Assessment of the Efficacy of Home Electronic Incarceration (HEI) as an Alternative to Custodial Confinement

(Technical Topic)

Can AI Save the Criminal Justice System from Its Biases?

(STS Topic)

A Thesis Project Prospectus

Submitted to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree

Bachelor of Science, Systems Engineering

Stella Catherine Banino

Fall, 2023

Technical Project Team Members
George Boulos
Chris Craft
Laura Phillips
Sally Sydnor

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Advisors

William Davis, Department of Engineering and Society
K. Preston White, Department of Systems and Information Engineering
Michael Smith, Department of Systems and Information Engineering
Loreto Alonzi, School of Data Science

Overview

As algorithms are increasingly integrated into the governing structures of our society, mounting evidence demonstrates that many of these algorithms and machine-learning models are fettered by the same biases that humans hold, hurting the same populations that human decision-makers have historically discriminated against (for some examples, see O'Neil, 2016). The impact of these biased algorithms is exacerbated by the acceptance of their outcomes as "value-neutral," which hides their sinister effects on marginalized populations (Stinson, 2022, p. 769).

In the past few years, many regions in the US have begun experimenting with how data-driven algorithms can help make the criminal justice system more efficient and equitable (Freeman et al., 2020, p. 2). Currently, Black defendants and offenders are treated more severely by the criminal justice system than their white counterparts due to a variety of interrelated factors, including implicit bias (see Tonry, 2010 for an explanation of some of these factors). This implicit bias in decision-making may be exacerbated by the fatigue common in the overloaded criminal justice system (Ma et al., 2013, p. 522). This author's study team, with the support of regional criminal justice officials, has proposed sentencing algorithms as a potential solution to this issue in the Albemarle-Charlottesville criminal justice system: a well-constructed algorithm that recommends criminal sentences based on crime severity, criminal law, and potential risk to public safety could take this cognitive load off of officers and prevent implicit bias from skewing results (Carew and Morrow, 2023). However, many of these algorithms developed by other teams appear to replicate the biases that are already entrenched in the American criminal justice system, largely due to the bias of the datasets they are trained on (Stinson, 2022, p. 764). This project seeks to answer the question of how these algorithms

should be constructed to reduce bias in the broader criminal justice system and produce more equitable results for marginalized populations, exploring both the technical and ethical dilemmas inherent in this endeavour.

Technical Topic

For over a decade, the Systems and Information Engineering Department has worked with the Albemarle-Charlottesville Regional Jail (ACRJ), Region 10 (R10), and the Office of Aid and Restoration (OAR) to research and improve the Albemarle-Charlottesville criminal justice system, specifically in post-arrest processes. This year, the team will be focusing on pre-trial home electronic monitoring (HEM) and post-trial home electronic incarceration (HEI), including conditions on its efficacy and impact on recidivism rates. The team is discussing specific goals with ACRJ, R10, and OAR to identify a research direction that will have the greatest positive impact on both broader criminal justice policy in the Albemarle-Charlottesville region and, more specifically, how inmates are handled before and after sentencing and before and after release.

One of the tasks the team may work on is developing a sentencing algorithm that can help determine who would be the best fit for HEI, based on factors including demographics, resources, and offense type. Previous research conducted by earlier iterations of this team found that HEI was helpful for both Albemarle-Charlottesville inmates and their communities; not only did inmates on HEI perform better mentally, socially, and economically during and after their incarceration period, but also, after release, they were less likely to commit another offense during the study period, keeping the community safer (Dornfeld et al., 2023). HEI also saves the ACRJ money and capacity (Kumer, 2023). Currently, the jail decides who goes on HEI through a rather ad-hoc process of file sorting (Kumer, 2023), so the team this year is looking to

streamline that process in a way that will make it easier to offer inmates HEI and expand the program.

One of the dimensions of HEI expansion the team may investigate is whether a defendant's mental health condition, especially experience of a serious mental illness (SMI), does, and whether it should, affect the rate at which defendants are recommended for release on HEI. The main criteria currently used to determine which inmates are recommended for HEI are whether the inmate has a stable home to return to and whether they're likely to violate their HEI conditions or commit another crime during their incarceration period (Kumer, 2023). Both of these criteria could prevent people with SMIs from entering the HEI program. First, people with SMIs are more likely to be homeless (Abt Associates, 2015) and thus many of them cannot participate in the HEI program on these grounds. Others live with family and friends who act as caretakers and may prefer their wards stay in the jail so the caretaker can take a much-needed break from caretaking duties (Kumer, 2023). As for the second HEI criterion, people with SMIs have higher recidivism rates than those without SMIs, both in the Albemarle-Charlottesville system (Corbin et al., 2022) and in the broader country, where recidivism rates for people with SMIs hover around 70% depending on the study (von Hemert, 2023). This could be due to a variety of factors, including poverty-related crimes and substance abuse, as both of these affect people with SMIs at a higher rate (von Hemert, 2023).

This second criterion in particular leads to an ethical issue - it would be accurate to say that a person with an SMI is more likely to commit another offense, but should the fact that a person has an SMI be used as a reason not to offer them HEI? There's no easy answer to that question. Currently, judges in the Albemarle-Charlottesville region are not told whether

defendants have an SMI during bond hearings, but are allowed to consider this information during sentencing trials (Carew and Morrow, 2023).

The team is working in part with data collected through the risk assessment algorithm COMPAS. COMPAS, and algorithms like it, use data including a person's criminal history and demographics to determine the likelihood they will commit another crime (Jackson and Mendoza, 2020). COMPAS outputs a scalar to indicate this likelihood, which is then used to determine allocations of services and resources aimed at preventing recidivism. More controversially (see, e.g., Carlson, 2017), this score is also used in sentencing to predict the risk the defendant poses to public safety, and thus whether or not their sentence should include prison time (Jackson and Mendoza, 2020). The animus for this controversy is the claim that these risk assessment algorithms, including COMPAS, overestimate the risk posed by Black defendants (see, e.g., Angwin et al., 2016). Even though race is not explicitly included in the algorithms, it forms a part of variables that can disproportionately target Black people. For example, past criminal history can seem race-neutral on its face, but since Black neighborhoods have a higher police presence than white neighborhoods, residents of those neighborhoods have more contact with police and are thus more likely to be charged with minor crimes (Alonzi et al., 2023). Partly in response to this controversy, the Albemarle-Charlottesville system doesn't use this risk assessment algorithm's output score, but it does use the individual, and arguably flawed variables to make case-by-case decisions on sentencing and HEI recommendations (Alonzi et al., 2023).

Science, Technology, and Society Topic

As the team works with and attempts to develop an algorithm to assist in sentencing, this science, technology, and society paper will examine the ethics and equity perspective of these algorithms. First, it will consider what outcomes would be desirable for different stakeholder

groups, what each group is ethically owed, and what constraints are placed on each group by ethical concerns. For example, the government's criminal justice system is motivated by a desire to increase public safety by lowering crime rates. They are owed compliance from civilians on the condition that they act lawfully and for the public benefit, and can only be expected to act within the bounds of what their resources allow. They are constrained in their investigation of crimes and enforcement of punishments by ethical concerns including the right to privacy. More in depth analysis will be applied to this stakeholder group and also to the public, defendants, and community groups that are involved in the criminal justice system. Any tensions between stakeholder groups will be considered, and weight will be given to each concern.

The second part of this paper will use this analysis of stakeholder ethics to consider what variables and information are ethical to include in sentencing algorithms and how the outputs of these algorithms should be used. It will examine questions like the one posed earlier in the paper about whether it is ethical to consider mental illness in sentencing decisions, along with other variables that may increase risk prediction accuracy, but could result in severer sentences on the basis of racial, economic, or other minority status. In answering this question, it will look at arguments like the one made by Skeem and Lowenkamp (2020, pp. 274-276) that including sensitive information like race in algorithms can actually improve outcomes for marginalized populations. It will also go beyond this to decide whether including information that does result in negative consequences for minorities is permissible and ethical if it increases predictive accuracy and improves public safety, as examined through the lens of stakeholder ethics.

Much of the work done so far in this field has focused on risk assessment tools like COMPAS and whether their outputs and use are or can be made equitable (see, e.g., Freeman et al., 2020). This paper will direct its main attention to sentencing algorithms, which provide

recommendations to judges on sentence severity given inputs that include many of the same variables as risk assessment tools, as well as additional information on evidence and criminal law (Stobbs et al., 2017, pp. 3-4). This paper will touch on some of the arguments written about risk assessment tools, specifically on transparency and data selection, and apply them to sentencing algorithms. These include Carlson's call for the full disclosure of algorithms used for sentencing (2017, pp. 324-329), Peeters and Schuilenberg's (2018, pp. 275-277) and Brayne and Christin (2021, pp. 619-622)'s critiques of algorithms' ability to shift accountability away from public-facing offices to less transparent areas of bureaucracy, and Freeman et al.'s claim (2020, p. 11) that variables and algorithmic formation methods can be selected in such a way that their use is more equitable than relying on human decision-makers. It will also consider sources that specifically compare the equitability of human sentencing vs. sentencing algorithms, including Bagaric and Hunter's 2022 literature review on the subject.

Conclusion

The technical and STS aspects of this project aim not only to streamline the criminal justice systems in Charlottesville and Albemarle County, but also to ensure equitable and effective outcomes for all stakeholders: defendants, inmates, community members, community organizations, courts, and jails. The team aims to develop an algorithm that can reduce implicit bias by shifting some of the decision-making load from human decision-makers to this algorithm, troubleshooting common ways that bias can enter this type of algorithm. The STS portion of this project will inform the technical portion, both in what information it is ethical for the algorithm to consider and how its output can be used to protect both defendants and public safety. Through research on both the technical and social aspects, this project will ideally

establish whether algorithmic sentencing is the best way to achieve these goals and if so, what methods should be employed towards this end.

References

Abt Associates. (n.d.). *2015 AHAR: Part 1 - PIT Estimates of Homelessness in the U.S.* Retrieved October 19, 2023, from https://www.hudexchange.info/resource/4832/2015-ahar-part-1-pit-estimates-of-homelessness

Alonzi, L. P., White, K. P., & Smith, M. (2023). *Capstone Project Meetings, Fall 2023* [Personal communication].

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (n.d.). *Machine Bias*. ProPublica. Retrieved October 18, 2023, from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bagaric, M., & Hunter, D. (2022). Enhancing the Integrity of the Sentencing Process through the Use of Artificial Intelligence. In J. Ryberg & J. V. Roberts (Eds.), *Sentencing and Artificial Intelligence* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780197539538.003.0007

Brayne, S., & Christin, A. (2021). Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts. *Social Problems*, *68*(3), 608–624. https://doi.org/10.1093/socpro/spaa004

Carlson, A. M. (2017). The Need for Transparency in the Age of Predictive Sentencing Algorithms. *Iowa Law Review*, *103*(1), 303–329.

Corbin, G., Dale, N., Deshpande, A., Korngiebel, K., Krablin, P., Wilt, E., Alonzi, L. P., Goodloe, N., Smith, M., & White, K. P. (2022). Evaluating Administered Differences of Brief Jail Mental Health Screener and Impacts of Diagnoses & Treatment of Linked

Inmates with Severe Mental Illness. *2022 Systems and Information Engineering Design Symposium (SIEDS)*, 351–356. https://doi.org/10.1109/SIEDS55548.2022.9799360

Carew, W. R., & Morrow, S. (2023, October 24). *Meeting between OAR and the UVA 2024 Capstone Team* [Personal communication].

Dornfeld, J., Hankinson, I., Hughes, L., Murphy, S., Peraka, R., Rawson, M., Alonzi, L. P., Smith, M., White, K. P., & Goodloe, N. (2023). Analyzing Efficacy of Home Electronic Incarceration on Return-to-Custody Rates for Inmates During the COVID-19 Pandemic. *2023 Systems and Information Engineering Design Symposium (SIEDS)*, 72–77. https://doi.org/10.1109/SIEDS58326.2023.10137861

Freeman, K. R., Hu, C., & Jannetta, J. (2021). Racial Equity and Criminal Justice Risk Assessment. *Urban Institute Justice Policy Center*.

Jackson, E., & Mendoza, C. (2020). Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not. *Harvard Data Science Review*, *2*(1). https://doi.org/10.1162/99608f92.1b3dadaa

Kumer, M. (2023, September 19). *Initial Interview with Martin Kumer, Superintendent of the Albemarle-Charlottesville Regional Jail* [Personal communication].

Ma, D. S., Correll, J., Wittenbrink, B., Bar-Anan, Y., Sriram, N., & Nosek, B. A. (2013). When Fatigue Turns Deadly: The Association Between Fatigue and Racial Bias in the Decision to Shoot. *Basic and Applied Social Psychology*, *35*(6), 515–524. https://doi.org/10.1080/01973533.2013.840630

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books.

Peeters, R., & Schuilenburg, M. (2018). Machine justice: Governing security through the bureaucracy of algorithms. *Information Polity: The International Journal of Government & Democracy in the Information Age*, *23*(3), 267–280. https://doi.org/10.3233/IP-180074

Stinson, C. (2022). Algorithms are not neutral: Bias in collaborative filtering. *AI and Ethics*, *2*(4), 763–770. https://doi.org/10.1007/s43681-022-00136-w

Stobbs, D. N., Hunter, D., & Bagaric, M. (2017). *Can Sentencing Be Enhanced by the Use of Artificial Intelligence?*

Tonry, M. (2010). The Social, Psychological, and Political Causes of Racial Disparities in the American Criminal Justice System. *Crime and Justice*, *39*, 273–312. https://doi.org/10.1086/653045

von Hemert, T. (2023, October 10). *The Mental Health Crisis and Solutions in the Charlottesville-Albemarle Area*.