

## ECONOMIC FINDING

# Measuring Racial Discrimination in Algorithms

*David Arnold, Assistant Teaching Professor, University of California, San Diego; Will Dobbie, Professor of Public Policy, Harvard Kennedy School; and Peter Hull, Assistant Professor, UChicago's Kenneth C. Griffin Dept. of Economics*

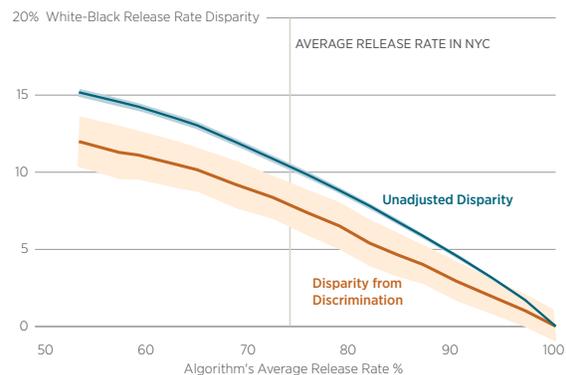
*A sophisticated machine-learning algorithm recommends the release of white defendants at a significantly higher rate in New York City than Black defendants, even after adjusting for differences in pretrial misconduct potential.*

Algorithms guide an increasingly large number of high-stakes decisions, including criminal risk assessment, resume screening, and medical testing. While such data-based decision-making may appear unbiased, there is increasing concern that it can entrench or worsen discrimination against legally protected groups. With algorithmic recommendations for pretrial release decisions, for example, a risk assessment tool may be viewed as racially discriminatory if it recommends white defendants be released before trial at a higher rate than Black defendants with equal risk of pretrial criminal misconduct.

How is it that discrimination can occur through logical, unfeeling, algorithms? The answer is in the data that feed the algorithms. Continuing with the pretrial release example, misconduct potential is only observed among the defendants who a judge chooses to release before trial. Such selection can introduce bias in algorithmic predictions but also complicate the measurement of algorithmic discrimination, since unobserved qualification cannot be conditioned on to compare white and Black treatment.

This paper develops new tools to overcome this selection challenge and measure algorithmic discrimination in New York City (NYC), home to one of the largest pretrial systems in the country. The method builds on previous techniques developed by the author to measure racial discrimination in actual bail judge decisions and leverages randomness in the assignment of judges to white and Black defendants. Applying their methods, the authors find that a sophisticated machine learning algorithm (which does not train directly on defendant race or ethnicity) recommends the release of white defendants at a significantly higher rate than Black defendants with identical pretrial misconduct potential.

**Figure 1 • Discrimination in Algorithmic Bail Decisions**



Notes: This figure plots the range of unadjusted racial disparities in algorithmic release rate recommendations, for different average release rates, along with the range of disparities due to racial discrimination. Algorithmic recommendations are from our baseline gradient-boosted decision tree model. Disparities from discrimination are computed as described in the text, using local linear estimates of the race-specific first and second moments. Shaded areas indicate pointwise 95 percent confidence intervals, computed by the bootstrapping procedure described in the working paper.

Specifically, when calibrated to the average NYC release rate of 73 percent, the algorithm recommends an 8-percentage point (11 percent) higher release rate for white defendants than equally qualified Black defendants. This unwarranted disparity explains 77 percent of the observed racial disparity in release recommendations, grows as the algorithm becomes more lenient, and is driven by discrimination among individuals who would engage in pretrial misconduct if released.