ARTICLES

Beyond Bias: Re-imagining the Terms of "Ethical AI" in Criminal Law

CHELSEA BARABAS*

ABSTRACT

Data-driven decision-making regimes, often branded as "artificial intelligence" (AI), are rapidly proliferating across the U.S. criminal legal system as a means of managing the risk of crime and addressing accusations of discriminatory practices. However, these data regimes have come under increased scrutiny, as critics point out the myriad ways that they reproduce or even amplify pre-existing biases in the criminal legal system. As a result, a new regulatory science is emerging to render AI "fair" in this context. This essay examines two general approaches to "algorithmic fairness" in criminal law: 1) formal fairness criteria that illustrate the trade-offs of different algorithmic design choices and 2) managerialist standards for maintaining a baseline of accuracy, transparency and validity in these systems. Attempts to render AI-branded tools more accurate by addressing narrow notions of "bias," miss the deeper methodological and epistemological issues regarding the fairness of the outcomes that these tools produce. The key questions are whether predictive tools reflect and reinforce punitive practices that drive disparate outcomes, and how data regimes interact with the penal ideology to naturalize these practices. The article concludes by calling for an abolitionist understanding of the role and function of the carceral state, in order to fundamentally reformulate the problems we aim to solve, the way we make claims based on existing data, and how we identify and fill gaps in the data regimes of the carceral state.

Keywords: artificial intelligence, abolition, algorithm, policing, surveillance, governance, justice

TABLE OF CONTENTS

* Chelsea is in the doctoral program of Media, Arts and Sciences at the Massachusetts Institute of Technology. Segments of this article appear as part of a book chapter in the Oxford Handbook of Ethics and AI (July, 2020).

I. INTRODUCTION

Data-driven decision-making regimes are rapidly proliferating across the U.S. criminal legal system as a means of managing the risk of crime and addressing accusations of discriminatory practices.[1] Yet these data regimes, often branded as artificial intelligence ("AI"), have come under increased scrutiny as critics point out the myriad ways that they can reproduce or even amplify pre-existing biases in the criminal legal system.[2] In response, technology vendors and scholars have proposed some technical adjustments for rendering these tools "unbiased," "valid," and "fair."[3]

These efforts have coalesced around two general approaches to, what has been widely branded as, "algorithmic fairness:" 1) the development of formal fairness criteria that illustrate the trade-offs of different algorithmic design choices and 2) the development of managerialist "best practices" for maintaining a baseline of accuracy, transparency, and validity in algorithmic systems. In both of these approaches, efforts to reduce bias and maximize accuracy figure prominently. Yet, technocratic notions of bias and accuracy serve as inadequate conceptual anchors for the debate regarding the ethical use of algorithms in criminal law, because they fail to interrogate the deeper normative, theoretical, and methodological premises of these tools.

Contemporary discourse regarding "fair, accountable, and transparent" algorithms represents the most recent incarnation of a long historical struggle over how to

1. JACKIE WANG, CARCERAL CAPITALISM 252 (2018).

2. *See generally* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671 (2016); Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROCEEDINGS MACHINE LEARNING RSCH. 1 (2018); Alexandra Chouldechova, *Fair Prediction With Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*, 5 BIG DATA 153, 153-63 (2017).

3. Richard Berk et al., *Fairness in Criminal Justice Risk Assessments: The State of the Art*, 47 SOC. METHODS RSCH. 1, 13 (2018) [hereinafter *Fairness in Risk Assessments*]; RICHARD BERK, MACHINE LEARNING RISK ASSESSMENTS IN CRIMINAL JUSTICE SETTINGS 115-26 (2019) [hereinafter MACHINE LEARNING RISK ASSESSMENTS]; Cynthia Dwork et al., *Fairness Through Awareness*, 3 INNOVATIONS THEORETICAL COMPUTER SCI. 214, 222 (2012); Jon Kleinberg et al., *Algorithmic Fairness*, 108 AEA PAPERS & PROC. 22, 23 (2018) [hereinafter *Algorithmic Fairness*]; Sharad Goel et al., *The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment, in* RESEARCH HANDBOOK ON BIG DATA LAW (forthcoming 2020).

properly interpret criminal legal data.[4] These data directly reflect the allocation of law enforcement resources and priorities, rather than rates of criminal activity across the population.[5] Yet, AI tools in the criminal legal system often rely on arrest and conviction data as a means of predicting criminal activity and threats to public safety.

By closely examining the current debate regarding the use of pretrial risk assessment, this article illustrates the ways that predominant practices of data collection, labeling, and analysis are produced by, and ultimately reproduce, a specific penal ideology that justifies unwarranted punishment and harm. Predictive tools reflect and reinforce punitive practices that drive racially disparate outcomes, by rendering these practices seemingly "objective and fair." This article calls for a significant reformulation of the key concepts and assumptions which undergird the adoption of algorithmic technologies in criminal law, shifting away from measuring criminal proclivities and towards understanding processes of criminalization, from supporting law and order to increasing community safety and self-determination, and from surveillance of risky populations to accountability of state officials.

II. AI AND STATISTICAL DISCOURSE IN CRIMINAL LAW

The term AI is not new. Since the 1950s, AI has been used in reference to a wide range of computational methods in both academia and popular culture.[6] Rather than attempt to draw clear methodological boundaries around what constitutes AI, we should understand the term as a sociotechnical concept, one which includes a diverse set of technocratic practices and logics.[7] In the context of the U.S. penal system, the term AI has emerged as a contested vehicle for "evidence-based reform," the most recent instance in a long history of state efforts to wield statistics as a means of depoliticizing state violence and reasserting legitimacy amid significant social change and unrest.[8] For over a century, crime statistics have been at the center of intense ideological struggles over how to characterize the role of the carceral state in maintaining public safety and managing criminal behavior.[9]

AI cannot be cleanly distinguished from other modes of data analysis within criminal law. The term AI has been used in reference to a hodgepodge of computational methods, which range from old school statistics like linear regression to new machine learning algorithms, which feed off unprecedented amounts of digital data.[10] The data used to build these systems are typically administrative records collected by local

4. KHALIL GIBRAN MUHAMMAD, THE CONDEMNATION OF BLACKNESS 5 (2011).

5. DELBERT S. ELLIOTT, LIES, DAMN LIES, AND ARREST STATISTICS, CTR. STUDY PREVENTION VIOLENCE 8 (1995).

6. PAMELA MCCORDUCK, MACHINES WHO THINK: A PERSONAL INQUIRY INTO THE HISTORY AND PROSPECTS OF ARTIFICIAL INTELLIGENCE 114 (2nd ed. 2009); Madeleine Clare Elish & Danah Boyd, *Situating Methods in the Magic of Big Data and AI*, 85 COMM. MONOGRAPHS 57, 62 (2018).

7. Elish & Boyd, *supra* note 6, at 62.

8. David Garland, *Criminological Knowledge and Its Relation to Power—Foucault's Genealogy and Criminology Today*, 32 BRITISH J. CRIMINOLOGY 403, 417 (1992); NAOMI MURAKAWA, THE FIRST CIVIL RIGHT: HOW LIBERALS BUILT PRISON AMERICA (2014); Tony Platt, *"Street" Crime—A View from the Left*, 9 CRIME & SOC. JUST. 26, 26 (1978).

9. MUHAMMAD, *supra* note 4, at 5.

10. Sarah Brayne, *Big Data Surveillance: The Case of Policing*, 82 AM. SOC. REV. 977, 977 (2017).

police departments and administrations of the court.[11] However, law enforcement agencies have also begun to partner with technology companies to integrate new forms of consumer surveillance, such as Amazon's Ring, into their data systems.[12] Generally speaking, most contemporary AI technologies aim to measure the strength of associations between a set of data inputs and an outcome of interest. These measurements are correlational at their core. Their outputs come in the form of probabilistic distributions to forecast or predict future events.

In the field of criminology, the interpretation of crime data has always served as a critical point of departure between positivist subfields of the discipline—seeking to measure and manage criminal behavior in "risky" populations—and critical scholars, who conceive of crime as primarily the by-product of criminalizing discourses and practices carried out by the carceral state. Critical scholars situate scholarship in the social sciences, and particularly mainstream criminology, as part of an interconnected set of practices which criminalize and contain Black life, while also serving as the very justification for the continuation of such practices.[13] In *The Condemnation of Blackness: Race, Crime and the Making of Modern America*, historian Khalil Gibran Muhammad traces the emergence of crime statistics as a central aspect of the discourse regarding the disparate treatment of different racial populations at the turn of the nineteenth century. The demographics of American cities were rapidly changing through industrialization and an influx of migration to urban hubs in the north when the nascent fields of sociology and criminology emerged, with a particular interest in studying the "defective and delinquent classes" arriving en masse.[14] Crime statistics and the study of criminal behavior justified a new "carceral bargain,"[15] where anti-black policies of segregation, racial violence, and incarceration systematically excluded African Americans from the broader public sphere.[16]

During a time when bioscientific theories regarding the moral inferiority of Black people were crumbling under empirical scrutiny, statistics regarding the overrepresentation of African Americans in U.S. prisons served as the empirical basis for pseudo-scientific arguments about the inherent criminality of the Black population. These theories were developed and propagated by a number of influential academics.[17] At the same time, these scholars, along with liberal reformers in the North, pointed to statistics about the overrepresentation of European immigrants in the penal system as a call to action for poverty alleviation during the Progressive Era. European criminality was attributed to structural inequalities and poverty, whereas

11. Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109, 1123-24 (2017).

12. Alfred Ng, *Amazon's helping police build a surveillance network with Ring doorbells*, C-NET (June 5, 2019), https://www.cnet.com/features/amazons-helping-police-build-a-surveillance-network-with-ring-doorbells/ [https://perma.cc/WML7-86YX].

13. Amna A. Akbar, *Toward a Radical Imagination of Law*, 93 N.Y.U. L. REV. 405, 410-11 (2018); MUHAMMAD, *supra* note 4.

14. MUHAMMAD, *supra* note 4, at 46.

15. Sharon Dolovich, *Exclusion and Control in the Carceral State*, 16 BERKELEY J. CRIM. L. 259, 274 (2011).

16. MUHAMMAD, *supra* note 4, at 94–95.

17. *Id.*

Black criminality was understood to be rooted in personal pathologies and inherent cultural inferiority.[18] According to Muhammad, this inconsistent interpretation of crime statistics laid the basis for a new "scientific" notion of Black criminality, while simultaneously creating new avenues for formerly criminalized populations of European immigrants to receive more robust social support.[19] As Black Americans were criminalized via statistical discourse, the public became increasingly sympathetic to the plights of poor European Americans as they assimilated into the category of whiteness via these same statistics. In this context, crime statistics were not only the by-product of racist ideas and policies, but the very justification for them.[20]

This inconsistent and racialized interpretation of crime statistics was challenged by a number of academics at the time, particularly African-American scholars coming out of the Atlanta School of Sociology.[21] They argued that crime statistics were more a reflection of racialized law enforcement practices and harsh social conditions than inherent differences in group proclivities towards crime. They were also quick to point out the inconsistencies in how crime statistics were interpreted, suggesting that Blackness had become a "glue that binds race to crime" such that white people can commit crime but Black people are deemed 'criminals.'[22] But these critiques were systematically suppressed by mainstream schools of thought led by white scholars at prestigious institutions, particularly the Chicago School of Sociology, which was immensely influential in shaping the theoretical and analytical foundations of criminology as an academic discipline.[23] As many scholars have argued, mainstream criminology became an intellectual prosthesis for the state, providing authoritative

18. This framing of racialized criminal behavior has prevailed in criminological discourse for well over a century. It figured prominently in high-profile policy reports, such as Daniel Moynihan's "The Negro Family: The Case for National Action" in the 1960's and persists up to today via mainstream theories of criminality, such as social learning theory. Some researchers argue that this framing of criminality has remained popular because it is supports individual-level crime reduction interventions, which are often viewed as more pragmatic than structural efforts to reduce the root drivers of crime, such as poverty and lack of opportunity. Seth J. Prins & Adam Reich, *Can We Avoid Reductionism in Risk Reduction?*, 22 THEORETICAL CRIMINOLOGY 258 (2018).

19. MUHAMMAD, *supra* note 4, at 75-76.

20. Muhammad provides a number of examples which illustrate the ways crime statistics were used to justify the exclusion of African-Americans from key public services and programs. For example, Mississippi governor James Vardaman argued in 1905 that:

> To school the negro is to increase his criminality. Official statistics do not lie, and they tell us that the Negroes who can read and write are more criminal than the illiterate. In New England, where they are best educated, they are four and a half times as criminals as they are in the Black Belt, where they are most ignorant. The more money for Negro education, the more Negro crime. This is the unmistakable showing of the United States Census.

Khalil Gibran Muhammad, *How Numbers Lie: Intersectional Violence and the Quantification of Race* 45:45 (2016), https://www.youtube.com/watch?v=br0ZYTGuW9M&t=2713s [https://perma.cc/C22S-76PU].

21. MUHAMMAD, *supra* note 4; ALDON MORRIS, THE SCHOLAR DENIED: W.E.B. DU BOIS AND THE BIRTH OF MODERN SOCIOLOGY 81 (2017).

22. MUHAMMAD, *supra* note 4, at 1.

23. MORRIS, *supra* note 21, at 2.

narratives in support of expansions in police power and incarceration by pathologizing individuals and their specific cultural contexts as "anti-social" or "at risk."[24]

III. THE CRISIS OF LEGITIMACY OF THE CARCERAL STATE

Contemporary debates regarding the adoption of AI in criminal law are occurring at a time when the legitimacy of the carceral state has come into question.[25] Over the last five decades, rates of incarceration in the United States have dramatically increased, reaching epic proportions. The U.S. incarcerates the largest number of people in the world, at a rate that is four times greater than the global average.[26] For people of color, the threat of incarceration is grossly disproportionate to their representation in the general population as African Americans are incarcerated more than six times the rate of Whites, and Latinx are incarcerated at more than double the rate of Whites.[27] A growing body of scholarship has documented the unprecedented scale and impact of discriminatory law enforcement practices[28] and racialized mass incarceration,[29] emphasizing that these developments are neither natural nor sustainable.[30] As these trends garner growing public concern, law enforcement agencies have sought to "upgrade" their tools and the popular discourse regarding their work.[31]

Over the last two decades, law enforcement strategies and tactics have undergone significant changes, fueling the adoption of controversial data-intensive surveillance technologies. In the wake of the September 11 terrorist attacks, the militarization of domestic law enforcement highlights a larger shift in policing, away from reaction to pre-emption and from deterrence to intelligence.[32] In this "securocratic" era of law enforcement,[33] there has been a massive expansion in digital data collection efforts,

24. Michelle Brown & Judah Schept, *New Abolition, Criminology and a Critical Carceral Studies*, 19 PUNISHMENT & SOC'Y 440, (2017); STANLEY COHEN, AGAINST CRIMINOLOGY 52-53 (2017); Garland, *supra* note 8, at 413; Tony Platt, *Prospects for a Radical Criminology in the United States*, 1 CRIME & SOC. JUST. 2 (1974).

25. Allegra M. McLeod, *Envisioning Abolition Democracy*, 132 HARV. L. REV. 1613, 1615 (2018).

26. Christopher Hartney, *U.S. Rates of Incarceration: A Global Perspective*, NAT'L COUNCIL ON CRIME & DELINQUENCY (2006), https://www.nccdglobal.org/sites/default/files/publication_pdf/factsheet-us-incarceration.pdf [https://perma.cc/9JAN-3YQ2].

27. *Id.*

28. Michael W. Sances & Hye Young You, *Who Pays for Government? Descriptive Representation and Exploitative Revenue Sources*, 79 J. POLS. 1090 (2017); *see also* Emma Pierson et al., *A Large-scale Analysis of Racial Disparities in Police Stops Across the United States*, NAT. HUMAN BEHAVIOR (conditionally accepted 2020).

29. Dolovich, *supra* note 15; Craig Haney, *The Psychological Impact of Incarceration: Implications for Post-prison Adjustment*, *in* PRISONERS ONCE REMOVED 33-66 (Jeremy Travis & Michelle Waul, eds., 2003).

30. MICHELLE ALEXANDER, THE NEW JIM CROW: MASS INCARCERATION IN THE AGE OF COLORBLINDNESS 60 (2012); Kelly Lytle Hernández, Khalil Gibran Muhammad & Heather Ann Thompson, *Introduction: Constructing the Carceral State*, 102 J. AM. HIST. 18, 19 (2015); MURAKAWA, *supra* note 8; WANG, *supra* note 1, at 297; BRUCE WESTERN, PUNISHMENT AND INEQUALITY IN AMERICA 14-15 (2006).

31. Ruha Benjamin, *Catching Our Breath: Critical Race STS and the Carceral Imagination*, 2 ENGAGING SCI. TECH. & SOC'Y 145, 145, 149 (2016).

32. STEPHEN GRAHAM, DISRUPTED CITIES 13-38 (2010).

33. Allen Feldman, *Securocratic Wars of Public Safety: Globalized Policing as Scopic Regime*, 6 INTERVENTIONS 330, 331 (2004).

including the adoption of closed circuit camera networks and acoustic sensors, federally funded police body cameras, and biometric data collection terminals at airports and border points.[34]

Forensic DNA databases have also rapidly proliferated.[35] Dragnet surveillance technologies, such as "stingray" cellular tracking devices, enable police to track an unprecedented amount of digital communications data without the owner's knowledge or consent.[36] "Smart" law enforcement technologies, such as GPS tracking devices and biometric monitors, have been heralded as the next major growth industry for "innovation solutions" in law enforcement.[37] These technologies collect massive amounts of data regarding the activities of individuals under an ever-widening net of state suspicion and control.

These developments are exacerbated by the fact that *access* to this data has also expanded across law enforcement and other state agencies, through the formation of cross-jurisdictional task forces, regional intelligence centers and shared gang databases.[38] Institutions that provide medical, financial, labor market, and educational services have also become "surveilling institutions"[39] for the carceral state, further blurring the line between basic social supports and carceral surveillance of poor populations and communities of color.[40]

Furthermore, in the aftermath of the 2007 financial crash, U.S. cities underwent significant demographic shifts, as urban centers became gentrified and the suburbs increasingly became home to low income communities of color.[41] These demographic changes precipitated a shift in law enforcement practices, which emphasize highly spacialized and militarized policing of communities of color, effectively transforming minority locales into prison-like spaces. As Daniel Kato argues,

> Because of the blurring of the urban/suburban divide, it became harder to identify "intruders," and hence a demand for a different kind of policing. The discursive

34. Daniel Kato, *Carceral State 2.0?: From Enclosure to Control & Punishment to Surveillance*, 39 NEW POL. SCI. 198, 204 (2017).

35. The proliferation of DNA databases is fueled by an expansion in the eligibility criteria for compulsory data collection in most states, from a small set of people convicted of violent felonies to individuals who have not yet been convicted of a crime. When the first DNA databases were started, the focus was on collecting biometric data from a small set of violent felons and sex offenders. But today it is commonplace for all people convicted of any crime, regardless of how serious the charge, to contribute DNA data. In at least twenty-nine states, even some categories of arrestees are required to provide DNA samples. Elizabeth E. Joh, *Policing by Numbers: Big Data and the Fourth Amendment*, 89 WASH L. REV. 35, 51 (2014).

36. ADAM BATES, STINGRAY: A NEW FRONTIER IN POLICE SURVEILLANCE, CATO INST. (2017), https://www.cato.org/sites/cato.org/files/pubs/pdf/pa-809-revised.pdf [https://perma.cc/3JQB-BHS5].

37. THE GEO GROUP, INC., *Electronic Monitoring, Complete Electronic Monitoring Solutions Provider* (2019), https://www.geogroup.com/Electronic_Monitoring [https://perma.cc/S98X-JB29].

38. Torin Monahan, *The Future of Security? Surveillance Operations at Homeland Security Fusion Centers*, 37 SOC. JUST. 84, 84 (2010); Sarah Brayne, *Surveillance and System Avoidance: Criminal Justice Contact and Institutional Attachment*, 79 AM. SOC. REV. 367, 368 (2014).

39. Brayne, *supra* note 38.

40. *Id.*; VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR 121 (2018).

41. LEIGH GALLAGHER, THE END OF THE SUBURBS: WHERE THE AMERICAN DREAM IS MOVING 177-80 (2014); Kato, *supra* note 34.

> terms of pre-emption that emerged out of the war on terror and the disproportionate targeting of people of color legitimated the shift in policing that was more active and less accountable.[42]

These demographic shifts brought about a new regime of heightened surveillance and punishment, one which is rooted in the police's power to surveil, search, and deploy lethal force.[43] The 2008 recession also resulted in reduced local law enforcement budgets, which precipitated the adoption of new surveillance and profiling technologies in an effort to pursue operational efficiencies and novel means of extracting revenue from local populations.[44]

Amidst these shifts in law enforcement and courtroom practices, abolitionist social movements have swelled, as more people seek to fundamentally challenge police brutality, state surveillance, and the use of lethal force against marginalized groups. In recent times, high-profile social movements such as Black Lives Matter explicitly align their cause with an abolitionist theory of change, arguing that policing has historically served as a force of violence in Black communities, which underpins a system of racial capitalism and fundamentally limits the life chances of people of color.[45]

In the context of widespread concerns about mass incarceration, a growing number of state and local advocacy groups are calling for the wholesale elimination of punitive practices that fuel detention, such as the use of cash bail.[46] In 2016, incarcerated workers organized one of the largest and most prolonged prison strikes in U.S. history. Activist scholars have characterized the strike as a part of a sophisticated abolitionist strategy to disrupt the reproduction of the carceral state apparatus by halting the daily activities that sustain prison operations.[47]

Moreover, social movements make explicit links between state surveillance technologies and the perpetuation of mass incarceration and police brutality. Grassroots organizations, such as the Stop LAPD Spying Coalition, Mijente, the Carceral Tech Resistance Network, Media Justice, and the Movement Alliance Project, are working

---

42. Kato, *supra* note 34, at 216.

43. While this era is marked by a heightened sense of digital surveillance in suburban areas, it is also important to remember that black communities, as well as other communities of color, have been subjected to surveillance technologies for a very long time. As Simone Browne explains, "surveillance is nothing new for black folks." Indeed, surveillance technologies have long coproduced notions of blackness, providing justification for the exclusion and confinement of black people from everyday life. SIMONE BROWNE, DARK MATTERS: ON THE SURVEILLANCE OF BLACKNESS 7 (2015).

44. Sances & You, *supra* note 28.

45. Akbar, *supra* note 13; Patrisse Khan-Cullors, *Abolition and Reparations: Histories of Resistance, Transformative Justice, and Accountability*, 132 HARV. L. REV. 1682 (2018); PATRISSE KHAN-CULLORS & ASHA BANDELE, WHEN THEY CALL YOU A TERRORIST: A BLACK LIVES MATTER MEMOIRE (2017); RUTH WILSON GILMORE, GOLDEN GULAG: PRISONS, SURPLUS, CRISIS, AND OPPOSITION IN GLOBALIZING CALIFORNIA 28 (2007); WANG, *supra* note 1, at 264.

46. COURTWATCH MA, *Stories from Court*, https://www.courtwatchma.org/stories-from-court.html [https://perma.cc/9UG4-48U6] (last visited Aug. 1, 2020); Eric Holder, Jr., *Memorandum re: Cook County's Wealth Based Pretrial System* (2017), http://www.chicagoappleseed.org/wp-content/uploads/2017/11/Holder_Cook-Countys-Wealth-Based-Pretrial-System-2017-07-12.pdf [https://perma.cc/N7JU-VP7M].

47. Alejo Stark, *Like a Game of Chess: The Prison Strike and Abolitionist Strategy*, ABOLITION J. (2018), https://abolitionjournal.org/like-a-game-of-chess/ [https://perma.cc/R343-ET3N].

at the local level to resist the adoption of data-driven technologies that legitimize harmful policing practices and undermine due process in the courts. Activist organizations and projects such as Data for Black Lives, the Black Futures Lab, and Our Data Bodies, have developed specific strategies to resist and invert prevailing narratives around the utility of government data, in order to reframe key debates around the use of data analytics for social justice. While many of the technological developments and law enforcement practices under fire today are not new, they have become a site of resistance for a growing number of social movements, which seek to challenge the carceral logics underpinning widespread criminalization and racialized premature death.[48]

These social movements are further bolstered by a growing body of key academic texts which delineate discriminatory police practices and reveal the criminogenic logics of the U.S. penal system. Uprisings in Ferguson, Missouri gave rise to research which calls attention to the ways police departments across the country systematically target Black communities for fines and fees in an effort to remain financially solvent.[49] Others have illustrated the numerous ways that the experience of incarceration strips individuals of the prosocial capacities and resources necessary to cope with the free world once they are released from prison.[50] This research compellingly illustrates the ways that practices within the U.S. criminal legal system are anchored in racialized state violence and the criminalization of poverty.

In this context, abolitionist leaders and scholars assert a political vision that centers the creation of lasting alternatives to punishment and imprisonment through the elimination of the tools and practices that legitimate and perpetuate the expansion of the carceral state.[51] Abolitionist challenges to this system are rooted deeply in the territory of epistemology — abolition aims to fundamentally reformulate the key concepts which guide criminological work, such as crime, punishment, and safety.[52] A practical aspect of this struggle has always been to challenge key assumptions that guide reform efforts, which have historically served to stabilize and expand the carceral state.

The goal of the abolitionist is to move beyond the default logics and assumptions of the carceral state, to address the foundational violence of law enforcement and courtroom practices. As Michelle Brown and Judah Schept argue, the abolitionist struggle is about "the dismantling, changing, and building anew—of the normative discourses and vocabularies, the ways of thinking and being, that constitute the conditions of the prison–industrial complex's ("PIC") possibility and which derive their legitimacy in part from criminology (Foucault 1980)."[53] Abolitionists seek to assert an alternative set of values to guide the pursuit of safe communities, such as self-determination and increased accountability for state authority figures. In pursuit of

48. GILMORE, *supra* note 45, at 28.
49. Sances & You, *supra* note 28.
50. Dolovich, *supra* note 15; Haney, *supra* note 29.
51. Mcleod *supra* note 25; Khan-Cullors *supra* note 45.
52. Brown & Schept, *supra* note 24.
53. *Id.* at 444.

these values, abolitionists ground their understanding of the carceral state in a distinct set of epistemological assumptions and ontological categories, ones which build from historically subjugated knowledges that are often omitted from the mainstream criminological canon.

IV. AI AS A DISCOURSE OF REFORM

In response to fundamental challenges made against the carceral state, law enforcement agents and policy makers have embraced a new discourse of reform, one which reframes the problems of mass incarceration and racialized state violence in terms of bias and administrative inefficiency. Reformers have called for expanded data collection as a pragmatic and politically neutral way to address these challenges.[54] The hope is that data-driven tools, often branded as AI, could render policing and courtroom practices fairer and more efficient by using data to create an objective view on current and future events. As state resources for new data platforms increase, even older tools, like actuarial risk assessment, are rebranding themselves as AI.

In spite of the shiny new brand, these proposals are consistent with a long lineage of technocratic reform efforts claiming to make the U.S. penal system more efficient, objective and fair. Liberal and conservative reform efforts throughout the twentieth century were grounded in a presumed benevolence of the carceral state as an institution that was struggling to meet its mandate of keeping communities safe, rather than as an institution for legitimizing racialized social exclusion and control.[55] To this end, bipartisan managerialist reforms are repeatedly embraced as a means of limiting bias and increasing the objectivity and efficiency of law enforcement practices.[56] As Murakawa argues, these reform efforts historically circumscribe the concept of racism in the penal system as the byproduct of flawed and individualized subjectivity. "Racism was an individual whim, an irrationality, and therefore racism could be corrected with 'state-building' in the Weberian sense—that is, the replacement of the personalized power of government officials with codified, standardized and formalized authority."[57]

These reforms are often couched in liberal or bipartisan rhetoric which remains committed to the state's language of "crime and punishment" even as it updates some of the key vocabulary to reflect changing moral intuitions over time.[58]

---

54. For example, in response to growing concerns over police brutality, FBI Director James Comey argued that, "the first step to understanding what is really going on is to gather more and better data related to those we arrest, those we confront, for breaking the law and jeopardizing public safety and those who confront us. Data seems a dry and boring word, but without it we cannot understand our world and make it better." James Comey, Address on Race and Law Enforcement at Georgetown University (Feb. 12, 2015) (transcript available at https://www.americanrhetoric.com/speeches/jamescomeygeorgetownraceandlaw.htm) [https://perma.cc/9TTP-E4EB].

55. MURAKAWA, *supra* note 8, at 79-82.

56. Goel et al., *supra* note 3; Jon Kleinberg et al., *Discrimination in the Age of Algorithms*, NAT'L BUREAU ECON. RSCH. 6, 37 (2019) [hereinafter *Discrimination in the Age of Algorithms*]; Cass R. Sunstein, *Algorithms: Correcting Biases*, 86 SOC. RSCH. 499 (2018).

57. MURAKAWA, *supra* note 8, at 11.

58. ALEXANDER, *supra* note 30, at 43; Alessandro Degiorgi, *Reform or Revolution: Thoughts on Liberal and Radical Criminologies*, 40 SOC. JUST. 131, 131–32 (2014).

As Wang notes, "'police science' is a way for police departments to rebrand themselves in the face of a crisis of legitimacy," pointing to internally generated data about arrests and incarcerations to justify their racially mediated practices.[59] During these times, governments create and financially support new research agendas, which frame issues in ways that protect the status of the penal system as a benevolent institution that simply needs new tools and training in order to meet its mandate of keeping communities safe.[60]

In this context, AI is a new brand of reform. AI often refers to the development of algorithms that are "trained" to recognize patterns and trends in large data sets. Over the last two decades, there has been a significant expansion in the interest in and availability of data within law enforcement agencies.[61] In an effort to gain access to large government data sets, technology companies have begun to develop tools that support and reproduce the operational logics of the carceral state. Private technology companies such as Securus, Amazon, and Palantir, have embraced law enforcement as their target customer, offering new analytics capacities that expand the state's crime control capabilities.

These collaborations are framed as "win-win" opportunities, wherein large tech companies gain a competitive edge in the race to develop state-of-the-art AI that is used in the service of creating a more efficient and expansive carceral apparatus. As a brand, the promise of AI is that it can equip court officials and law enforcement with superhuman capabilities—AI is a machine that can consume and impartially learn from the digital traces of human experience without ever growing tired,[62] an ideal solution to the problem of personal bias and administrative inefficiency.

These reforms are further bolstered by the enthusiastic participation of many civil society organizations and members of the academy. Philanthropic organizations from across the political spectrum have funded a number of high-profile initiatives aimed at reform through the adoption of "smart," "evidence based" and "data driven" practices.[63] In addition to spending millions of dollars to fund local pilot projects around the country, many of these initiatives hire teams of academic researchers to evaluate project sites using historical law enforcement data. These foundations have long served as key brokers of research partnerships between local governments and scholars, which tend to strengthen and institutionalize research

59. WANG, *supra* note 1, at 51.

60. MURAKAWA, *supra* note 8, at 81.

61. Brayne, *supra* note 38.

62. Christopher Rigano, *Using Artificial Intelligence to Address Criminal Justice Needs*, NAT'L. INST. JUST. (2018), https://www.nij.gov:443/journals/280/Pages/using-artificial-intelligence-to-address-criminal-justice-needs.aspx [https://perma.cc/MTE2-F5YS].

63. For example, private foundations such as the MacArthur Foundation, the Arnold Foundation and the Koch Foundation have all invested millions of dollars on criminal legal system reform initiatives which embrace this technocratic language of reform. Kristen Stoler, *Texas Billionaire John Arnold Gives $39 Million to Reform America's Broken Bail System*, FORBES (Mar. 19, 2019), https://www.forbes.com/sites/kristinstoller/2019/03/19/texas-billionaire-john-arnold-gives-39-million-to-reform-americas-broken-bail-system/#36c2cae31c13 [https://perma.cc/ZC3X-FDZU]; Molly Ball, *Do The Koch Brothers Really Care About Criminal-Justice Reform?*, ATLANTIC (Mar. 3, 2015), https://www.theatlantic.com/politics/archive/2015/03/do-the-koch-brothers-really-care-about-criminal-justice-reform/386615/ [https://perma.cc/9KM5-QYCF].

agendas that avoid structural criticisms in favor of research that supports increased efficiency and consistency of established systems.[64]

In addition, a new generation of mathematics and computer science scholars have assumed influential positions in debates regarding the use of algorithms and data analytics in criminal legal system reform.[65] These scholars tend to have minimal domain knowledge regarding the criminal legal system. They are often members of influential research organizations at prestigious universities,[66] which believe engineering and "data science" are the key skill sets needed to "drive social impact through technical innovation" in a wide range of high stakes policy domains, including the criminal legal system.[67]

Data collection and the study of "at risk" populations is often framed as a pragmatic approach to rehabilitation—if only we could better understand the populations most at risk, then we can more effectively address their problems in order to save them from themselves.[68] During these times, what Jock Young calls the "elective affinity" between positivist criminological work and the state is most clearly revealed through their orientation around state-generated data that is used in the study of crime and criminal populations.[69] As I will discuss in greater detail in the following section, these scholars uncritically accept problematic assumptions and theoretical framings from mainstream criminology, which results in the perpetuation of fundamentally flawed research under the guise of rigorous technical work.

V. THE PROMISES AND PERILS OF ALGORITHMIC REFORM

A. The promises: fairness, accuracy, and transparency.

In response to a growing number of studies measuring the disparate impact of criminal legal system practices on racial minorities,[70] a number of leaders from across

64. Platt, *supra* note 24.

65. *See, e.g.*, Sam Corbett-Davies & Sharad Goel, *The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning*, ARXIV (2018), https://arxiv.org/abs/1808.00023 [https://perma.cc/BT5J-EEWF]; Dwork et al., *supra* note 3; Kleinberg et al., *Algorithmic Fairness*, *supra* note 3; Zachary C. Lipton, Alexandra Chouldechova & Julian McAuley, *Does Mitigating ML's Disparate Impact Require Disparate Treatment?*, 1050 STATS. 19 (2017).

66. Examples include the Chicago Crime Lab, Stanford Computational Policy Lab, and the University of Southern California's Center for Artificial Intelligence in Society.

67. STANFORD COMPUTATIONAL POL'Y LAB, *Driving Social Impact Through Technical Innovation*, https://policylab.stanford.edu/ [https://perma.cc/6SDV-NJED] (last visited Aug. 1, 2020).

68. Dr. Muhammad argues that implicit in this framing is the assumption that the best way for us to understand the tensions between law enforcement and minority communities today is through closer scrutiny of policed populations. By studying "at risk" populations, we can better understand *why* law enforcement responds in the way they do and help them to more effectively intervene in the lives of risky individuals. This framing is centered squarely in a "politics of personal responsibility," which has a very long history in US carceral discourse – one that has long justified violence against and criminalization of racial minorities. Khalil Gibran Muhammad, *The Condemnation of Blackness - Khalil Gibran Muhammad Book Talk* 7:30 YOUTUBE (Aug. 3, 2015), https://www.youtube.com/watch?v=STKb-ai6874&t=392s [https://perma.cc/S39M-7HHD].

69. Brown & Schept, *supra* note 24, at 443.

70. *See, e.g.*, Will Dobbie, Jacob Goldin & Crystal S. Yang, *The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges*, 108 AM. ECON. REV. 201 (2018); WESTERN, *supra* note 30, at 15-18; Pierson et al., *supra* note 28.

the political spectrum have called for the adoption of scientific tools that could increase the accuracy and efficiency of criminal legal system operations by checking the implicit bias of officials. In this context, the hope is that regression and machine learning algorithms can be used to course correct the cognitive pitfalls and implicit biases of key decision makers in the system by presenting evidence-based claims about the likelihood of future events.

By framing the issue of disparate impact in this way, academics and government officials effectively circumscribe the issue of racial disparity in terms of individually held beliefs and preferences, rather than as the by-product of widespread organizational practices and cultural norms.[71] Technology firms have also embraced unconscious bias as a social challenge, which they can effectively overcome with the help of data-driven technology. As Hoffmann points out, implicit bias is understood as a phenomenon which "is somehow apart from us yet can infect our decision-making . . . as opposed to something that is variously, but systematically, cultivated and maintained."[72] In this uncritical framing of the problem, historical crime data are characterized as objective facts, a neutral "view from nowhere"[73] that stands in stark contrast to the flawed, fickle, and opaque subjectivity of human decision makers.

For example, this discourse is driving the rapid proliferation of pretrial risk assessments across the United States, where they are sold as a means of overriding judges' intuitive decision making processes, through which they may "erroneously, and unwittingly, introduce bias through acquired stereotypes"[74] or succumb to well-known cognitive pitfalls, such as "availability bias."[75] Proponents of pretrial risk assessment point to a growing literature in behavioral science to illustrate the common cognitive fallacies of legal decision makers in order to make the case for why actuarial tools could support more objective and accurate decisions.[76]

In this context, contentious social issues, such as massive increases in pretrial detention rates, are reframed as data processing challenges, in which key decision makers (judges, police officers, etc.) would benefit from tools that help them to distinguish "signal from noise" when making time-sensitive decisions about potentially dangerous individuals.[77] Risk assessment instruments purportedly hone in on the most predictive factors of an outcome of interest, helping to minimize the occurrence of "false positives and false negatives" in decisions over time. This is particularly

71. MURAKAWA, *supra* note 8; *see generally* Anna Lauren Hoffmann, *Where Fairness Fails: On Data, Algorithms, and the Limits of Antidiscrimination Discourse*, 22 INFO. COMM. & SOC'Y 900 (2019).

72. Hoffmann, *supra* note 71, at 905.

73. Donna Haraway, *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective*, 14 FEMINIST STUD. 575, 575-99 (1988).

74. Matthew DeMichele et al., *The Intuitive-Override Model: Nudging Judges Toward Pretrial Risk Assessment Instruments* 9 (2018).

75. Sunstein, *supra* note 56.

76. Goel et al., *supra* note 3, at 116; Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, *Blinking on the Bench: How Judges Decide Cases*, 93 CORNELL L. REV. 1 (2007); Kleinberg et al., *Discrimination in the Age of Algorithms*, *supra* note 56; Berk et al., *Fairness in Risk Assessments*, *supra* note 3.

77. DeMichele et al., *supra* note 74, at 9; Goel et al., *supra* note 3; Jon Kleinberg et al., *Human Decisions and Machine Predictions*, 133 Q.J. ECON. 237, 238 (2018) [hereinafter *Human Decisions and Machine Predictions*]; Sunstein, *supra* note 56.

important in contexts where risk management is framed in terms of high stakes, life-and-death situations where there is little room for error.[78] As a result, predictive accuracy is often held up as a key selling point of these tools. In fact, accuracy has become a fetishized measure of a tool's worth. In cases of life and death, it does not matter *why* a prediction is accurate, so long as it is.[79]

Skeptics of algorithmic tools in criminal law have also centered accuracy and bias in their criticisms. There are a growing number of researchers who investigate the ways that protected class attributes, such as race and gender, mediate the accuracy of outputs produced by algorithmic tools, including risk assessment, predictive policing, and facial recognition software.[80] A number of high-profile studies have argued that, not only are algorithmic tools in criminal law not very accurate, but the burden of that inaccuracy is disproportionately borne by historically marginalized groups, who are often subject to higher false positive rates.[81] This discrepancy in accuracy is usually talked about in terms of bias—critics argue that algorithmic tools run the risk of reproducing or amplifying pre-existing biases in the system.

These concerns have given rise to an influential community of researchers from both academia and industry who have formed a new regulatory science[82] under the rubric of "fair, accountable, and transparent algorithms" ("FAccT algorithms"). In this research community, criminal law applications have served as some of the most prominent thought exercises used to illustrate the trade-offs in the technical design and implementation of algorithmic tools. These efforts have coalesced around two general approaches to, what has been widely branded as, "algorithmic fairness:" 1) the development of formal fairness criteria that illustrate the trade-offs of different algorithmic interventions and 2) the development of managerialist "best practices" for maintaining a baseline of accuracy, transparency and validity in algorithmic systems. In the following sections, I outline these two approaches in greater detail before ultimately arguing that technocratic conceptions of bias and accuracy are inadequate

78. In spite of the fact that, on average, less than eight percent of pretrial defendants are arrested for a violent crime while awaiting trial, the fear of rape or murder is repeatedly mentioned in the academic literature, which presents the risk of being assaulted, raped or killed as an important issue to consider alongside the well documented harms of detention. These scholars repeatedly bring up these rare crimes as a point of contrast to the well documented harms of pretrial incarceration. This trope is also invoked in one-on-one interviews I've had with judges, as well as in the mainstream press when covering the issue of pretrial release. DeMichele et al., *supra* note 74, at 15; Goel et al., *supra* note 3; Sunstein, *supra* note 56.

79. For example, in a 2013 talk, a prominent statistician and criminologist argued, "I'm not trying to explain criminal behavior, I'm trying to forecast it. If shoe size or sunspots predicts that a person's gonna commit a homicide I want to use that information, even if I have no idea why it works." Richard Berk, *Forecasting Criminal Behavior and Crime Victimization* 9:40, YOUTUBE (Feb. 12, 2013), https://www.youtube.com/watch?v=rolFHPegLVQ&t=105s [https://perma.cc/2UV5-R59F].

80. Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/238X-8GKA]; Buolamwini & Gebru, *supra* note 2; Kristian Lum & James Johndrow, *A Statistical Framework for Fair Predictive Algorithms*, ARXIV (Oct. 25, 2016), http://arxiv.org/abs/1610.08077 [https://perma.cc/2MSE-RTQ4].

81. Angwin et al., *supra* note 80; Buolamwini & Gebru, *supra* note 2; Jacob Snow, *Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots*, AMERICAN CIVIL LIBERTIES UNION (Jul. 26, 2018), https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28 [https://perma.cc/5L4K-L2SC].

82. SHEILA JASANOFF, THE FIFTH BRANCH: SCIENCE ADVISERS AS POLICYMAKERS 77 (2d ed. 1994).

conceptual anchors for this discussion, because they fail to interrogate the deeper normative, theoretical, and methodological premises of predictive systems.

B. Formalized fairness criteria

The initial focus of the FAccT algorithm community has been to map mathematical formalisms onto complex legal concepts such as discrimination, disparate impact, equal opportunity, and affirmative action.[83] In doing so, researchers claim that the goal is to create a foundation for more robust debates about the social desirability of algorithmic tools by providing "conceptual precision" regarding a tool's fairness.[84]

A number of researchers have pointed out that some fairness criteria are mutually incompatible,[85] which has given rise to deeper questions about what criteria of fairness should be met and across what groups.[86] These limitations have been framed in terms of trade-offs which must be debated and resolved on a case-by-case basis. Formalizing these trade-offs is widely considered to be a pragmatic approach to criminal legal system reform. For example, a prominent criminologist recently co-wrote a paper with a number of computer scientists, arguing that "one cannot expect any . . . tool to reverse centuries of racial injustice or gender inequality" in the criminal legal system.[87] Instead, they argued, the job of technical researchers is to delineate the trade-offs of different decisions in quantifiable terms, so that they can be transparently adjusted to reflect the preferences and values of different communities. The hope is that formalized definitions of fairness will increase the transparency of various trade-offs and bolster a more inclusive public discourse regarding the social desirability of these tools.[88]

In the context of criminal law, "equity," "fairness," and "accuracy" are defined as mathematical formalisms that are pitted against one another as competing values that can never be fully resolved unless there is a fundamental change in "base rates" of criminal activity across groups.[89] These researchers argue that differences in false positive rates across racial populations have more to do with real differences in the prevalence of criminal activity across those groups than with bias in the algorithm.[90] In these arguments, "unequal base rates" of criminal activity are uncritically characterized as an endemic issue across protected groups, rather than as a by-product of

83. Kleinberg et al., *Algorithmic Fairness*, *supra* note 3; Barocas & Selbst, *supra* note 2; Moritz Hardt, Eric Price & Nati Srebro, *Equality of Opportunity in Supervised Learning*, 29 ADVANCES NEURAL INFO. PROCESSING SYSTEMS 3315, 3315–23 (2016); Dwork et al., *supra* note 3.

84. Berk et al., *Fairness in Risk Assessments*, *supra* note 3; BERK, MACHINE LEARNING RISK ASSESSMENTS, *supra* note 3, at 57.

85. Berk et al., *Fairness in Risk Assessments*, *supra* note 3, at 116; Chouldechova, *supra* note 2.

86. For example, it is theoretically impossible to design a classifier that simultaneously satisfies false positive parity and "predictive value parity," or equal calibration across protected classes. Chouldechova, *supra* note 2. Berk et al., *Fairness in Risk Assessments*, *supra* note 3, at 19.

87. Berk et al., *Fairness in Risk Assessments*, *supra* note 3, at 35.

88. *Id.* at 30.

89. BERK, MACHINE LEARNING RISK ASSESSMENTS, *supra* note 3, at 123; Chouldechova, *supra* note 2; Corbett-Davies & Goel, *supra* note 65, at 12.

90. BERK, MACHINE LEARNING RISK ASSESSMENTS, *supra* note 3, at 123; Corbett-Davies & Goel, *supra* note 65, at 12.

discriminatory policing and courtroom practices.[91] As a result, a number of scholars have warned that attempts to balance false positive and false negative rates in tools like risk assessments could result in higher rates of victimization in communities of color, because it would result in less accurate risk classifications. In these arguments, the risk of violent crime, such as murder, is frequently invoked. As Berk argues,

> by far the leading cause of death among young African-American males is homicide. The most likely perpetrators of those homicides are other young African-American males. There are legitimate concerns about fair risk assessments for accused perpetrators, but no such concerns about the consequences of fair risk assessments for their possible victims. Is that fair?[92]

Berk's assertion is based on a false binary distinction between "victims" and "perpetrators" of violent crime, in spite of a growing body of literature which has established a significant overlap across victim and offender populations—yesterday's victim is likely to become tomorrow's perpetrator, and vice versa.[93] Moreover, Berk uses statistics about the violent death of African Americans as a justification for racial profiling and pre-emptive detention in African-American communities.

Muhammad calls this rhetorical strategy the "violence card," whereby proponents of race-based profiling present statistics that, on their face, seem to speak for themselves. According to these people, Muhammad argues, "by knowing that this is what Black people do to each other, we need not have further conversation about any responsibility that lies outside the Black community" for their treatment. This tactic is based on a very long tradition of respectability politics in U.S. carceral discourse. Harsh, preemptive policing practices, such as stop and frisk, are rationalized as a means of protecting communities from themselves.[94] Berk uses this line of thinking to posit that "accurate" risk assessments serve the best interests of overpoliced communities, even if it means subjecting them to higher rates of false positive misidentification.

C. Managerialist "best practices"

In light of the tensions between accuracy and other definitions of fairness, many researchers in the FAccT community have made a "don't let the perfect be the enemy of the good" appeal, arguing that a tool's social value is best understood in comparative terms—do algorithmic decision-making aids produce more accurate predictions

---

91. NAT'L ACADEMY OF SCIENCES, *Using ML in Criminal Justice Risk Assessments - The Frontiers of Machine Learning*, YOUTUBE (Feb. 14, 2017), https://www.youtube.com/watch?v=gdEPPRhNu34 [https://perma.cc/8H4U-JJQA]; Anthony W. Flores et al., *False Positives, False Negatives, and False Analyses: A Rejoinder to Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased against Blacks*, 80 FED. PROBATION 38, 41 (2016).

92. BERK, MACHINE LEARNING RISK ASSESSMENTS, *supra* note 3, at 125.

93. Wesley G. Jennings et al., *On the Overlap Between Victimization and Offending: A Review of the Literature*, 17 AGGRESSIVE & VIOLENT BEHAVIOR 16, 16–26 (2012).

94. What these statistics eschew is the fact that the majority of individuals who are arrested as a result of policies like "stop and frisk," are arrested for non-violent, petty offences, such as riding a bicycle on the sidewalk, public intoxication, loitering, etc. These offences have nothing to do with increasing the safety of African American communities. Muhammad, *supra* note 68, at 12:30.

than a human decision maker would make on their own?[95] These researchers characterize criticisms of algorithmic bias as impractical and perfectionist, arguing that current statistical practices, while not perfectly accurate, provide a pragmatic means of improving the overall accuracy and transparency of high stakes decisions in criminal law.[96]

In order to bolster these claims, these scholars point to literature from the behavioral sciences in order to argue that, by and large, algorithms outperform human decision makers in accurately predicting outcomes like recidivism.[97] Others try to empirically test this "human versus machine" formulation with historical crime data. However, these studies have proven challenging to do in most criminal legal system scenarios, because counterfactual data are not available for measuring the comparative accuracy of different decisions (i.e. we do not know if an incarcerated person would have gone on to commit another crime had they been released). This challenge has not stopped researchers from taking elaborate measures to impute missing data in order to make bold claims about whether or not an algorithmic prediction is more accurate than a human forecast.[98] These researchers posit that algorithms have the potential to serve as a force of equity, if only appropriate safeguards could be put into place to minimize overall bias and maximize accuracy of their predictions.

To this end, some have sought to address issues of bias and accuracy in criminal legal system algorithms by reformulating them in terms of narrower technical issues such as "sample bias," which can be addressed by regularly re-validating predictive models with data from local jurisdictions.[99] Scholars have pointed out that such practices are crucial for understanding the impact of changing conditions and specific policy interventions over time.[100] Thus, there is a growing body of literature within both academia and industry which aims to outline such standards and best practices for the ethical implementation of predictive algorithms.[101] Generally speaking, these frameworks aim to minimize specific types of bias (sample bias, label bias, etc.) procedurally, through semi-regular validations of predictive models, in order to maximize their purported accuracy and minimize well-established forms of statistical bias. While these procedures are an important first step towards addressing a specific subset of issues regarding a tool's validity and generalizability, they are insufficient for

95. BERK, MACHINE LEARNING RISK ASSESSMENTS, *supra* note 3, at 165; Goel et al., *supra* note 3, at 2; Sunstein, *supra* note 56, at 504; Jared Sylvester & Edward Raff, *What About Applied Fairness?*, ARXIV (2018), https://arxiv.org/abs/1806.05250 [https://perma.cc/VBR6-FWYY].

96. BERK, MACHINE LEARNING RISK ASSESSMENTS, *supra* note 3, at 116.

97. DeMichele et al., *supra* note 74, at 9; Goel et al., *supra* note 3, at 2; Sunstein, *supra* note 56, at 502.

98. Jon Kleinberg et al., *Human Decisions and Machine Predictions*, *supra* note 77, at 6.

99. Goel et al., *supra* note 3, at 7; Kleinberg et al., *Discrimination in the Age of Algorithms*, *supra* note 56, at 19.

100. John L. Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1793 (2018); Megan Stevenson, *Assessing Risk Assessment in Action*, 103 MINN. L. REV. 303, 375 (2018).

101. *Arnold Ventures Statement of Principles on Pretrial Justice*, ARNOLD FOUNDATION, https://www.arnoldventures.org/work/pretrial-justice/ [https://perma.cc/R7WS-BZR7] (last visited Aug. 1, 2020) [hereinafter *Statement of Principles*]; Christopher Bavitz et al., *Assessing the Assessments: Lessons from Early State Experiences In the Procurement and Implementation of Risk Assessment Tools*, BERKMAN KLEIN CENT. RSCH. PUBL. (2018); BERK, MACHINE LEARNING RISK ASSESSMENTS, *supra* note 3, at 155-61.

addressing deeper issues regarding the way claims are constructed based on the available data. In the following section, I argue bias and accuracy are inadequate conceptual anchors for discussing the social implications of these tools, since they fail to interrogate the deeper theoretical and methodological premises of these data-intensive, algorithmically mediated systems.

D. The perils: misattribution of agency and the conflation of arrest with danger.

All of the above arguments regarding accuracy and objectivity are built on a shared epistemological assumption that arrest, conviction, and incarceration data reflect individual and population-level criminal activity, rather than law enforcement activity.[102] The deeper historical disparities in how the police and court officials treat different groups and pursue various types of crime are not considered.[103] Numerous researchers have pointed out the fundamental measurement errors that occur when people uncritically characterize criminal legal system data solely in terms of an individual's proclivity toward crime.[104] Scholars have long argued that crime statistics are partial and biased, and their incompleteness is delineated clearly along power lines.[105] Arrest statistics are best understood as measurements of law enforcement practices. These practices tend to focus on "street crimes" carried out in low income communities of color while neglecting other illegal activities that are carried out in more affluent and white contexts.[106] Similarly, conviction and incarceration data primarily reflect the decision-making habits of relevant actors, such as judges, prosecutors, and probation officers, rather than a defendant's criminal proclivities or guilt.[107]

In light of these criticisms, scholars should systematically recharacterize arrest, conviction, and incarceration data, as data that can inform important conversations regarding the disparate impact of specific policing and courtroom practices, rather than individual proclivities toward crime. Such a recharacterization fundamentally shifts the attribution of agency and responsibility, away from the "antisocial behavior" of "risky individuals" and towards a carceral system that surveils, arrests, prosecutes and incarcerates people in disparate ways. Arrest, conviction, and incarceration data are not accurate measures of crime, and arrest is not synonymous with danger or potential harm to the community.

Yet mainstream characterizations of police and court data continue to fuel deeply problematic conflations between arrest and dangerousness.[108] For example, Kleinberg

102. ELLIOTT, *supra* note 5, at 1.

103. *Id.*

104. *Id.*; Dolovich, *supra* note 15, at 265; David A. Harris, *The Reality of Racial Disparity in Criminal Justice: The Significance of Data Collection*, 66 L. & CONTEMPORARY PROBLEMS 71, 79 (2003); Prins et al., *supra* note 18, at 4.

105. Platt, *supra* note 8; Brown & Schept, *supra* note 24.

106. Laura Nader, *Crime as a Category—Domestic and Globalized*, *in* CRIME'S POWER 55–76 (Philip C. Parnell & Stephanie C. Kane, eds., 2003); Platt, *supra* note 8.

107. Harris, *supra* note 104.

108. *See* Lauryn P. Gouldin, *Distangling Flight Risk from Dangerousness*, 2016 B.Y.U. L. REV. 837, 852 (2016); Paula Maurutto & Kelly Hannah-Moffat, *Assembling Risk and the Restructuring of Penal Control*, 46 BRITISH J. CRIMINOLOGY 438, 438–54 (2006).

et al. conflate arrest statistics, even for minor technical violations, with criminal activity in order to impute data about the probability of a defendant recidivating while awaiting trial.[109] These calculations are used to bolster an argument regarding whether or not pretrial risk assessments are more accurate than judges at predicting future crime. What this conversation eschews is that pretrial detention is not constitutionally permissible, unless the judge finds clear and convincing evidence that the defendant poses a significant risk of flight or danger to the community. As Robinson and Koepke point out, the concept of "dangerousness" is ill defined in the courts, leading to a massive expansion in the use of pretrial detention,[110] despite the persistently low incidence of re-arrest for violent crime.[111]

Given the extremely low rates of pretrial arrest for violent crime, developing actuarial risk assessments that meaningfully differentiate the pretrial population is quite challenging. Risk assessments cannot identify people who are more likely than not to commit a violent crime. In fact, the vast majority of individuals, even those classified with the highest levels of risk, will not be arrested for a violent crime while awaiting trial.[112] Consider the dataset used to build the Public Safety Assessment (PSA): more than 92% of the people who were flagged for pretrial violence did not get arrested for a violent crime.[113] Yet this reality is masked by descriptions in PSA materials and presentations that characterize high-risk defendants as "three times more likely to be arrested for a violent crime."[114] While technically true, such a framing masks the disturbing fact that the average difference in arrest rates between high and low risk defendants on tools like the PSA is less than six percentage points—a paltry 7.2% of the high risk group are arrested, in contrast to 2.4% of the low risk group.[115]

If these tools were calibrated as accurately as possible, then they would predict that every person was unlikely to commit a violent crime while on pretrial release. Instead, risk assessments sacrifice accuracy and generate substantially more false positives (people who are flagged for violence but do not go on to commit a violent

109. Kleinberg et al., *Human Decisions and Machine Predictions*, *supra* note 77, at 3.

110. Koepke & Robinson, *supra* note 100, at 1742.

111. In many jurisdictions, such as Kentucky, Washington DC, and Cook County, the rate of arrest for violent crime during pretrial has been reported to be as low as less than one or two percent. For tools which specifically aim to measure re-arrest for a violent offense, such as the PSA and COMPAS, the vast majority of defendants (about 92%) are predicted to not be arrested for a violent offence while awaiting trial. Sandra G. Mayson, *Dangerous defendants*, 127 YALE L.J. 490, 514 (2017); *see also* ILLINOIS CIRCUIT COURT OF COOK COUNTY, *Model Bond Court Initiative* (2018), http://www.cookcountycourt.org/HOME/ModelBond CourtInitiative.aspx [https://perma.cc/H3ZL-6QBT] (last visited Aug. 1, 2020); Stevenson, *supra* note 100; COURT SERVS. & OFFENDER SUPERVISION AGENCY D.C., FY 2016 AGENCY FINANCIAL REPORT 27 (Nov. 15, 2016), https://www.psa.gov/sites/default/files/FY2016%20CSOSA%20AFR%20FINAL%2011-15-2016.pdf [https://perma.cc/G9DR-ZGPK].

112. ARNOLD FOUNDATION, Presentation, PSA RESULTS: FOR REFERENCE WHEN CREATING A RELEASE CONDITIONS MATRIX slide 3 (2019), https://advancingpretrial.org/implementation/guides/ [https://perma.cc/V374-NF66] [hereinafter Presentation]; Sarah Lustbader, *Risk Assessment Tools Are Flawed—Should We Throw Them Away?*, APPEAL (July 25, 2019), https://theappeal.org/risk-assessment-tools-are-flawed-should-we-throw-them-away/ [https://perma.cc/9ZX5-D24W].

113. ARNOLD FOUNDATION, Presentation, *supra* note 112, at 3.

114. Billie Grobe, Justice System Partners, Plenary Speech at the annual conference of the National Association of Pretrial Services Agencies (Sept. 11, 2017) (field notes on file with author).

115. ARNOLD FOUNDATION, Presentation, *supra* note 112, at 3.

crime) than true positives (people who are flagged for violence and do go on to be arrested for a violent crime.[116] Consequently, violence risk assessments could easily lead judges to overestimate the risk of pretrial violence and detain more people than is justified.[117]

Moreover, current risk assessment instruments are unable to distinguish one person's risk of violence from another's. In statistics, predictions are made within a range of likelihood, rather than as a single point estimate. For example, a predictive algorithm might confidently estimate a person's risk of arrest as somewhere between a range of five and fifteen percent. Studies demonstrate that predictive models can only make reliable predictions about a person's risk of violence within very large ranges of likelihood, such as twenty to sixty percent.[118] As a result, virtually everyone's range of likelihood overlaps. When everyone is similar, it becomes impossible to differentiate people with relatively low or high risks of violence. At present, there is no statistical remedy to this problem.

In light of these challenges, a number of pretrial risk assessments provide an additional risk score for "new criminal activity," which estimates the likelihood that a defendant will be re-arrested for *any* offence while awaiting trial.[119] This general recidivism score is sometimes provided as a supplementary point of consideration to decision makers, in spite of the fact that the Supreme Court and state high courts have not recognized the likelihood of non-violent re-arrest as a constitutionally permissible reason for detaining someone prior to their trial.[120] Beyond flight risk, the Supreme Court to date has only approved pretrial detention when someone is accused of "a serious crime [and] presents a demonstrable danger to the community."[121]

In addition, a number of tools actively conflate the likelihood of arrest for any infraction with dangerousness. For example, the Colorado Pretrial Assessment Tool (CPAT) defines a risk to "public safety" as any "new criminal filing," including for

116. Angwin et al., *supra* note 80, at 11.

117. For example, a recent study found that people significantly overestimate the recidivism rate for individuals who are labeled as "moderate-high" or "high" risk on a risk assessment. Participants greatly overestimated the true recidivism rate for those assessed as moderate-high risk category – the true rate was less than fifty percent of what participants predicted. *See* Daniel A. Krauss, Gabriel I. Cook & Lukas Klapatch, *Risk Assessment Communication Difficulties: An Empirical Examination of the Effects of Categorical Versus Probabilistic Risk Communication in Sexually Violent Predator Decisions*, 36 BEHAVIORAL SCI. & L. 532, 532-53 (2018).

118. Stephen D. Hart & David J. Cooke, *Another Look at the (Im-)precision of Individual Risk Estimates Made Using Actuarial Risk Assessment Instruments*, 31 BEHAVIORAL SCI. & L. 81, 92-93 (2013).

119. For example, pretrial risk assessments from Florida, Virginia, Ohio, Indiana and the Federal courts all purport to evaluate the public safety risks. Yet, they define their outcome for public safety as arrest for any new crime, not just violent offenses. Mayson, *supra* note 111, at 124.

120. Ensuring a person's appearance in court has historically been the driver behind judges' decisions regarding bail and pretrial detention. *See* Stack v. Boyle, 342 U.S. 1, 5 n.3 (1951) ("If the defendant is admitted to bail, the amount thereof shall be such as in the judgment of the commissioner or court or judge or justice will insure [sic] the presence of the defendant . . . ."). Before the Supreme Court's 1987 decision in *United States v. Salerno*, flight risk was the only legitimate (i.e., constitutional) basis for detaining a defendant before trial. 481 U.S. 739 (1987).

121. United States v. Salerno, 481 U.S. 739, 750 (1987).

traffic stops and municipal offenses.[122] Still other tools, such as the Nevada Pretrial Risk Assessment, merge flight and dangerousness into one aggregate risk score, which poses an additional set of challenges.[123] Developers of these assessments often define their outcomes in terms of general recidivism to produce stronger associations between the inputs and outcome variables in their statistical models. This results in a significant expansion in the number of defendants who are rated as "moderate" or "high" risk for pretrial failure and fuels a widespread conflation of re-arrest for any infraction with dangerousness.

The conflation of generalized risk of arrest with dangerousness has serious implications for how police and judges interact with individuals caught in the criminal legal system. In the case of pretrial risk assessment, it can lead to unwarranted detention of people who have not been convicted of a crime. This has serious ripple effects in terms of housing and employment instability, the disruption of social support structures, and the increased likelihood of conviction.[124] Proponents of risk assessment often make perfunctory acknowledgements of these issues regarding pretrial detention, but then minimize these impacts against more weighty concerns of community safety, citing the potential for murder, rape, and assault if the person were released.[125]

Ironically, these invocations of violence are likely to fuel the very logical fallacies that these scholars purport to mitigate through algorithms. By placing a widespread practice like pretrial detention beside the very rare occurrence of violent crime, scholars like Sunstein fuel an "availability bias."[126] The perception of violent crime is heightened due to the frequency with which it is invoked in mainstream and academic discourse regarding pretrial release.[127] This heightened sense of danger is further fueled by widespread beliefs held by practitioners within the criminal legal system justice practitioners, who frequently associate the number of prior arrests with an individual's proclivity towards violence.[128] Rather than challenging this assumption, academic articles, industry white papers, and official government documents tend to reinforce this conflation by using general re-arrest data as a proxy for

122. TIMOTHY SCHNACKE, "MODEL" BAIL LAWS: RE-DRAWING THE LINE BETWEEN PRETRIAL RELEASE AND DETENTION, CTR. LEGAL & EVIDENCE-BASED PRACTICES 109, 202 (Apr. 18, 2017), https://university.pretrial.org/viewdocument/model-bail-laws-re-drawing-the-l [https://perma.cc/7E9V-GTV3].

123. Scholars have warned that combining flight and dangerousness into one score can lead to an overestimation of both types of risk and make it challenging to identify effective risk mitigating interventions. Gouldin, *supra* note 108, at 844.

124. *See generally* Dobbie, Goldin & Yang, *supra* note 70; Arpit Gupta et al., *The Heavy Costs of High Bail: Evidence from Judge Randomization*, 45 J. LEG. STUD. 471, 471-505 (2016); Paul Heaton, Sandra Mayson et al., *The Downstream Consequences of Misdemeanor Pretrial Detention*, 69 STAN. L. REV. 711 (2017).

125. BERK, MACHINE LEARNING RISK ASSESSMENTS, *supra* note 3, at 34-35; Sunstein, *supra* note 56.

126. Sunstein, *supra* note 56.

127. For example, Sunstein argues that "If defendants are incarcerated, the long-term consequences can be very severe. Their lives can be ruined. But if defendants are released, they might flee the jurisdiction or commit crimes. People might be assaulted, raped, or killed." This kind of side by side comparison of detention versus murder makes the prospect of unwarranted detention seem rather minor in comparison to the risk of lost life. Sunstein, *supra* note 56, at 500.

128. This theme has come up repeatedly in interviews I have conducted with judges and other pretrial officials within the criminal legal system, who frequently conflated prior arrest history with public safety risk.

danger, or by providing separate scores for "new criminal activity" alongside more modest estimates of violent re-arrest.[129]

This "fundamental misattribution of agency"[130] and conflation of re-arrest with danger renders moot conversations regarding the accuracy of pretrial risk assessment forecasts, because the data used to measure accuracy is simply not representative of the outcome of interest. Some researchers recognize the limits of the available data, but insist on making claims of crime prediction by framing the problem as a question of "sample bias" or "label bias." To make these tools more accurate and valid, these authors provide some quick technical fixes for addressing these narrow conceptualizations of bias.[131] Such efforts only reinforce the false association between arrest history and dangerousness and further conceal the fundamental misattribution of agency.

This insistence on misleading framings of police and court data is fueled by the crucial rhetorical role they play in justifying punitive decisions. More representative framings of the data would produce less powerful claims, or they would give rise to research questions that directly challenge the practices and logics of the carceral state. The political economy of algorithmic systems rests largely on the fundamental misattribution of agency to make authoritative claims about an individual's criminal proclivities, which fuel and legitimize decisions to punish.[132] In this way, predictive algorithms that are based on these widespread mischaracterizations of the data underpin the moral economy that justifies the exclusion and repression of marginalized populations through the construction of "risky" or "deviant" profiles.[133]

Critical criminologists have long argued that these interpretations of crime data are performative enactments of power structures, ones which fundamentally shape the discourse, methods, and epistemological assumptions of criminology and law enforcement practices.[134] Yet, positivist subfields of the discipline continue to build predictive models to forecast "dangerousness," "new criminal activity," and "recidivism" based on this data. This issue has been a recurring tension within the field of criminology since the turn of the nineteenth century, when the meaning of arrest and incarceration statistics from the 1890 census were debated by early scholars of crime.[135] In the 1920's and 30's, actuarial methods of forecasting criminal behavior relied heavily on incorrect framings of arrest, conviction, and incarceration in order to make fallacious claims about crime prediction.[136] In the wake of the civil rights

129. DeMichele et al., *supra* note 74, at 26; Kleinberg et al., *Human Decisions and Machine Predictions*, *supra* note 77, at 237; SCHNACKE, *supra* note 122.

130. This phrase was first introduced to me by my colleague Rodrigo Ochigame at MIT and is his term for describing the misattribution of criminal legal system data to the behaviors and pathologies of individuals who are being prosecuted by the system.

131. Goel et al., *supra* note 3.

132. Harris, *supra* note 104; MUHAMMAD, *supra* note 4.

133. Dolovich, *supra* note 15; Harris, *supra* note 104; Michael J. Lynch, *The Power of Oppression: Understanding the History of Criminology as a Science of Oppression*, 9 CRITICAL CRIMINOLOGY 144, 144–52 (2000).

134. Brown & Schept, *supra* note 24; Platt, *supra* note 8.

135. MUHAMMAD, *supra* note 4.

136. BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 88 (2008).

movement of the 1960's, critical criminologists argued that drastic increases in official crime statistics were more a by-product of administrative changes in how crimes were reported than a result of real spikes in crime.[137] They pointed to alternative sources of crime data in order to resist the conflation of racial unrest with criminality in the late 1960s.[138]

More recently, David Harris cites numerous examples in which law enforcement officials have used arrest and incarceration statistics to justify racial profiling in the 1990's.[139] Officials pointed to statistics that reflect the overrepresentation of African Americans and Latinx in jails in order to justify racial profiling. They argued that their officers stopped and searched a disproportionate number of minorities, not because of racial animus, but because, quite simply the data showed that "that's where the criminals are."[140] These officials used arrest and incarceration data as a substitute for crime rate and, in doing so, laid the foundation for the state's own recursive logic, whereby it used internally generated numbers about arrest and incarceration as a justification for continuing the very practices that fueled those numbers.[141]

Discourse regarding "fair, accountable, and transparent" AI is the most recent incarnation of this historical struggle over the interpretation of legal system data. To date, the lion's share of research in this area uncritically embraces the epistemological assumptions of mainstream criminology. In doing so, they continue a long tradition of centering reforms in the "sciences of oppression" which seek to profile and surveil marginalized communities.[142] This scholarship not only provides a mechanism for the confinement and control of the "dangerous classes," but also creates the very processes through which these populations are turned into deviants to be controlled and feared.

Attempts to render these tools more accurate by addressing narrow notions of "bias" simply miss the deeper methodological and epistemological issues regarding the fairness of these tools. As Hoffmann argues, we must grapple with the ways data-intensive, algorithmically mediated systems reinforce certain discursive frames over

137. Vesla M. Weaver, *Frontlash: Race and the Development of Punitive Crime Policy*, 21 STUD. AM. POLIT. DEV. 230, 245-46 (2007).

138. Platt, *supra* note 8.

139. Harris, *supra* note 104.

140. What these officials conveniently overlooked were data which revealed that the "false positive" rate was much higher for African-American males, meaning that the number of times that they searched that population and found nothing was much higher than other racial groups. *Id.* at 79.; Pierson et al., *supra* note 28.

141. A recent study documenting the perspectives of line prosecutors in four different jurisdictions across the United States found similar justifications to questions regarding racial disparities in arrest and conviction. While many prosecutors expressed an understanding of how disparate levels of policing could lead to different levels of involvement with the criminal legal system, they also emphasized inherent "racial differences in criminal behavior," and expressed doubt that prosecutors could do anything to reduce such disparities. BESIKI LUKA KUTATELADZE ET AL., PROSECUTORIAL ATTITUDES, PERSPECTIVES, AND PRIORITIES: INSIGHTS FROM THE INSIDE (2018), http://www.safetyandjusticechallenge.org/wp-content/uploads/2018/12/FIU-Loyola-_MacAthruth-Prosecution-Project-Report-One-PDF.pdf [https://perma.cc/XA6X-R5G6].

142. Lynch, *supra* note 133. For example, Arnold Ventures purports with their National partnership for Pretrial Justice to "promote racial justice" while investing in pre-trial practices that will continue to disproportionately affect marginalized communities. ARNOLD FOUNDATION, Statement of Principles, supra note 101.

others—only then can we begin to unpack the ways such systems shape and constrain our ability to collectively pursue particular visions of justice.[143] Efforts to increase the accuracy of predictive systems run the risk of circumscribing these deeper ideological and epistemological struggles within a narrow technocratic debate about how to make these tools more valid, accurate, and fair.

VI. THE WAY FORWARD: EMBRACING AN ABOLITIONIST WORLDVIEW

In the current political moment, the conversation regarding "FAccT" algorithms has proven highly influential in shaping state and federal legislative efforts for reform. In the case of pretrial risk assessment, a number of states have passed legislation that acknowledges the risk of bias in the risk assessment tools. They call for the establishment of oversight committees and standards to ensure that specific types of bias are minimized, through semi-regular validation using updated data from local jurisdictions.[144] These efforts are very much aligned with the narrow formulation of "bias" embraced by many in the FAccT community, eschewing deeper concerns regarding the epistemological soundness of these tools.

In this context, community advocates have to strike a balance between calling for wholesale moratoriums on the use of some technologies in criminal law, while also trying to mitigate harm in places where those tools have already been adopted. For example, in a public letter regarding the use of risk assessment, a coalition of civil rights organizations emphasized that the connection between risk assessment and decarceral policies was tenuous at best.[145] They went on to make a fundamental critique regarding the use of arrest data to measure individual risk, arguing that "decades of research have shown that such data primarily document the behavior and decisions of police officers and prosecutors, rather than the individuals or groups that the data are claiming to describe."[146]

In addition to this fundamental critique regarding the epistemological validity of risk assessments, the authors spend subsequent pages outlining guidelines for minimizing the harm of risk assessments in places where they are already adopted.[147] These guidelines include recommendations for increased transparency and third-party validation, as well as more pointed suggestions for how to tailor risk assessments to meet de-carceral ends.[148] These recommendations have less to do with the technical design of the risk assessments and more to do with implementing procedural safeguards to ensure a presumption of innocence is maintained in the courts.

---

143. Hoffmann, *supra* note 71.

144. SARAH DESMARAIS & EVAN LOWDER, PRETRIAL RISK ASSESSMENT TOOLS: A PRIMER FOR JUDGES, PROSECUTORS, AND DEFENSE ATTORNEYS, SAFETY & JUST. CHALLENGE (2019), http://www.safetyandjusticechallenge.org/wp-content/uploads/2019/02/Pretrial-Risk-Assessment-Primer-February-2019.pdf [https://perma.cc/6AX2-V65E].

145. LEADERSHIP CONFERENCE ON CIVIL AND HUMAN RIGHTS, THE USE OF PRETRIAL "RISK ASSESSMENT" INSTRUMENTS: A SHARED STATEMENT OF CIVIL RIGHTS CONCERNS (2018), http://civilrightsdocs.info/pdf/criminal-justice/Pretrial-Risk-Assessment-Full.pdf [https://perma.cc/35Q8-NM93].

146. *Id.* at 1.

147. *Id.* at 2-9.

148. *Id.*

Other community organizations call for a shift away from assessing "risk" to evaluating "need" in the hopes of changing the logic that underlies actuarial assessments, towards more rehabilitative ends.[149] However, this strategy should be pursued with caution. Attempts to use risk/needs assessments for rehabilitation can easily slip into more punitive practices of profiling and criminalization.[150] In the case of pretrial assessment, some organizations are pushing back against the framing of pretrial interventions as a service. For example, Chicago community advocates argue "punishment is not a service" by documenting the various ways pretrial interventions, such as electronic monitoring and mandatory curfews, disrupt the livelihoods and home life of defendants awaiting trial.[151] This work illustrates the ways that progressive framings of data science for rehabilitative intervention run the risk of fueling an expansion of the carceral state, by focusing on the measurement of defendants' pathologies and deficiencies, rather than reframing analysis in terms of the disparate violence of the carceral state.

Yet, a growing number of thinkers and advocates are resisting AI applications on more fundamental terms. For example, abolitionist and scholar Nabil Hassein problematizes efforts to increase the representation of dark-skinned individuals in the training data for facial recognition software,[152] arguing that efforts to render such software more accurate through the inclusion of underrepresented faces would do more harm than good. As Hassein argues,

> The reality for the foreseeable future is that the people who control and deploy facial recognition technology at any consequential scale will predominantly be our oppressors. Why should we desire our faces to be legible for efficient automated processing by systems of their design? . . . The struggle for liberation is not a struggle for diversity and inclusion — it is a struggle for decolonization, reparations, and self-determination.[153]

In contrast to mainstream efforts to render algorithms more accurate, Hassein positions the limits of AI within a structural critique of the carceral state as a

149. SOUTHERNERS ON NEW GROUND, *Durham Judges Refuse to Eradicate the Use of Money Bail Or Include Needs Assessment Model in New Policy* (2019), http://southernersonnewground.org/2019/03/breaking-durham-judges-refuse-eradicate-use-money-bail-include-needs-assessment-model-new-policy/ [https://perma.cc/DR7W-YPPB].

150. Professor Ferguson illustrates this in the case of using predictive algorithms to create a "strategic subjects" list that identified individuals who are at risk of being involved in gun violence. The project was initially framed as a community health intervention, whereby at-risk subjects would be targeted for support services to minimize the incidence of gun violence. However, the algorithm quickly evolved into a tool used for targeting and arresting individuals who were considered "suspects" in incidence of gun violence. In instances where these individuals were charged with a specific crime, they were punished more severely than people who were not on such a list. Ferguson, *supra* note 11, at 1109.

151. CHICAGO COMMUNITY BOND FUND, PUNISHMENT IS NOT A "SERVICE:" THE INJUSTICE OF PRETRIAL CONDITIONS IN COOK COUNTY (2017), https://chicagobond.org/wp-content/uploads/2018/10/pretrialreport.pdf. [https://perma.cc/P698-SZ42].

152. Ruchir Puri, *Mitigating Bias in AI Models*, IBM RSCH. BLOG (2018), https://www.ibm.com/blogs/research/2018/02/mitigating-bias-ai-models/ [https://perma.cc/V4WE-LQU6].

153. Nabil Hassein, *Against Black Inclusion in Facial Recognition*, DIGITAL TALKING DRUM (Aug. 15, 2017), https://digitaltalkingdrum.com/2017/08/15/against-black-inclusion-in-facial-recognition/ [https://perma.cc/5BMW-FTDZ].

fundamentally punitive system of racialized exclusion and social control. According to Hassein, efforts to render the carceral system more efficient through the deployment of more accurate technology are harmful, and in direct conflict with the goal of promoting thriving and inclusive communities. For historically marginalized people, inclusion in facial recognition will likely result in their exclusion from broader society. Hassein proposes a different set of values on which to base a counter-imaginary about the future of the carceral state, one which centers on the pursuit of agency and healing within historically marginalized communities.

Importantly, this perspective doesn't advocate for the wholesale rejection of data and technology in the criminal legal system. Rather, it argues for the refusal of key concepts and assumptions that drive AI models in this context. An abolitionist reimagining of AI in criminal law would require shifting away from measuring criminal behavior and towards understanding processes of criminalization, from supporting law and order towards increasing community safety and self-determination, and from surveilling risky populations towards holding accountable state officials.

Shifting these fundamental assumptions is very important, as they inform 1) how we diagnose the problems we aim to solve, 2) what data we consider important for understanding the problem and 3) how we make claims based on the available data. To truly address the ethical stakes of artificial intelligence, we must engage with an abolitionist "sociotechnical imaginary" in order to redefine "not only what is attainable through science and technology, but also of how life ought, or ought not, be lived."[154] This requires a fundamental shift in the narrative tools we use to diagnose the problems of the carceral state and construct subjects of analysis from available data.[155] Rather than use data to profile and manage "risky populations," we should build systems to evaluate the impacts of key policies and decision-making practices, as well as build the infrastructure needed to increase accountability for the authority figures who drive outcomes.

For example, in the case of pretrial reform, various attempts have been made to decrease judges' use of cash bail as a means of detaining defendants prior to their trial. These include state supreme court orders that mandate that judges inquire about a defendant's ability to pay prior to setting bail, as well as statutory guidelines that attempt to significantly reduce the use of cash bail for defendants who are categorized as low- or moderate-risk on a risk assessment.[156] Very little research exists regarding

154. SHEILA JASANOFF & SANG-HYUN KIM, DREAMSCAPES OF MODERNITY: SOCIOTECHNICAL IMAGINARIES AND THE FABRICATION OF POWER 4 (2015).

155. As Ruha Benjamin explains, "An abolitionist toolkit, in this way, is concerned not only with emerging technologies, but also with the everyday production, deployment and interpretation of data. Such toolkits can be focused on computational interventions, but they do not have to be. In fact *narrative tools* are essential." RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE 363 (2019).

156. Ky. Rev. Stat. Ann. § 431.066(2) (codifying H.B. 463) (instructing judges to consider the risk assessment when considering release and bail); Ky. Rev. Stat. Ann. § 431.066(3) (instructing release on unsecured bond or own recognizance for low risk defendants); Ky. Rev. Stat. Ann. § 431.066(4) (instructing release on unsecured bond or own recognizance for moderate risk defendants with possible supervision, monitoring or other conditions of release); Ky. Rev. Stat. Ann. § 27A.096(1,2,3) (instructing judges to follow guidelines set by the Supreme Court on pretrial release or supervision for moderate and high risk defendants); Brangan v.

the impact of these reforms, but initial evaluations suggest that these efforts do not translate into significant or sustained changes in courtroom practices.[157] In most jurisdictions, the impact of these reforms is simply not known, because the requisite data needed to answer basic questions about a given policy's impact are either not collected or not made available to researchers or the public for scrutiny.[158]

In response to the lack of available data, community organizations across the U.S. are undertaking grassroots data collection efforts to gather information key to answering basic questions about these pretrial reforms.[159] These data collection efforts are examples of Nader's classic concept of "studying up" to understand how power and responsibility are exercised and drive social outcomes.[160] In some cases, the courts already collect data that could be used to understand judge behavior. However, the interpretation of that data is limited to the realm of defendant criminality, as data that is integrated into risk assessment instruments. To date, court docket data has not been used to understand, say, judges' compliance with a state supreme court order regarding bail, even though it could be used in that way. When researchers and community groups attempt to access court data for these purposes, they encounter a variety of administrative obstacles.[161] As a result, some researchers and community groups have undertaken labor intensive data collection initiatives in order to collect information that the courts already collect but will not provide.

Other data that are key to understanding court practices are simply absent from government records because they are not considered by the state to be useful or important. As James Scott argues "builders of the modern nation-state do not merely describe, observe, and map; they strive to shape a people and landscape that will fit their techniques of observation . . . there are virtually no other facts for the state than those that are contained in documents."[162] The process of data collection is a process of rendering some phenomena visible and other phenomena invisible. In the carceral state, data regarding the decisions and actions of key state officials, such as judges, prosecutors, and police officers, are scant. This makes it very challenging to build robust systems of accountability around their actions. Thus, court watching initiatives around the country are seeking to create new data sets, ones which enable them to build stronger accountability for decision makers and ensure that reform efforts

Commonwealth, 80 N.E.3D 949 (Mass. 2017) (holding that judges must issue findings of fact when setting unaffordable bail for indigent defendants).

157. Stevenson, *supra* note 100.

158. *Id.*; Koepke & Robinson, *supra* note 100.

159. COURT WATCH NYC, *About Court Watch NYC*, https://www.courtwatchnyc.org/about [https://perma.cc/HSH6-9RBX] (last visited Aug. 1, 2020); COURTWATCH MA, *Community, Accountability, Justice.*, https://www.courtwatchma.org/ [https://perma.cc/RFK9-EQEZ] (last visited Aug. 1, 2020).

160. LAURA NADER, UP THE ANTHROPOLOGIST: PERSPECTIVES GAINED FROM STUDYING UP 284 (1972).

161. For example, in my experiences working with members of the public who participated in courtwatch programs, I discovered that volunteers' requests for access to public data, such as docket information, were frequently denied, without an explanation. Similarly, as an academic researcher, I have faced significant delays in accessing data from administrations of the court after they discovered that the goal of my data analysis was to scrutinize the behavior patterns of specific authority figures, such as judges.

162. JAMES C. SCOTT, SEEING LIKE A STATE: HOW CERTAIN SCHEMES TO IMPROVE THE HUMAN CONDITION HAVE FAILED 82–83 (1998).

translate into real changes in the daily practices of officials within the criminal legal system.[163]

In addition to collecting data regarding the behavior of authority figures, court watch initiatives also expand the types of data which are collected to measure the harm and benefit of new carceral technologies, such as electronic monitoring and mandatory drug testing. This includes written and oral testimonies from impacted individuals regarding the challenges these technologies impose on their lives. This is a notable departure from most academic research which tries to evaluate the harms and benefits of technocratic interventions in the criminal legal system. As Kilgore notes, research on technology like electronic monitors frequently leaves out the perspectives and lived experiences of the people who are subjected to these interventions.[164] Many of these evaluations are based on very narrow outcome measures, such as "recidivism," which are measured using police and court data that is misinterpreted in terms of defendant behavior.

There's a clear hierarchy of evidence when it comes to making "evidence-based" claims about what works in the carceral state. As Marston and Watts argue, "far from being a neutral concept, evidence-based policy is a powerful metaphor in shaping what forms of knowledge are considered closest to the 'truth' in decision-making processes and policy argument."[165] In the case of court watching and other community data collection efforts, advocates and directly impacted individuals are expanding the ontological categories of analysis in conversations regarding "what works" in the criminal legal system. In doing so, they strive to paint a more complete picture of how carceral technologies impact the lives of those subjected to them.

These efforts are less concerned with guaranteeing some elusive notion of "objectivity" and more interested in collecting and disseminating information that can be used to enhance the public's understanding about what policies to adopt and which officials to elect to office. This work seeks to wrest back control over what is considered authoritative knowledge and expertise by centering the voices and perspectives of ordinary citizens and the lived experiences of impacted communities. In doing so, they illustrate the ways that one might engage with data-driven regimes within the carceral state from an abolitionist perspective.

An abolitionist understanding of the role and function of the carceral state provides us with a first-principles framework to fundamentally reformulate the questions we ask, the way we characterize existing data, and how we identify and fill gaps in existing data regimes of the carceral state. The igniting of an abolitionist sociotechnical imaginary is especially important in the current political moment, when the term

163. For example, CourtWatch MA launched a First 100 Days monitoring campaign in the first quarter of 2019. The goal of the campaign was to collect data regarding the decisions and requests made by assistant district attorneys in the courts, in order to track the extent to which they were upholding specific campaign promises that the newly elected District Attorney Rachael Rollins had made on the campaign trail. COURTWATCH MA, *supra* note 159.

164. JAMES KILGORE, ELECTRONIC MONITORING IS NOT THE ANSWER: CRIT. REFLECTIONS ON A FLAWED ALTERNATIVE, URBANA-CHAMPAIGN INDEP. MEDIA CTR. (2015).

165. Greg Marston & Rob Watts, *Tampering with the Evidence: A Critical Appraisal of Evidence-based Policy-making*, 3 DRAWING BOARD 143, 163 (2003).

"artificial intelligence" has been deployed as a means of justifying and de-politicizing the expansion of state and private surveillance amidst a growing crisis of legitimacy for the U.S. prison industrial complex. Under the authoritative rubric of "evidence-based reform" and "artificial intelligence," law enforcement officials have reframed contentious social issues in terms of technocratic shortcomings, or issues of information access and interpretation. Efforts to increase the accuracy of predictive and evaluative systems run the risk of circumscribing deeper ideological and epistemological struggles within a narrow technocratic debate about how to make these processes more valid, accurate, and fair.

The key questions are whether predictive tools reflect and reinforce punitive practices that drive disparate outcomes, and how data regimes interact with the penal ideology to naturalize these practices. Conversations regarding the ethical stakes of AI in criminal law must interrogate the default logics and assumptions of the carceral state, in order to address the foundational violence of law enforcement and courtroom practices. Only then can we hope to re-imagine the use of data and technology to explore and substantiate a political vision that centers the creation of lasting alternatives to punishment and imprisonment, by increasing community safety and centering values of self-determination and healing in marginalized communities.