The Accuracy, Fairness, and Limits of Predicting Recidivism

Authors: Dressel, Julia & Farid, Hany

Source: Science Advances, Vol. 4, No. 1, eaao5580 (2018)

DOI/URL: https://www.science.org/doi/10.1126/sciadv.aao5580

KEY ANNOTATED PASSAGES

[Abstract [KEY CLAIM]]

The widely used commercial risk assessment software COMPAS is no more accurate or fair than predictions made by people with little or no criminal justice expertise — a 137-feature proprietary algorithm does no better than untrained crowdworkers at predicting recidivism.

[§Results — Crowd accuracy]

People from a popular online crowdsourcing marketplace — who can reasonably be assumed to have little to no expertise in criminal justice — are as accurate and fair as COMPAS at predicting recidivism (~67% both). COMPAS adds no accuracy beyond random people; it adds only opacity, proprietary control, and systematic racial bias.

[§Results — Racial bias]

Both people and COMPAS still over-predict recidivism for Black defendants and under-predict for white defendants — racial bias is not an artifact of the algorithm alone but of the training data. AI does not fix systemic racism; it encodes and amplifies it at scale.

[§Methods — Simplicity]

The same accuracy can be achieved with a simple linear classifier with only two features — the complexity of 137 features provides no accuracy benefit, yet enables Northpointe to claim proprietary protection preventing judicial review of the algorithm used in sentencing.

[§Discussion — Implications]

Risk assessment instruments like COMPAS are used for pretrial, parole, and sentencing decisions affecting human liberty — a deployed AI system with documented racial bias is a mechanism of wrongful imprisonment at scale.

RELEVANCE TO POSITION PAPER

Cited in §2 (Documented Harms) for 'wrongful imprisonment'. Demonstrates COMPAS, a commercial AI recidivism prediction algorithm used in US sentencing, is no more accurate than untrained crowdworkers at 67%, systematically over-predicts risk for Black defendants, and is used in decisions determining incarceration — a documented case of AI-caused wrongful imprisonment.