original data and the algorithm are also both flawed since they take into account too many factors, some of which have little value.

Reason Against Demographic Parity/ Equal Opportunity:

“depending on the relationship between a protected attribute and the data, certain definitions of fairness can actually increase discrimination””we provide the formal definition of counterfactual fairness, which enforces that a distribution over possible predictions for an individual should remain unchanged in a world where an individual’s protected attributes had been different in a causal sense”

(Counterfactual fairness ,Matt J. Kusner and Joshua R. Loftus and Chris Russell and Ricardo Silva, 2017,arXiv:1703.06856v1)

Reference:

1. Jobin, Anna; Ienca, Marcello; Vayena, Effy (2 September 2019). "The global landscape of AI ethics guidelines". Nature Machine Intelligence. 1 (9): 389–399.
2. Alexander, Rudolph; Gyamerah, Jacquelyn (September 1997). "Differential Punishing of African Americans and Whites Who Possess Drugs: A Just Policy or a Continuation of the Past?". Journal of Black Studies. 28 (1): 97–111

1. [↑](#footnote-ref-0)
2. [↑](#footnote-ref-1)