Abstract

Artificial intelligence and machine learning (AI/ML) models are increasingly utilised in every aspect of life and society due to their superhuman abilities to digest large amounts of data and find obscure patterns and correlations. One contentious area of this technological application is in the criminal justice system, where AI/ML is used as a recommendation or decision-making support tool. These applications are particularly popular in the United States of America (USA), the nation with the highest rate of incarceration and correctional budget, to aid in managing overcrowded and overspending facilities. Angwin et al.'s (2016) ground-breaking study found the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) model to be biased against Black defendants and sparked an influential academic debate around algorithmic bias and fairness. This study aims to fill the gap in the scholarship by focusing on the content of COMPAS's recidivism risk assessment questionnaire through a qualitative content analysis within the conceptual framework of Critical Race Theory (CRT). The findings presented in this research are twofold: (1) almost half of the COMPAS questions were opinion-based, thus reducing quantitative neutrality, and (2) there were significant proxy factors for race that could have led to biased results in the model. Implications of these findings are discussed.

Keywords

Algorithmic fairness, AI/ML models in policing, AI/ML models in the criminal justice system, Policing in the USA, Recidivism risk assessments, COMPAS assessment, Algorithmic bias, Disparate impact, Critical Race Theory and AI/ML, Critical content analysis