DISCREDITED DATA

Ngozi Okidegbe†

Jurisdictions are increasingly employing pretrial algorithms as a solution to the racial and socioeconomic inequities in the bail system. But in practice, pretrial algorithms have reproduced the very inequities they were intended to correct. Scholars have diagnosed this problem as the biased data problem: pretrial algorithms generate racially and socioeconomically biased predictions because they are constructed and trained with biased data.

This Article contends that biased data is not the sole cause of algorithmic discrimination. Another reason pretrial algorithms produce biased results is that they are exclusively built and trained with data from carceral knowledge sources—the police, pretrial services agencies, and the court system. Redressing this problem will require a paradigmatic shift away from carceral knowledge sources toward non-

† Associate Professor of Law & Assistant Professor of Computing and Data Science, Boston University. I appreciate the helpful feedback and support I received from Ifeoma Ajunwa, Amna Akbar, Chaz Arnett, Kristen Bell, Jack M. Balkin, Rachel E. Barkow, Kenitra Brown, Hannah Bloch-Wehba, I. Bennett Capers, Jenny Carroll, Carliss Chatman, Bryan H. Choi, Rebecca Crootof, Colin Doyle, Jessica Eaglin, Andrew Ferguson, Jonathan Glater, Leigh Goodmark, Lauryn P. Gouldin, Nikolas Guggenberger, Aziz Z. Huq, Thea Johnson, Thomas E. Kadri, Zachary D. Kaufman, Jennifer Laurin, Daniel S. McConkie, Pamela Metzger, Jamelia Morgan, Rachel Moran, Justin Murray, Emily R.D. Murphy, Alicia Solow-Niederman, Najarian Peters, Francesca Procaccini Rashida Richardson, Anna Roberts, Jenny Roberts, Maybell Romero, Ric Simmons, Jocelyn Simonson, Megan Stevenson, Stewart Sterk, I. India Thusi, Matt Wansley, Rebecca Wexler, Patricia Williams, Jonathan P. Witmer-Rich, and Corey Rayburn Yung. This Article benefited greatly from presentations and discussions at Northern Illinois University College of Law's Criminal Law Works in Progress series, the 2020 Joint Lutie A. Lytle Black Women Law Faculty and John Mercer Langston Writing Winter Workshop, the 2020 Junior Law & Tech Workshop, the 2021 Vanderbilt Criminal Justice Roundtable, Center for Professional and Applied Ethics at UNCC, CrimFest 2020, the AALS 2021 Works-in-Progress Panel for the Criminal Justice Section, the 2021 Cardozo Junior Faculty Workshop, the Women of Color Collective Junior Legal Scholars Workshop, 2021 Junior Faculty Summer Workshop Series at Ohio State University's Moritz College of Law, 2022 American Constitution Society's Schmooze, and faculty workshops at Boston University School of Law, University of Connecticut School of Law, University of Hawaii's William S. Richardson School of Law, UC Hastings Law, New York Law School, and California Western School of Law. I would also like to thank Laura Catherine McCrary and Bella Pori for their excellent research assistance and for extraordinary editorial assistance by the members of the Cornell Law Review, especially Kyle Burrus, Shayla Hinson, and Milagros Pilipiak.

carceral knowledge sources. This Article explores knowledge produced by communities most impacted by the criminal legal system ("community knowledge sources") as one category of non-carceral knowledge sources worth utilizing. Though data derived from community knowledge sources have traditionally been discredited and excluded in the construction of pretrial algorithms, tapping into them offers a path toward developing algorithms that have the potential to produce racially and socioeconomically just outcomes.

INTRODUCTION

The bail system is rife with racial and socioeconomic inequities. Life-altering decisions about pretrial release, surveillance, and detention are often based on race and financial ability to pay rather than on a defendant's risk of pretrial misconduct.[1] These decisions can have severe physical, psychological, financial, and socioeconomic consequences for defendants, their families, and their communities.[2] One proffered cause of this inequity is the significant discretion afforded to bail judges, which has enabled race and class bias to taint the bail determination process.[3] In an effort to redress this problem, jurisdictions across the country are increasingly adopting pretrial risk assessment algorithms.[4] These algorithms are designed to predict the likelihood that a defendant will miss a court appearance or be arrested for pretrial crime if released before

---

1 *See* David Arnold, Will Dobbie & Crystal S. Yang, *Racial Bias in Bail Decisions*, 133 Q.J. ECON. 1885, 1892, 1906, 1917, 1917 n.14 (2018) (noting that Black defendants are released at lower rates than similarly situated white defendants); COLIN DOYLE, CHIRAAG BAINS & BROOK HOPKINS, BAIL REFORM: A GUIDE FOR STATE AND LOCAL POLICYMAKERS 7, 12, 23, (2019), https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=9a804d1d-f9be-e0f0-b7cd-cf487ec70339&forceDialog=0 [https://perma.cc/A6PS-X5BG] (discussing the fact that "a defendant's release depends upon an ability to pay" and that "[w]ealthy defendants walk free while poor defendants languish in jail"). In terms of pretrial misconduct, the bail system is designed to release defendants except those posing a risk of non-appearance, obstruction of justice, and, in most jurisdictions, danger to public safety. Shima Baradaran Baughman, *Dividing Bail Reform*, 105 IOWA L. REV. 947, 984 (2020).

2 *See* Russell M. Gold, *Jail as Injunction*, 107 GEO. L.J. 501, 501 (2019) (documenting the various harms of pretrial incarceration on a defendant.).

3 One critical strand of bail scholarship has explored how unfettered judicial discretion contributes to racial inequities. *See, e.g.*, Crystal S. Yang, *Toward an Optimal Bail System*, 92 N.Y.U. L. REV. 1399, 1408–09 (2017) (contending that racial inequities in release and in detention rates is largely due to the current state of discretionary bail determinations).

4 Sandra G. Mayson, *Bias in, Bias out*, 128 YALE L.J. 2218, 2221 (2019) ("Over the last five years, criminal justice risk assessment has spread rapidly."); Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 61–62 (2017) ("Predictive technologies increasingly appear at every stage of the criminal justice process."); PRETRIAL JUST. INST., THE STATE OF PRETRIAL JUSTICE IN AMERICA 3, 13 (2017), https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=484affbc-d944-5abb-535f-b171d091a3c8&forceDialog=0 [https://perma.cc/X2YG-URT2] ("25% of people living in the United States now reside in a jurisdiction that uses a validated evidence-based pretrial assessment.").

trial.[5] The hope is that bail judges will condition bail decisions on the predictions produced by these algorithms, thereby reducing the racial and socioeconomic biases held by decisionmakers that often drive outcomes.[6] And increasingly, algorithms are influencing bail decisions.[7]

But in practice, the increased use of pretrial algorithms has tended to reproduce existing racial and socioeconomic

---

5 DOYLE, BAINS & HOPKINS, *supra* note 1, at 13–14. Though this Article focuses exclusively on the use of risk assessment algorithms in bail, there is a body of scholarship on their use in sentencing. *See, e.g.*, Eaglin, *supra* note 4 (examining how developers construct algorithms that produce recidivism risk); Erin Collins, *Punishing Risk*, 107 GEO. L.J. 57, 58–63 (2018) (illuminating the unintended consequences of actuarial sentencing); John Monahan, *Risk Assessment in Sentencing*, *in* 4 REFORMING CRIMINAL JUSTICE: PUNISHMENT, INCARCERATION, AND RELEASE 77, 77–78 (Erik Luna ed., 2017), https://law.asu.edu/sites/default/files/pdf/academy_for_justice/5_Criminal_Justice_Reform_Vol_4_Risk-Assessment-in-Sentencing.pdf [https://perma.cc/TKS9-TET3] (discussing the use of risk assessment in sentencing generally and recommending changes to reduce mass incarceration). Predictive technologies are also being used in policing and prosecution. *See, e.g.*, ANDREW GUTHRIE FERGUSON, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT 29–32, 62–69 (2017) (discussing the use of predictive policing technology in the criminal justice system); Andrew Guthrie Ferguson, *Predictive Prosecution*, 51 WAKE FOREST L. Rev. 705, 705–08 (2016) (explaining that risk predictions shape prosecutors' positions at bail, charging, and sentencing).

6 Megan Stevenson & Sandra G. Mayson, *Pretrial Detention and Bail*, *in* 3 REFORMING CRIMINAL JUSTICE: PRETRIAL AND TRIAL PROCESSES 21, 34 (Erik Luna ed., 2017), https://law.asu.edu/sites/default/files/pdf/academy_for_justice/2_Reforming-Criminal-Justice_Vol_3_Pretrial-Detention-and-Bail.pdf [https://perma.cc/4S3A-EK6Y] ("There is reason to be optimistic about the actuarial turn in pretrial practice. Risk-assessment tools should reduce the subjective, irrational bias that distorts judicial decision-making."); Christopher Slobogin, *Preventive Justice: How Algorithms, Parole Boards, and Limiting Retributivism Could End Mass Incarceration*, 56 WAKE FOREST L. REV. 97, 105 (2021) ("[T]he quantified results of well-validated [risk assessment instruments] can provide a concrete, rational basis for diversion or release. If, as recommended . . . adherence to those results is *required* in most circumstances, the human urge to incapacitate those in the law's grasp can be even more effectively resisted because decision-makers must obey the objective facts.").

7 Lauryn P. Gouldin, *Defining Flight Risk*, 85 U. CHI. L. REV. 677, 713, 718–19 (2018); Sandra G. Mayson, *Dangerous Defendants*, 127 YALE L.J. 490, 492–93 (2018); Chaz Arnett, *From Decarceration to E-Carceration*, 41 CARDOZO L. REV. 641, 651–52 (2019); Jenny E. Carroll, *Pretrial Detention in the Time of COVID-19*, 115 NW. U. L. REV. 59, 65 (2020).

inequities.[8] There are many reasons for this phenomenon.[9] In previous scholarship, I have written about how these inequities are fueled in part by the fact that communities most harmed by the criminal legal system's operation are unable to stop, shape, or oversee these algorithms.[10] In this Article, I focus on the role

---

8 Dorothy E. Roberts, *Digitizing the Carceral State*, 132 HARV. L. REV. 1695, 1699, 1708, 1713–14 (2019) (reviewing VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018)) ("Moreover, [these algorithms'] forecasts of the future are based on data that were produced by existing racial discrimination in systems such as policing, housing, education, health care, and public assistance. The future predicted by today's algorithms, therefore, is pre-determined to correspond to past racial inequality."); PARTNERSHIP ON AI, REPORT ON ALGORITHMIC RISK ASSESSMENT TOOLS IN THE U.S. CRIMINAL JUSTICE SYSTEM 3, 15, 18–19 (2019), https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/ [https://perma.cc/PN9D-68Q5] ("Although the use of these tools is in part motivated by the desire to mitigate existing human fallibility in the criminal justice system, it is a serious misunderstanding to view tools as objective or neutral simply because they are based on data. While formulas and statistical models provide some degree of consistency and replicability, they still share or amplify many weaknesses of human decision-making."); Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/4G83-MDAS] (documenting the racial disparities produced by the use of the COMPAS algorithm in Florida). *But see* David Arnold, Will S. Dobbie & Peter Hull, *Measuring Racial Discrimination in Algorithms* 3 (Nat'l Bureau of Econ. Rsch., Working Paper No. 28222, 2021) (questioning the existence of algorithmic discrimination and instead contending that there is insufficient data around the extent to which algorithms produce racially biased and predictions).

9 A major reason that algorithms produce biased results is that they are constructed with biased data. For more on the biased data diagnosis, see CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 199–200 (2016); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 673–76 (2016) (discussing the role of bias in data and what can be done about it); Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109, 1119–20, 1148 (2017) (noting the problem of biased data in the context of predictive policing); Nizan Geslevich Packin & Yafit Lev-Aretz, *Learning Algorithms and Discrimination*, *in* RESEARCH HANDBOOK ON THE LAW OF ARTIFICIAL INTELLIGENCE 88, 109–11 (Woodrow Barfield & Ugo Pagallo eds., 2018); Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1025–27 (2017) (arguing that "if we believe that the real-world facts, on which algorithms are trained and operate, are deeply suffused with invidious discrimination, then our prescription to the problem of racist or sexist algorithms is *algorithmic affirmative action*").

10 I explore this subject matter more comprehensively in a prior article: Ngozi Okidegbe, *The Democratizing Potential of Algorithms?*, 53 CONN. L. REV. 739 (2022). In *The Democratizing Potential of Algorithms?*, my prescription is that most impacted communities should be endowed with power over if and how algorithms are adopted, implemented, and overseen. One unaddressed issue concerned the knowledge sources used in algorithmic construction. As explored in this Article, failure to contend with the dominance of carceral knowledge sources in algorithmic construction will lead to the creation of algorithms that reproduce existing inequities regardless of whether the paradigm governing these tools involves or is controlled by most impacted communities.

of data, specifically the fact that these algorithms have a "data source selection problem." They are built on and trained exclusively with data derived from criminal legal institutions, such as the police, pretrial services agencies,[11] and the court system (institutions that I refer to as "carceral knowledge sources"[12]). Arrest data, conviction data, and court appearance records are examples of the kinds of data these sources produce.[13] It is well established that these sources produce data that are infected with racial and socioeconomic bias.[14] The exclusive reliance on data from such sources has

---

[11] Pretrial services agencies are state, county, city, or nonprofit agencies under government contract that collect and produce information about pretrial detainees in a specific jurisdiction that is designed for use by bail judges. One example of a pretrial services agency is the New York City Criminal Justice Agency that operates in New York City. *See* Freda F. Solomon & Russell F. Ferri, *Reducing Unnecessary Pretrial Detention: CJA's Manhattan Supervised Release Program*, 42 N.Y.C. CRIM. JUST. AGENCY 1 (2017), https://www.nycja.org/publications/brief-reducing-unnecessary-pretrial-detention-cjas-manhattan-supervised-release-program [https://perma.cc/CF9L-TMUA].

[12] I use the term "carceral knowledge sources" to refer to the data derived from the knowledge produced by political and social systems that formally control or promote punishment and incarceration. This definition borrows from the definition of carcerality provided by Tracy Lachica Buenavista, *Model (Undocumented) Minorities and "Illegal" Immigrants: Centering Asian Americans and US Carcerality in Undocumented Student Discourse*, 21 RACE ETHNICITY & EDUC. 78, 78 (2018).

[13] For example, nearly all pretrial algorithms are built with arrest data. *See* STAN. L. SCH. POL'Y LAB, RISK ASSESSMENT FACTSHEET: PUBLIC SAFETY ASSESSMENT (PSA) 1 (2019), https://www-cdn.law.stanford.edu/wp-content/uploads/2019/05/PSA-Sheet-CC-Final-5.10-CC-Upload.pdf [https://perma.cc/BVL3-CC6A] [hereinafter STAN. L. SCH. POL'Y LAB, PSA] (noting that the PSA was developed using arrest data); STAN. L. SCH. POL'Y LAB, RISK ASSESSMENT FACTSHEET: CORRECTIONAL OFFENDER MANAGEMENT PROFILING FOR ALTERNATIVE SOLUTIONS (COMPAS) PRETRIAL RELEASE RISK SCALE—II (PRRS-II) 1 (2019), https://www-cdn.law.stanford.edu/wp-content/uploads/2019/06/COMPAS-PRRS-II-Factsheet-Final-6.20.pdf [https://perma.cc/RH57-VBKN] [hereinafter STAN. L. SCH. POL'Y LAB, COMPAS PRRS-II] (noting that the COMPAS PRRS-II was developed using arrest data); STAN. L. SCH. POL'Y LAB, RISK ASSESSMENT FACTSHEET: VIRGINIA PRETRIAL RISK ASSESSMENT INSTRUMENT (VPRAI) 1 (2019), https://www-cdn.law.stanford.edu/wp-content/uploads/2019/06/VPRAI-Factsheet-FINAL-6-20.pdf [https://perma.cc/JE6V-L54W] [hereinafter STAN. L. SCH. POL'Y LAB, VPRAI] (noting that the VPRAI was developed using arrest data).

[14] Many scholars have discussed the racialized and socioeconomically disparate nature of data produced by the criminal legal system. *See, e.g.*, Anna Roberts, *Arrests as Guilt*, 70 ALA. L. REV. 987, 1020–21 (2019) [hereinafter Roberts, *Arrests*] (discussing the disproportionate production of arrests in poor and racially marginalized areas); Anna Roberts, *Convictions as Guilt*, 88 FORDHAM L. REV. 2501, 2509 (2020) [hereinafter Roberts, *Convictions*] (discussing the racial and socioeconomic inequities around how convictions are produced); THOMAS H. COHEN & BRIAN A. REAVES, BUREAU OF JUST. STAT., PRETRIAL RELEASE OF FELONY DEFENDANTS IN STATE COURTS 1, 17 (2007), https://www.bjs.gov/content/pub/pdf/prfdsc.pdf [https://perma.cc/SB3R-SLK9] (illustrating the racial and socioeconomic disparity in the data produced about pretrial non-appearance); I.

created algorithms that maintain the very inequities that their implementation was designed to dismantle.[15]

The data source selection problem is not exclusive to pretrial algorithms.[16] Most predictive technologies in use in the criminal legal system tend to be built and trained with data from carceral knowledge sources.[17] But conversations about algorithmic discrimination have largely failed to examine the data source selection problem and instead have focused on the biases in the data from currently used knowledge sources. I refer to this line of reasoning as the "biased data diagnosis."[18] Scholars and policymakers who have accepted the biased data diagnosis have encouraged the development of better datasets,[19] technical adjustments to algorithmic systems,[20] and better data auditing practices[21] to root out bias. But

---

India Thusi, *Radical Feminist Harms on Sex Workers*, 22 LEWIS & CLARK L. REV. 185, 185–86, 187 n.6, 224–25 (2018) (noting the racial, gendered, and class inequities around prostitution convictions); Jamelia N. Morgan, *Rethinking Disorderly Conduct*, 109 CALIF. L. REV. 1637, 1641–44 (noting the racial, class, and ableist inequities around disorderly conduct convictions).

15 *See generally* Jessica M. Eaglin, *Technologically Distorted Conceptions of Punishment*, 97 WASH. U. L. REV. 483 (2019) (problematizing the failure of sentencing algorithms to redress mass incarceration); RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE (2019) (exploring the sociopolitical reasons for which technologies support the inequities that they are implemented to correct); Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 CARDOZO L. REV. 1671 (2020) (noting this paradox in the employment context).

16 It is important to note that this Article is primarily focused on the data source selection problem in predictive technologies, though the problem is pervasive within the criminal legal system itself. The effect that the data source selection problem has on the criminal legal system and current efforts are considered in depth in future work. *See* Ngozi Okidegbe, *Beyond Carceral Data* (Aug. 20, 2022) (unpublished manuscript) (on file with author).

17 *See, e.g.*, Rashida Richardson, Jason M. Schultz & Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. ONLINE 15, 15, 21 (2019) (noting how police predictive systems are primarily constructed with data produced by policing); Ferguson, *supra* note 9, at 1123–24, 1137–38 (describing the data used to build predictive policing systems); Eaglin, *supra* note 4, at 74 (noting how sentencing algorithms rely on information produced by the criminal legal system).

18 For an example of scholarship based on biased data diagnosis, see Mayson, *supra* note 4, at 2281–87.

19 Frank Pasquale, *Data-Informed Duties in AI Development*, 119 COLUM. L. REV. 1917, 1917, 1931–35 (2019) (advocating for a legal and regulatory framework to compel developers to use more complete and representative datasets).

20 *See* Jason R. Bent, *Is Algorithmic Affirmative Action Legal?*, 108 GEO. L.J. 803, 814, 817, 819–20 (2020) (exploring the technical adjustments advocated to redress algorithmic discrimination).

21 Richardson, Schultz & Crawford, *supra* note 17, at 41, 47–48; Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 110 (2017) (proposing the use of "algorithmic impact statements" to audit and correct the

focusing on ways to reduce bias in the data currently in use has inadvertently "locked in"[22] the problem and it has obscured a key cause of biased data: the choice to exclusively use carceral knowledge sources.[23] The failure to engage with the systemic problem around carceral knowledge sources will continue to hamper our ability to fully contend with the ways in which algorithmic systems maintain and reproduce societal inequities.

This Article is one of the first to explore the data source selection problem and how it could be redressed. Its central claim is that the exclusive use of carceral knowledge sources limits the capacity for algorithms to redress historical and current inequities in the pretrial system by tethering pretrial algorithms to data from the very same institutions responsible for the current bail crisis. Redressing this problem will require replacing the dominance of carceral knowledge sources in algorithmic construction with non-carceral knowledge sources.[24] Knowledge produced by those most affected by the criminal legal system—which I refer to as "community knowledge sources"—is one such type of non-carceral knowledge source that has the potential to reduce existing inequities.[25] I focus on a specific kind of data from community knowledge sources: the qualitative data about the criminal legal system produced by currently and formerly incarcerated people hailing from communities most harmed by mass criminalization and incarceration.[26] One barrier to this proposal is that developers have traditionally excluded and discredited data from non-carceral knowledge sources in the construction of pretrial algorithms.[27] But overcoming this

---

discriminatory impact of new technologies, using the example of policing algorithms).

22 Rebecca Crootof, *"Cyborg Justice" and the Risk of Technological-Legal Lock-In*, 119 COLUM. L. REV. F. 233, 235 (2019) (using "technological-legal lock-in" to refer to the "translating [of] rules and decisionmaking procedures into algorithms [which] grants them a new kind of permanency which creates an additional barrier to legal evolution").

23 *See infra* subpart III.A.

24 The term "non-carceral knowledge sources" refers to knowledge sources that are not formally connected to the political and social systems that control or facilitate punishment and incarceration and that produce data that is not required for the criminal legal system's current functioning.

25 *See infra* Part III.

26 *See infra* subpart III.B.

27 One justification that has been given for the exclusion of community knowledge sources in algorithmic construction is that these sources produce qualitative data, while developers only use quantitative data to build algorithmic systems. However, the problem with this justification is that developers do use qualitative data from carceral knowledge sources to build currently employed

barrier is a precondition to creating algorithms that have the potential to orient the criminal legal system away from its racist, classist, and punitive tendencies.

This Article makes three contributions to the existing literature. First, it fills a gap in the current scholarly discourse around algorithmic discrimination. Though a growing number of scholars (myself included) have begun studying and identifying the institutional, systemic, and sociopolitical dimensions of algorithmic discrimination, less attention has been paid to its epistemic dimension.[28] By examining the knowledge sources behind the data in use, this Article expands the contours of the traditional critique about algorithmic discrimination and connects the critique to wider conversations around epistemic oppression and subjugated knowledge.[29] Second, through unpacking the data source selection problem, it offers an explanation for why certain knowledge is normatively understood to be "data" that are credible for use in algorithmic systems, while other knowledge is not. The very notion of which knowledge is "credited data" or "discredited data" is a source of bias itself, affecting the extent to which algorithmic systems can be de-biased. Third, it positions community knowledge sources as knowledge sources that warrant scholarly attention. It juxtaposes the exclusive use of carceral knowledge sources with the non-use of community knowledge sources to reveal how the data source selection problem has served racial ends.

This Article proceeds in four parts. Part I details the rise of pretrial algorithms as a proposed solution to racial and socioeconomic inequities in the bail system. It also explains the algorithmic construction process and demonstrates how it relies exclusively on data from carceral knowledge sources. Part II details how current accounts about algorithmic discrimination have missed the data source selection problem.

---

algorithms. For more information on this aspect of algorithmic construction, see *infra* subpart III.A.

28 For examples of scholarship on this point, see Ngozi Okidegbe, *When They Hear Us: Race, Algorithms and the Practice of Criminal Law*, 29 KAN. J.L. & PUB. POL'Y 329, 331–34 (2020) (discussing the challenge that the proliferation of algorithms poses to the pursuit of racial justice in the criminal justice system); Rashida Richardson, *Racial Segregation and the Data-Driven Society: How Our Failure to Reckon with Root Causes Perpetuates Separate and Unequal Realities*, 36 BERKELEY TECH. L.J. 1051, 1070–90 (2022) (identifying the role that systemic racial segregation plays in algorithmic discrimination); BENJAMIN, *supra* note 15, at 1679 (identifying the sociopolitical conditions that promote the use of algorithms that produce racially inequitable results).

29 *See infra* subpart III.A.

Part III argues for the use of non-carceral knowledge sources in algorithmic construction. By examining the qualitative data from currently and formerly incarcerated people, it provides a concrete example of one of the kinds of data that can be derived from the knowledge produced by community knowledge sources. Part IV responds to some potential objections to the proposal presented in this Article.

For helpful reading, three caveats are in order. First, this Article is primarily focused on the data source selection problem in predictive technologies, though the problem is pervasive within the criminal legal system itself. The effect that the data source selection problem has on the criminal legal system and on current criminal legal reform efforts is considered in depth in future work.[30] Second, though this Article advocates for dismantling the dominance of carceral knowledge sources in algorithmic construction, it leaves open the question of whether there is a role for carceral knowledge sources (albeit minimalized) in algorithmic construction in the future.[31] Third, resolving the data source selection problem requires a deep rethinking around data and around the goals and objectives of predictive technologies and the systems to which they are destined for use in.[32] This rethinking requires input from a variety of stakeholders including data scientists, programmers, statisticians, criminologists, sociologists, policymakers, and communities most impacted by the criminal legal system. For this reason, this Article does not provide a roadmap as to how to operationalize its proposal. The point of surfacing the data source selection problem is to start an important conversation around how the exclusive use of knowledge sources historically and currently implicated in mass incarceration within algorithmic construction operates as a barrier to creating algorithms in line with racial and socioeconomic justice.

---

30 Okidegbe, *supra* note 16.

31 I plan to take up this question in depth in future work. *See* Ngozi Okidegbe, *Abolitionist Algorithms?* (Oct. 1, 2022) (unpublished manuscript) (on file with author).

32 Erin Collins, *Abolishing the Evidence-Based Paradigm*, BYU L. REV. (forthcoming 2022) (contending that evidence-based tools are unable to meaningfully reform the criminal legal system, since these tools are steeped in the same neo-liberal ideology and epistemology responsible for mass incarceration and race and class subordination within the criminal legal system).

I
CONTEXTUALIZATION

This Part contextualizes the rise of pretrial algorithms and details the process by which these algorithms are constructed.

A. The Rise of Pretrial Algorithms

The use of pretrial algorithms is on the rise. The increased use of pretrial algorithms has triggered significant resistance from scholars, policymakers, and racial justice activists.[33] Their concerns have coalesced around the fact that these pretrial algorithms are often implemented without meaningful democratic input, public notice, or oversight, despite the political, financial, and racial and socioeconomic impacts that these technologies have on enacting jurisdictions.[34] Recently, this growing resistance was one of the reasons for the defeat of SB-10 in California, which would have mandated the statewide use of risk assessment algorithms in the pretrial system.[35] Despite this resistance, the trend toward pretrial algorithms persists.

---

33 *See* Sean Allan Hill II, *Bail Reform and the (False) Racial Promise of Algorithmic Risk Assessment*, 68 UCLA L. REV. 910, 963–68 (2021) (discussing how pretrial algorithms are technically flawed, encourage divestment from Black and Latinx communities, and maintain a veneer of impartiality while erroneously classifying Black people as higher-risk); Okidegbe, *supra* note 10, at 747 (arguing that algorithmic governance "serves merely to entrench and to legitimate the existing democratic exclusion experienced by racially marginalized people in the crafting and implementation of criminal laws and policies"). It is important to note that these concerns have also been noted in regard to use of predictive technologies in other parts of the criminal legal system. *See* Eaglin, *supra* note 4, at 95–96, 105, 108–09 (discussing the problem in the context of sentencing algorithms); Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 803 (2014) (discussing the use of risk assessment tools on constitutional and policy grounds with a focus on their use in sentencing); Hannah Bloch-Wehba, *Visible Policing: Technology, Transparency, and Democratic Control*, 109 CALIF. L. REV. 917, 954–956 (2021) (discussing this problem in the context of policing algorithms); Kate Weisburd, *Punitive Surveillance*, 108 VA. L. REV. 147, 147, 173–84 (2022) (discussing this problem in the context of electronic monitoring technologies). Additionally, there is a burgeoning set of scholarship considering human preferences around algorithmic decision-making. *See generally* Derek E. Bambauer & Michael Risch, *Worse Than Human?*, 53 ARIZ. ST. L.J. 1091, 1091, 1093–95 (2021) (outlining consumer preferences about the role of algorithms).

34 *See* Okidegbe, supra note 10, at 743.

35 *See* Patrick McGreevy, *Prop. 25, Which Would Have Abolished California's Cash Bail System, Is Rejected by Voters*, L.A. TIMES, https://www.latimes.com/california/story/2020-11-03/2020-california-election-prop-25-results [https://perma.cc/XBR8-39AR] (last updated Nov. 4, 2020) (noting that the defeat was attributed to an alliance between the bail bond industry and grassroots organizations).

Pretrial algorithms are one of the many pretrial reform measures being implemented around the country. They are part of a growing movement to use data-informed predictive technologies to correct the racial and socioeconomic biases that have ushered in racialized mass incarceration.[36] The term "pretrial algorithm" is currently used to refer to any assessment that employs an actuarial method, big data, and information about a defendant to predict the likelihood that, if released pending the disposition of their criminal case, the defendant will fail to appear or be arrested for a pretrial crime.[37] The predictions produced or "risk scores" are used by bail judges as a factor in determining a defendant's pretrial release eligibility.[38]

The increased reliance on pretrial algorithms has triggered an intense debate in bail reform circles. Algorithm reformers claim that these algorithms can decrease unwarranted disparities within the pretrial system by enabling bail judges to identify and release low-risk defendants without resorting to the racial and socioeconomic heuristics that have fueled the problem.[39] Moreover, as Chris Slobogin has noted, the promise is that these algorithms could provide jurisdictions with the "quantitative clarity and authority" needed to reduce the incarcerated population.[40] Bernard Harcourt and other

---

36 *See* Itay Ravid & Amit Haim, *Progressive Algorithms*, 12 U.C. IRVINE L. REV. 527, 542–45 (2022) (critiquing the trend of employing algorithms to reform the criminal legal system); Stevenson & Mayson, *supra* note 6, at 23, 30 ("Jurisdictions around the country are now rewriting their pretrial law and policy. They aspire to reduce pretrial detention rates, as well as racial and socioeconomic disparities in the pretrial system, without increasing rates of non-appearance or pretrial crime. . . . To accomplish this, jurisdictions are implementing actuarial risk assessment and reducing the use of money bail as a mediator of release.").

37 I am adopting the definition provided by Mayson, *supra* note 4, at 2228. It is important to note that the risk that a pretrial algorithm is designed to forecast is connected to the enacting jurisdiction's bail law, policy, and practice. For instance, New York City's pretrial algorithm is designed only to gauge non-appearance risk, since the bail regime in the state of New York only detains defendants for non-appearance risk. *See* Crystal S. Yang & Will Dobbie, *Equal Protection Under Algorithms: A New Statistical and Legal Framework*, 119 MICH. L. REV. 291, 358 (2020).

38 *See* Arnett, *supra* note 7, at 651–52.

39 *See* Stevenson & Mayson, *supra* note 6, at 23, 30; Yang, *supra* note 3, at 1401–04 (2017). Discussing the use of these tools in sentencing and policing, Chris Slobogin has argued that risk assessment tools are preferable to unstructured judgment since they can provide more transparent, accurate, and consistent conclusions on risk. *See* Christopher Slobogin, *Principles of Risk Assessment: Sentencing and Policing*, 15 OHIO ST. J. CRIM. L. 583, 586 (2018).

40 CHRISTOPHER SLOBOGIN, JUST ALGORITHMS: USING SCIENCE TO REDUCE INCARCERATION AND INFORM A JURISPRUDENCE OF RISK 158–59 (2021) ("Risk and needs assessment instruments are crucial tools for pinpointing the hundreds of

algorithm critics, in contrast, worry that these algorithms threaten to reproduce inequities under the veneer of scientific objectivity.[41] And these concerns have begun to bear fruit. Recent studies have shown that these algorithms have maintained racial disparities within the pretrial systems of enacting jurisdictions, even while reducing overall incarceration.[42] This is because these algorithms tend to produce biased predictions that maintain the overincarceration of poor and racially marginalized people within the pretrial system.[43]

---

thousands of arrestees and offenders who can, with relative safety, be diverted to community programs or be released with no restrictions. Without the quantitative clarity and authority of these instruments, governments will have neither the wherewithal nor the will to make serious inroads on our incarcerated populations.").

41 Bernard E. Harcourt, *Risk as a Proxy for Race: The Dangers of Risk Assessment*, 27 FED. SENT'G. REP. 237, 237 (2015) ("[R]isk today has collapsed into prior criminal history, and prior criminal history has become a proxy for race. The combination of these two trends means that using risk-assessment tools is going to significantly exacerbate the unacceptable racial disparities in our criminal justice system."); *see also* Roberts, *supra* note 8, at 1699; Eaglin, *supra* note 15, at 487 ("The introduction of sentencing technologies facilitated interpreting those inequities as natural. As such, sentencing technologies reified structural racism under the auspice of scientific objectivity."). Though focused on sentencing technologies, Eaglin's criticism equally applies to pretrial algorithms. Additionally, critics contend that these systems may make no difference in enacting jurisdictions. *See* Starr, *supra* note 33, at 851–52 (contending that there is no persuasive evidence that use of algorithmic systems changes outcomes at all).

42 The technology's use has corresponded with a decrease in an enacting jurisdiction's pretrial detainee population, yet that decrease has not altered the disproportionate percentage of Black defendants in pretrial incarceration. The experience in New Jersey is illustrative. *See* GLENN A. GRANT, ADMIN OFF. OF THE CTS., 2018 REPORT TO THE GOVERNOR AND THE LEGISLATURE 4, 26–27 (2019), https://njcourts.gov/courts/assets/criminal/2018cjrannual.pdf?c=DSE [https://perma.cc/2W3V-WUDU] (noting that the racial percentage of Blacks in jail in 2018 was the same as 2012. However, between this period, the Latinx population declined by 2% and the white population increased by 2%.). In 2017, New Jersey implemented PSA statewide, overhauling its cash-based system in favor of a system of detention based on risk. *See id.* at 3, 49. The switch was lauded as a success, with the pretrial population falling by 19% in its first year and by 13% in its second year. *See id.* at 38. However, the racial makeup of the New Jersey prison population has remained constant, despite the reform. *See id.* at 7.

43 The 2016 ProPublica-Northpointe debate is illustrative. In their study on the use of COMPAS in the bail hearings of 7,000 defendants in Broward County, Florida, ProPublica compared the risk classification that the algorithm assigned each defendant with their actual committance of pretrial crime within the two years following their bail hearing. *See* Angwin, Larson, Mattu & Kirchner, *supra* note 8. ProPublica researchers concluded that COMPAS was racially biased, after finding that it erroneously flagged Black defendants as at high risk for pretrial crime more often than white defendants, who were correspondingly mistakenly flagged as at low risk for pretrial crime compared to Black defendants. *See id.* Northpointe defended itself by emphasizing COMPAS's compliance with the

The next section of this Article introduces the kinds of algorithms being used in the pretrial process and details their construction.[44] Its aim is to show that developers only draw data from one category of knowledge sources when constructing these systems: carceral knowledge sources. By showing this exclusive reliance on carceral knowledge sources within algorithmic construction, this section presents the data source selection problem and sets the stage for the discussion in Part II around how current approaches to algorithmic discrimination have failed to account for and redress this problem.

B. The Construction of Pretrial Algorithms

Though risk prediction has a long and controversial history in the criminal legal system, it is a relatively new component of pretrial decision-making. Prior to the 1960s, risk assessment algorithms were primarily used to inform sentencing and parole decisions.[45] This changed in the 1970s and 1980s with the introduction of preventative pretrial incarceration, which allowed bail judges to detain defendants posing a danger to public safety pending the outcome of their case.[46] The scheme was first introduced into the federal system with the passage of the 1984 Federal Bail Reform Act,[47] and most states followed suit.[48] The introduction of preventative pretrial incarceration operated to spur the development of data-informed pretrial

metric of predictive parity. *See* WILLIAM DIETERICH, CHRISTINA MENDOZA & TIM BRENNAN, COMPAS RISK SCALES: DEMONSTRATING ACCURACY EQUITY AND PREDICTIVE PARITY 21 (2016), https://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf [http://perma.cc/L7VU-T4BT]. Predictive parity in this context refers to the fact that two defendants labelled with a COMPAS high risk classification committed similar rates of pretrial crime. *See id.* Recent studies have demonstrated the impossibility of achieving both statistical parity and predictive parity in an algorithmic system, resulting in most pretrial algorithms complying only with the metric of predictive parity. *See id.* at 2–3, 9.

44 For an extensive synopsis of the algorithmic construction process, see Eaglin, *supra* note 4, at 67–88; Mayson, *supra* note 7, at 509.

45 *See* Alicia Solow-Niederman, YooJung Choi & Guy Van den Broeck, *The Institutional Life of Algorithmic Risk Assessment*, 34 BERKELEY TECH. L.J. 705, 710–11 (2019).

46 *See id.* at 712–13.

47 18 U.S.C. § 3142(e).

48 *See* John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1740–42 (2018). New York is a notable exception. Bail judges are not allowed to detain a defendant in New York for dangerousness. *See* INSHA RAHMAN, NEW YORK, NEW YORK: HIGHLIGHTS OF THE 2019 BAIL REFORM LAW 8 (2019), https://www.vera.org/downloads/publications/new-york-new-york-2019-bail-reform-law-highlights.pdf [https://perma.cc/8AC6-ADKN].

algorithms.[49] Today's pretrial algorithms differ from their earlier iterations in that they are automated and statistically derived.[50]

Pretrial algorithms bear a strong resemblance to their sentencing counterparts. Both are designed using "actuarial—or—statistical analysis of data-driven observations"[51] about behaviors of defendants, who failed to appear at court or were arrested for a crime while on pretrial release.[52] Developers collect and input these data into a statistical model that identifies correlations between a trait of the defendant and the occurrence of pretrial misconduct, in the form of nonappearance at trial or an arrest for crime.[53] After identifying the traits that are closely correlated with these specific behaviors, developers label such traits as risk factors, which they rely on to create the pretrial algorithm.[54] These risk factors are then ranked in the pretrial algorithm and each risk factor is assigned "a number of points corresponding to how closely it is correlated with the [nonappearance or arrest] in the group data."[55] The risk factors that end up being part of the algorithm are either static risk factors or dynamic risk factors.[56] A static risk factor concerns a trait that a defendant possesses and cannot change.[57] An example of a static risk factor is a defendant's criminal record.[58] A dynamic risk factor is an alterable trait that a defendant currently possesses but could be altered with targeted intervention.[59] An example of a dynamic factor would be drug dependency, which a defendant could cease to have with tailored interventions.[60]

Both public and private entities have occupied the field of pretrial algorithms. The two most used pretrial algorithms, the Public Safety Assessment ("PSA") and the Correctional Offender Management Profiling for Alternative Sanctions ("COMPAS"), are privately developed and owned. PSA was developed by the nonprofit organization Laura and Arnold

---

49 *See* Solow-Niederman, Choi & Van den Broeck, *supra* note 44, at 712–13.
50 Mayson, *supra* note 7, at 508–09.
51 Eaglin, *supra* note 4, at 68.
52 *See* Mayson, *supra* note 7, at 509–510.
53 *Id.* at 509.
54 *See id.*
55 *See id.*
56 *See* Eaglin, *supra* note 15, at 491.
57 *See id.*
58 *See id.*
59 *See id.*
60 *See id.*

Foundation, now known as Arnold Ventures.[61] Arnold Ventures offers the PSA model for free.[62] Several states including New Jersey,[63] Kentucky,[64] and Arizona,[65] and at least fifty-nine counties outside of those states have adopted a version of the PSA that has been validated in their state.[66] COMPAS was designed by Northpointe, a for-profit company, now known as Equivant.[67] COMPAS is used in eleven counties.[68] Some states have developed state specific pretrial risk assessment algorithms. For instance, the Ohio Department of Rehabilitation and Correction and the University of Cincinnati's Center for Criminal Justice Research developed the Ohio Risk Assessment System Pretrial Assessment Tool ("ORAS-PAT") for Ohio.[69] The Virginia Department of Criminal Justice Services developed the Virginia Pretrial Risk Assessment Instrument ("VPRAI"), which is used statewide.[70] Despite being developed specifically for use in a particular state, both algorithms are in use nationally.[71] ORAS-PAT has been adopted by five states and at least forty-eight counties.[72] VPRAI is used in at least forty-three counties outside of Virginia.[73]

Additionally, some jurisdictions have developed algorithms under public-private partnerships. For example, New York City Mayor's Office of Criminal Justice contracted with New York City's Criminal Justice Agency (a nonprofit organization), the University of Chicago's Crime Lab New York, idea42 (a behavioral science firm), and Luminosity (a for-profit

---

61 *See* STAN. L. SCH. POL'Y LAB, PSA, *supra* note 13, at 1.

62 *See* LAURA & JOHN ARNOLD FOUND., PUBLIC SAFETY ASSESSMENT: RISK FACTORS AND FORMULA 4 (2016), https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/PSA-Risk-Factors-and-Formula.pdf [https://perma.cc/7WFR-GY8N].

63 ROGER K. WARREN & SUSAN KEILITZ, NAT'L CTR. FOR STATE CTS., PRETRIAL PREVENTIVE DETENTION 6–7 (2020), https://nvcourts.gov/AOC/Committees_and_Commissions/Evidence/Documents/Reports_and_Studies/Pretrial_Preventive_Detention_White_Paper_4_24_2020/ [https://perma.cc/DKA4-EFB5] (detailing the use of risk assessment for preventive detention).

64 *How Many Jurisdictions Use Each Tool?*, MAPPING PRETRIAL INJUSTICE, https://pretrialrisk.com/national-landscape/how-many-jurisdictions-use-each-tool/ [https://perma.cc/R25D-D573] (last visited Feb. 21, 2021).

65 WARREN & KEILITZ, *supra* note 63, at 10–11.

66 *How Many Jurisdictions Use Each Tool?*, *supra* note 64.

67 *See* STAN. L. SCH. POL'Y LAB, COMPAS PRRS-II, *supra* note 13, at 1.

68 *How Many Jurisdictions Use Each Tool?*, *supra* note 64.

69 *Common Pretrial Risk Assessments*, MAPPING PRETRIAL INJUSTICE, https://pretrialrisk.com/the-basics/common-prai/ [https://perma.cc/P5K2-8VSA] (last visited Feb. 21, 2021).

70 *See How Many Jurisdictions Use Each Tool?*, *supra* note 64.

71 *See id.*

72 *Id.*

73 *Id.*

organization) to modify its pretrial algorithm.[74] The new algorithm is currently in use. Unlike most pretrial algorithms, New York City's algorithm only predicts nonappearance risk, not risk of arrest for pretrial crime.[75]

The following subsection explores the knowledge sources used by developers to create pretrial algorithms. The decision around which knowledge sources are chosen is critical, since the choice impacts important aspects of the algorithm's construction. First, the choice of knowledge sources will influence how developers conceptualize the bad outcome that the algorithm is designed to predict. As Jessica Eaglin has pointed out, developers have to first define a target bad outcome in order to observe which traits correlate to it.[76] If the only knowledge sources used are carceral ones, then developers will define the target bad outcome as a pretrial defendant's non-appearance in court or arrest for pretrial crime without factoring in the devastating, life-altering consequences that pretrial incarceration can impose.[77] Second, the choice of knowledge sources will determine the kind of data with which the algorithm will be built with. The data used will predetermine the type of traits that are found to be statistically correlated with the bad outcome that the developers are targeting.[78] This is because the statistical model used to build the algorithm can only observe the traits of

---

74 LUMINOSITY & U. CHI. CRIME LAB N.Y., UPDATING THE NEW YORK CITY CRIMINAL JUSTICE AGENCY RELEASE ASSESSMENT: MAINTAINING HIGH COURT APPEARANCE RATES, REDUCING UNNECESSARY PRETRIAL DETENTION, AND REDUCING DISPARITY 1–2 (2020), https://www.nycja.org/assets/Updating-the-NYC-Criminal-Justice-Agency-Release-Assessment-Final-Report-June-2020.pdf [https://perma.cc/fxq9-llw2]. For context, New York State passed sweeping legislation in 2019 to curtail the practice of conditioning pretrial release on cash bail for most nonviolent offenses, but the state permits the use of algorithms in the pretrial context. FY 2020 NEW YORK STATE EXECUTIVE BUDGET: PUBLIC PROTECTION AND GENERAL GOVERNMENT ARTICLE VII LEGISLATION 207–08 (2019), https://www.budget.ny.gov/pubs/archive/fy20/exec/artvii/ppggartvii.pdf [https://perma.cc/24W4-SS4B]. This allowance has enabled New York City to continue to use its pretrial algorithm that was first implemented in 2003. *See* LUMINOSITY & U. CHI. CRIME LAB N.Y., *supra*, at 3, 29 (outlining changes to the 2003 CJA implemented in 2019).

75 Yang & Dobbie, *supra* note 37, at 358.

76 *See* Eaglin, *supra* note 4, at 75.

77 *See* Shima Baradaran Baughman, *Costs of Pretrial Detention*, 97 B.U. L. REV. 1, 5–6 (2017) (noting the consequences of pretrial incarceration, including that "[d]etainees are often victims of humiliation, rape, and other violent acts while incarcerated, and they also suffer added anxiety, stress, and a lower quality of life as a result" (footnote omitted)); *see also* Megan T. Stevenson & Sandra G. Mayson, *Pretrial Detention and the Value of Liberty*, 108 VA. L. REV. 709, 717–24 (2022) (describing how the benefits of pretrial incarceration must be considered in light of the severe deprivation of liberty that it inflicts on pretrial detainees).

78 *See* Mayson, *Dangerous Defendants*, *supra* note 7, at 509.

pretrial defendants that the original knowledge sources listed as traits.[79] Third, the combination of the first two consequences will affect the predictive model that is ultimately created and the extent to which the model upholds or dismantles existing inequities in the bail system. For this reason, the kinds of knowledge sources used in algorithmic construction bear on the larger question of algorithmic discrimination.

1. *Data Collection and Carceral Knowledge Sources*

Developers have chosen to develop pretrial algorithms with big data.[80] For example, the PSA model was developed by using a dataset consisting of 750,000 defendants on pretrial release from approximately 300 different jurisdictions across the United States over a ten year period.[81] COMPAS, on the other hand, was the product of an analysis performed on a dataset consisting of 2,831 felony defendants on pretrial release in Kent County, Michigan, over a three year period.[82] State-specific algorithms tend to be created with smaller datasets. For example, VPRAI was developed by using a dataset consisting of 1,971 pretrial defendants arrested while on release in Virginia between July 1, 1998 and June 30, 1999.[83] ORAS-PAT's model was created with a dataset of 1,837 individuals, only 452 of whom were defendants on pretrial release.[84]

The data used in pretrial algorithms originate exclusively from carceral knowledge sources. Prior scholarship has explored how predictive technologies rely on the data produced

---

79 *See id.*

80 The definition of big data is contested. This Article uses the term as used in Karen Yeung, *'Hypernudge': Big Data as a Mode of Regulation by Design*, 20 INFO., COMMC'N & SOC'Y 118, 119 (2017) ("Big Data is essentially shorthand for the combination of a technology and a process. The technology is a configuration of information-processing hardware capable of sifting, sorting and interrogating vast quantities of data very quickly. The process involves mining data for patterns, distilling the patterns into predictive analytics and applying the analytics to new data." (citation omitted)).

81 STAN. L. SCH. POL'Y LAB, PSA, *supra* note 13, at 1.

82 STAN. L. SCH. POL'Y LAB, COMPAS PRRS-II, *supra* note 13, at 1.

83 MARIE VANNOSTRAND & KENNETH J. ROSE, PRETRIAL RISK ASSESSMENT IN VIRGINIA 7 (2009), https://www.dcjs.virginia.gov/sites/dcjs.virginia.gov/files/publications/corrections/virginia-pretrial-risk-assessment-report.pdf [https://perma.cc/Q9SZ-87JN].

84 EDWARD LATESSA, PAULA SMITH, RICHARD LEMKE, MATTHEW MAKARIOS & CHRISTOPHER LOWENKAMP, CREATION AND VALIDATION OF THE OHIO RISK ASSESSMENT SYSTEM 13–14 (2009), https://www.ocjs.ohio.gov/ORAS_FinalReport.pdf [https://perma.cc/L9LX-GDJ4].

by criminal legal institutions.[85] For instance, Rashida Richardson, Jason Schultz, and Kate Crawford have identified the data derived from policing and police officers' conduct as the primary data used in predictive policing systems.[86] But criminal legal institutions are not merely data creators. Criminal legal institutions are also carceral knowledge sources that function as hegemonic sites for the production and validation of knowledge. One of their purposes is to produce knowledge about public safety that drives criminal legal policies, practices, and outcomes.[87] And like all knowledge, the knowledge produced and validated by these sources is inherently incomplete and is constructed in relation to the political, economic, and social standpoints of its creators.[88] For this reason, the data derived from the knowledge produced by these sources tends to reflect the status quo and taken for granted assumptions around the relationship between public safety and its promotion through incarceration.[89] A key discourse in the pretrial context supported by the data from these sources is the defining of public safety in relation to the defendant's flight risk or crime risk while excluding other equally important components of public safety, such as securing the safety of that defendant, their family, and their community.[90] For this reason, data from carceral knowledge sources has promoted the incarceration of defendants whose incarceration produces more harms than benefits.[91] In recent decades, social movements and grassroot organizations have

---

85 *See, e.g.*, Ferguson, *supra* note 9, at 1146–47 (discussing how predictive policing technologies are built from police produced data on crime).

86 *See* Richardson, Schultz & Crawford, *supra* note 17, at 21.

87 *See* LIAT BEN-MOSHE, DECARCERATING DISABILITY: DEINSTITUTIONALIZATION AND PRISON ABOLITION 116 (2020). It is important to note that the notion that incarceration protects "public safety" comes from the fact that technocrats have tended to define public safety in the narrow terms of "freedom from injury to one's person and to one's property, in particular from violent crime or events" despite the fact that this definition excludes consideration of how the criminal legal system and other oppressive structures render marginalized groups politically, socially, and economically insecure and unsafe. Barry Friedman, *What Is Public Safety*, 102 B.U. L. REV. 725, 728 (2022).

88 *See* PATRICIA HILL COLLINS, BLACK FEMINIST THOUGHT: KNOWLEDGE, CONSCIOUSNESS, AND THE POLITICS OF EMPOWERMENT 251–52 (2d ed., 2000) (noting that knowledge production practices are not neutral and instead are shaped by the intersecting privilege or oppression that the producer experiences in society).

89 *See* BEN-MOSHE, *supra* note 87, at 116.

90 *See* I. India Thusi, *Policing is Not a Good*, 110 GEO. L.J. ONLINE 226, 226, 230–31, 233 (2022) (noting how the current conceptualization of public safety in the criminal legal system fails to account for the destabilizing impact that criminal legal institutions have on poor and racialized communities).

91 *See* Yang, *supra* note 3, at 1417 (noting that pretrial release and detention decisions fail to account for the individual or social costs of pretrial incarceration).

opposed the prevailing assumptions produced by carceral knowledge sources about the public safety benefits of incarceration.[92] Drawing on the lived experiences of communities most harmed by the criminal legal system, these movements and organizations have produced and validated their own knowledge about public safety that rejects its protection through mass incarceration, criminalization, and surveillance.[93] The data derived from these non-carceral knowledge sources, which are discussed in further detail in Part III, are not used in algorithmic construction. The following subsections will discuss in depth the carceral knowledge sources used to construct pretrial algorithms, which are: the police, pretrial services agencies, and the court system. Its aim is to explore the kinds of data that these sources produce.

a. *The Police*

The police are a critical knowledge source for the construction of pretrial algorithms. Nearly all pretrial algorithms are developed using data generated by the police. The use of police data is the reason for which most pretrial algorithms use arrests as a risk factor. For instance, COMPAS takes into account a defendant's prior arrests in its prediction of a defendant's crime risk.[94] The PSA, on the other hand, does not factor in a defendant's prior arrests but does account for some of the information that police collect on arrest, such as a defendant's age at arrest.[95] This information is used by the algorithm to estimate a defendant's risk for pretrial misconduct.[96] There are two reasons why the use of police data has become so commonplace in algorithmic construction. First, the police generate large amounts of data about pretrial defendants. This is because, as Rashida Richardson, Jason Schultz, and Kate Crawford note, most individuals enter the criminal legal system through policing.[97] Second, the data generated from policing influences bail determinations, since

---

92 *See* Jocelyn Simonson, *Bail Nullification*, 115 MICH. L. REV. 585, 589–91, 634 (2017) (noting that community bail funds can better assess the public safety implications that the pretrial incarceration of a community member poses to their community).

93 *See* Amna A. Akbar, *Toward a Radical Imagination of Law*, 93 N.Y.U. L. REV. 405, 408, 426–28 (2018).

94 *See* STAN. L. SCH. POL'Y LAB, COMPAS PRRS-II, *supra* note 13, at 2.

95 *See Frequently Asked Questions*, ADVANCING PRETRIAL POL'Y & RSCH., https://advancingpretrial.org/appr/faq/#what-psa [https://perma.cc/2MHV-XAA5] (last visited Feb.]21, 2021).

96 *See id.*

97 Richardson, Schultz & Crawford, *supra* note 17, at 23.

they are relied upon by other criminal legal actors, such as prosecutors, pretrial services agents, and bail judges.[98]

The police produce two kinds of data. The first kind is produced during police interactions and encounters that do not lead to an arrest.[99] In this set of circumstances, individuals are stopped by the police and asked a series of questions designed to elicit a variety of information about the stopped individual.[100] This information can include their name, their address, their state identification, and/or their geographical location in which they are found.[101] These stops can also involve searches, which themselves provide the police with additional information about the individual.[102] Much of this information is rendered into data that are then logged into police databases that are maintained on either the local, state, or national level.[103] These data are readily accessible to all parts of the criminal legal system and is often released in an anonymous and aggregated form to the public, different levels of government, public entities, and algorithm developers.[104]

Arrests are the second kind of data produced by the police. The record of a person's arrest is dutifully inputted into police databases and shared among other parts of the criminal legal system, even if that arrest does not lead to charges, prosecution, or a conviction.[105] In this set of circumstances, police typically gather information about the arrested individual as well as the time, place, and the alleged crime.[106] Arrest data are the largest subset of data produced by the

---

98 *See id.*

99 *See* Wayne A. Logan & Andrew Guthrie Ferguson, *Policing Criminal Justice Data*, 101 MINN. L. REV. 541, 556–58 (2016).

100 *See id.* at 556.

101 *See, e.g.*, MARIE PRYOR, FARHANG HEYDARI, PHILIP ATIBA GOFF & BARRY FRIEDMAN, COLLECTING, ANALYZING, AND RESPONDING TO STOP DATA: A GUIDEBOOK FOR LAW ENFORCEMENT AGENCIES, GOVERNMENT, AND COMMUNITIES 38, 52–54 (2020), https://policingequity.org/images/pdfs-doc/COPS-Guidebook_Final_Release_Version_2-compressed.pdf [https://perma.cc/9HHG-7W3T] (listing data that California agencies must collect during stops); Emma Pierson et al., *A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States*, 4 NATURE HUM. BEHAV. 736, 737 (2020) (compiling and analyzing data from traffic stops conducted across the United States).

102 *See* PRYOR, HEYDARI, GOFF & FRIEDMAN, *supra* note 101, at 14.

103 *See* Pierson et al., *supra* note 101, at 736.

104 *See* Logan & Ferguson, *supra* note 99, at 549, 549 n.30 (citing Daniel J. Steinbock, *Designating the Dangerous: From Blacklists to Watch Lists*, 30 SEATTLE L. REV. 65, 69–77 (2006)).

105 *See id.* at 557–58, 565–66.

106 *See* David Eitle, *The Influence of Mandatory Arrest Policies, Police Organizational Characteristics, and Situational Variables on the Probability of Arrest in Domestic Violence Cases*, 51 CRIM. & DELINQ. 573, 580 (2005).

police. For instance, in 2018 alone, law enforcement agents arrested over 10.3 million people.[107]

The data produced by the police are often conceptualized as neutral and objective. There are problems with this conceptualization. First, data produced by the police are often incomplete.[108] As Andrew Ferguson has noted, most crime is not reported and is not caught by the police.[109] Second, police can produce erroneous data. For example, as Anna Roberts has noted, arrests can be wrongfully produced.[110] Third, police data can be the product of illegal practices. Rashida Richardson, Jason M. Schultz, and Kate Crawford have discussed how arrests that are the product of unethical and unconstitutional practices are kept in police databases alongside arrests produced by legal conduct.[111] Moreover, since policing is concentrated in low-income communities of color, policing data are often the product of racially and socioeconomically disparate policing practices that are unethical and inconsistent with the public safety goals of crime control.[112] Since racially marginalized people are particularly vulnerable to being stopped by the police, they are overrepresented in national, state, and local police databases.[113] Racially marginalized people are also disproportionately arrested as compared to their white counterparts. An example of such racial and class disproportionality concerns drug crimes. Even though white and Black people possess drugs at a similar rate, Black people are disproportionately arrested for drug possession in comparison to their white counterparts.[114] To compound the issue, Wayne Logan and Andrew Ferguson's work has shown that, despite the errors in police databases, there exists no viable framework to correct data that are incomplete;

---

107 *2018 Crime in the United States: Persons Arrested*, FED. BUREAU OF INVESTIGATION, https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/topic-pages/persons-arrested [https://perma.cc/KAF4-Y9FA] (last visited Feb. 21, 2021).

108 *See* Ferguson, *supra* note 9, at 1146–47.

109 *See id.*

110 *See* Roberts, *Arrests*, *supra* note 14, at 990–94 (noting that arrests do not necessarily equate to factual guilt).

111 *See* Richardson, Schultz & Crawford, *supra* note 17, at 20, 49–54.

112 *See* Thusi, *supra* note 90, at 241–43 (noting how policing in racially marginalized communities produces harms that undermine public safety in these communities).

113 *See* Pierson et al., *supra* note 101, at 737.

114 *See* Shima Baradaran, *Race, Prediction, and Discretion*, 81 GEO. WASH. L. REV. 157, 189–90 (2013) (finding that "[B]lack defendants are more often arrested for drug crimes even though all races commit drug crimes equally").

inaccurate; and/or derivative of unconstitutional, racialized, and/or classist practices.[115] The combination of these three problems distorts the reliability and validity of the data produced by policing. Nevertheless, such data remain important to algorithmic construction.

b. *Pretrial Services Agencies*[116]

Pretrial services agencies are another crucial source of data for pretrial algorithms, since their mission is to collect data about defendants in order to influence bail determinations.[117] These agencies are charged by a jurisdiction with collecting data about charged defendants. This collected data pertains to the defendant's address, employment status, prior arrests and convictions, pending charges, length of residence in the geographical region, drug use, mental health history, familial situation, community ties, and drug and alcohol use.[118] In some jurisdictions, pretrial services agents are also required to perform a risk assessment on the defendant using the pretrial algorithm in use in their jurisdiction.[119] These data collected by pretrial services agencies are inputted into a pretrial report that is used by bail judges.[120] It is also maintained in pretrial databases that are

---

115 *See* Logan & Ferguson, *supra* note 99, at 585–91.

116 It is important to note that pretrial services agencies were originally developed for a decarceration purpose. They have their origins in the Manhattan Bail Project, launched by the Vera Foundation (now Vera Institute of Justice) and New York University in 1961, which was designed to provide courts with information about a defendant's ties in a community in an effort to secure pretrial release without bond for defendants with strong community ties. *See* Donna Makowiecki, *U.S. Pretrial Services: A Place in History*, 76 FED. PROBATION 10, 10–11 (2012).

117 However, modern pretrial services agencies tend to operate as an arm of the carceral state.

It is important that these agencies are also called bureaus or departments.

118 Cynthia E. Jones, *"Give Us Free": Addressing Racial Disparities in Bail Determinations*, 16 N.Y.U. J. LEGIS. & PUB. POL'Y 919, 923 (2013).

119 *See, e.g.*, *Release Assessment*, N.Y.C. CRIM. JUST. AGENCY, https://www.nycja.org/release-assessment [https://perma.cc/U29J-2PH3] (last visited Feb. 22, 2021).

120 It is important to note that pretrial reports are carceral data, even though they are in part generated from the information that pretrial services agents collect from defendants. The reason for this relates to the process by which these reports are produced. First, pretrial reports are the product of a curated set of questions that agents pose to defendants. These questions are designed to elicit information about each defendant's risk profile, a profile that is itself a product of a carceral way of knowing about public safety and its protection with incarceration. For this reason, agents do not ask questions that would facilitate an assessment about the extent to which pretrial incarceration may affect the safety of a particular defendant, their family, or their community. Second, to compound the issue, the answers provided by defendants are filtered through and

accessible to other parts of the government and are routinely released to algorithm developers in an aggregated and anonymous form.

The data produced by pretrial services agencies are often held out as objective and neutral. However, the data has the same problems that affect the accuracy, reliability, and validity of police data. First, because pretrial services agents are under time pressures and resource constraints, they are inadvertently liable to commit errors about the collected information. These errors undermine the accuracy and completeness of the data kept in their databases.[121] Second, the racial and socioeconomic biases held by individual pretrial services agents can impact the data produced. One recent study found that pretrial services agents are more likely to label Black defendants as having an unstable family situation than white defendants during the screening stage.[122] Though these issues should warrant deeper scrutiny around the use of these data, developers routinely draw on these data without auditing them for accuracy, completeness, or bias.

c. *The Court System*

The data produced by the court system features prominently in algorithmic construction. Courts produce two kinds of data that developers draw upon in the construction process. Convictions are one kind. They are produced by the court system following a guilty verdict or an accepted plea deal.[123] Regardless of how the conviction is produced, convictions are shared with and inputted into databases

---

interpreted by agents. This interpretive process ensures that answers by a defendant that are inconsistent with or irrelevant to a carceral way of knowing about public safety are omitted from the ultimate report. An example of this omitting occurs in relation to a defendant's caretaking responsibilities. The fact that a defendant is the primary custodial parent to a minor child is rarely mentioned in pretrial reports. The final reason for which these reports are carceral concerns the subordinated position of the defendant in relation to a pretrial services agent. Though involved in the generation of these reports, defendants are not conceptualized as knowledge producers whose knowledge should contribute to the dominant knowledge production and validation processes around public safety. Instead, defendants are viewed as objects of inquiry, whose knowledge is useful to the extent that it conforms with the governing paradigm in the pretrial system. The combination of these three factors disqualifies pretrial reports from being non-carceral data.

121 Jones, *supra* note 118, at 943.

122 *See* Traci Schlesinger, *Racial Disparities in Pretrial Diversion: An Analysis of Outcomes Among Men Charged with Felonies and Processed in State Courts*, 3 RACE & JUST. 210, 211 (2013).

123 *See* Roberts, *Convictions*, *supra* note 14, at 2511.

throughout the criminal legal system. Convictions, such as those for sexual assault, are also included in administrative databases at the local and state levels.[124] Bench warrants are another kind of data generated by the court system. Judges issue bench warrants when a defendant fails to appear for a court hearing. Judges have significant discretion in determining whether to issue a bench warrant for a specific defendant.[125] Both kinds of data are procured in an aggregated and anonymous form by developers either from the court system itself or from police or other agency databases.

As with data from the police and pretrial services agencies, the data produced by the court system are not infallible. First, convictions can be wrongfully produced, a problem that disproportionately occurs for racially and socioeconomically marginalized people.[126] Second, bench warrant data can also be produced wrongfully, since judges may issue bench warrants based on incorrect or unreliable information.[127] Moreover, it is important to think about this issue in light of the system by which courts keep track of a defendant's appearance record. Not only do court appearance records suffer from inaccuracies, system actors, as Lauryn Gouldin has problematized, rarely provide the reason behind a defendant's non-appearance at a specific hearing. This omission is important, because, as Gouldin notes, it renders it impossible to glean from the data which defendants engaged in pretrial flight and which ones had innocent reasons for failing to appear at court and had no intention to flee.[128] This leads to a conflation of non-appearance and flight, which ultimately hampers the predictive validity of algorithms built with this data. The combination of these problems causes algorithms to poorly assess the flight risk of defendants, particularly poor and racialized defendants.[129]

d. *Carceral Knowledge Sources and Algorithmic Discrimination*

The exclusive reliance on carceral knowledge sources is an important component to the problem of algorithmic discrimination. First, since these sources tend to produce

---

124 Logan & Ferguson, *supra* note 99, at 553.
125 Gouldin, *supra* note 7, at 690–96.
126 Roberts, *Convictions*, *supra* note 14, at 2509–10.
127 Gouldin, *supra* note 7, at 680–81.
128 *Id.* at 686–89.
129 *See id.* at 710.

incomplete, inaccurate, and racially and socioeconomically disparate data about the individuals subjected to the power of the criminal legal system, their exclusive use in algorithmic construction creates algorithms posed to produce existing inequities. Second, the exclusive reliance on these sources departs from the practice of bail judges. Even though bail judges are required to account for data produced by carceral knowledge sources, they are permitted to draw upon data from other knowledge sources. Bail judges routinely factor in data from non-carceral knowledge sources when determining pretrial release eligibility. For example, studies have shown that bail judges have historically drawn upon qualitative data or their own personal experiences to count the fact that a woman is the primary caregiver of a minor child as a factor in favor of pretrial release.[130] By departing from the practice of drawing on other knowledge sources, developers have facilitated the creation of algorithms that do not include mitigating factors, an exclusion that operates to maintain the current inequities around incarceration by concentrating it among poor and racially marginalized defendants. For these reasons, the exclusive use of carceral knowledge sources operates as a barrier to the creation of pretrial algorithms that could produce outcomes in line with racial and socioeconomic justice in the pretrial system.

II
PROBLEMS WITH CURRENT APPROACHES TO ALGORITHMIC DISCRIMINATION

The exclusive use of carceral knowledge sources is the reason for which algorithmic construction suffers from a data source selection problem. The failure to utilize non-carceral knowledge sources has destined algorithms created under the current paradigm to maintain the racial and socioeconomic status quo in the pretrial system. This Part lays out the biased data diagnosis for algorithmic discrimination. It shows how this diagnosis and current proposals in response have failed to address the data source selection problem. By surfacing this failure, I demonstrate that approaches emanating from the biased data diagnosis are not equipped to fully redress the problem of algorithmic discrimination.

---

130 *See, e.g.*, K.B. Turner & James B. Johnson, *The Effect of Gender on the Judicial Pretrial Decision of Bail Amount Set*, 70 FED. PROB. 56, 57–58 (2006).

This Part is organized into three subparts. Subpart A discusses the growing recognition around the problem of algorithmic discrimination. Subpart B explores the biased data diagnosis. Subpart C discusses the inability of this diagnosis to contend with the data source selection problem by looking at three approaches developed in response to it: (1) the better data approach, (2) the technical adjustment approach, and (3) the auditing approach. By unpacking the limits of the biased data diagnosis, the intent is not to suggest that the resolution of the data source selection problem will provide a complete solution to algorithmic discrimination. As discussed in Part III, fixing the data source selection problem does not guarantee a particular outcome. Pretrial algorithms developed with non-carceral knowledge sources may still produce racially and socioeconomically biased outputs in the bail setting. However, addressing the data source selection problem is a precondition for opening up the possibility for the development of algorithms that produce outcomes that could be in line with redressing racialized mass pretrial incarceration. Given this, the aim of this Part is to highlight how current approaches to algorithmic discrimination cannot counter or fix the data source selection problem and its ensuing consequences.

A. Consensus Around Algorithmic Discrimination

Before discussing the biased data diagnosis and current approaches to algorithmic discrimination, it is necessary to note the consensus around resolving the problem of algorithmic discrimination. When algorithmic technologies were in their infancy, there had been a contentious debate about whether algorithms could discriminate. Some touted the technology as neutral, objective, and a substantial improvement to human decision-making, which was susceptible to subjectivity, bias, and irrationality.[131] However, as governmental processes have become increasingly subjected to algorithmic decision-making, resolving the problem of algorithmic discrimination has taken on a renewed

[131] O'NEIL, *supra* note 9, at 86–89, 197 (problematizing prior conceptualizations of algorithms as being more objective or less biased than human decisionmakers); Kate Crawford, *Think Again: Big Data*, FOREIGN POL'Y (May 10, 2013), https://foreignpolicy.com/2013/05/10/think-again-big-data [https://perma.cc/2HVM-SGHY] (problematizing the idea that big data is objective); Ajunwa, *supra* note 15, at 1685–86 (disputing the notion of data objectivity).

importance.[132] It has become increasingly apparent that the biases in algorithmic systems can lead to ruinous outcomes for individuals and can usher historic biases into the future, thereby reproducing and entrenching existing inequities.[133] One reason for this is that, as Danielle Citron notes, automation bias can lead decision-makers to defer and give effect to the outcomes produced by these systems.[134] Moreover, as Sonia Katyal observes, the scale of harm that can be caused by biased algorithms vastly outstrips the harm of an individual human decision-maker, since these systems tend to apply uniformly to an entire jurisdiction.[135] These problems are not easily correctable, since these algorithmic systems tend to evade private and public accountability mechanisms.[136] This growing recognition has fueled a whole category of scholarship as technocrats struggle to comprehend—let alone resolve—the problem of algorithmic discrimination.[137] A prominent feature of this scholarship has been the biased data diagnosis, which is discussed in the following section.

B. Biased Data Diagnosis

As efforts to redress algorithmic discrimination accelerate, many have coalesced on the biased data diagnosis for algorithmic discrimination. The biased data diagnosis centers on the data currently used in algorithmic systems. It holds that predictive technologies reproduce the biases contained in the data used for their construction.[138] Since algorithmic

---

132 Mayson, *supra* note 4, at 2223 (noting that there has been a proliferation of articles about algorithmic discrimination and its redress).

133 EUBANKS, *supra* note 8, at 141–42, 152–54 (noting the adverse racial and socioeconomic impact that algorithms have in the child welfare context); SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM 12–15 (2018) (noting how the biased algorithms used by online search engines cause societal harms on racial and socioeconomic grounds).

134 Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1271–72 (2008).

135 *See* Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV 54, 69 (2019).

136 *Id.* at 98 (stressing the need to develop private accountability mechanisms to regulate algorithms); Madalina Busuioc, *Accountable Artificial Intelligence: Holding Algorithms to Account*, 81 PUB. ADMIN. REV. 825, 827, 833–34 (2020) (exploring various public accountability mechanisms that can govern algorithms).

137 *See* Thomas B. Nachbar, *Algorithmic Fairness, Algorithmic Discrimination*, 48 FLA. ST. U. L. REV. 509, 509–10 (2021); Yang & Dobbie, *supra* note 37, at 291–92; Ignacio N. Cofone, *Algorithmic Discrimination Is an Information Problem*, 70 HASTINGS L. J. 1389, 1389 (2019); Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Discrimination in the Age of Algorithms*, 10 J. LEGAL ANALYSIS 113, 113–14 (2019); Ajunwa, *supra* note 15, at 1686.

138 Mayson, *supra* note 4, at 2294–95.

systems are built and trained with incomplete and unrepresentative data that have been shaped by societal inequities, they generate forecasts that reflect these inequities.[139] In other words, "bias in, bias out."[140]

The biased data diagnosis has encouraged technocrats to focus on the data currently used in algorithmic systems. This focus has produced useful insights into how problems in datasets employed in algorithmic construction can cause algorithmic systems to disproportionately produce inaccurate and biased predictions about racially marginalized, socioeconomically disadvantaged, and other marginalized groups in society.[141] There are three kinds of problems that the biased data diagnosis has surfaced. The first kind of problem occurs when an algorithm is constructed with data containing errors and omissions. As some have noted, errors and omissions can lead to distortions in the dataset that can hamper the accuracy of the algorithm's predictions.[142] For instance, incorrect or incomplete data about a specific racial group can cause an algorithm reliant on such data to produce faulty predictions about that group.[143] This kind of problem can be mitigated through the utilization of more representative and complete datasets, which is discussed in more detail in Part II.C.1.

The second kind of problem occurs when an algorithm uses data from a knowledge source that has no internal mechanism to ensure that the data is not tainted by the discriminatory practices of institutional actors at that knowledge source. As Sonia Katyal notes, an algorithm can produce biased results when it is built on data that is reflective of the structural discriminatory practices at the knowledge source from which the data originated.[144] For instance, the study conducted by Rashida Richardson, Jason Schultz, and

---

139 Roberts, *supra* note 8, at 1713 ("[T]heir forecasts of the future are based on data that were produced by existing racial discrimination in systems such as policing, housing, education, health care, and public assistance. The future predicted by today's algorithms, therefore, is predetermined to correspond to past racial inequality.").

140 Kristian Lum & William Isaac, *To Predict and Serve?*, 13 SIGNIFICANCE 14, 16, 18 (2016) (describing and demonstrating how algorithmic outputs reinforce bias); Mayson, *supra* note 4, at 2224.

141 Katyal, *supra* note 135, at 75.

142 *See* KATE CRAWFORD ET AL., THE AI NOW REPORT: THE SOCIAL AND ECONOMIC IMPLICATIONS OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN THE NEAR-TERM 6–7 (2016), https://ainowinstitute.org/AI_Now_2016_Report.pdf [https://perma.cc/2NVS-WKLU].

143 Katyal, *supra* note 135, at 124.

144 *Id.* at 68.

Kate Crawford detailed how predictive policing technologies produce racially inequitable results partly because they are trained on data that originated from discriminatory, unethical, and unconstitutional policing practices.[145] This kind of problem can sometimes be mitigated through the adoption of sophisticated auditing practices, which is discussed in more detail in Part II.C.3.

The third kind of problem occurs when an algorithm uses data that are over-representative of a specific racial group. One example of this problem arises in the context of predictive policing technologies. Since policing is concentrated in poor and racially marginalized neighborhoods, systems constructed with data reflecting this reality will produce outcomes advocating for the disproportionate deployment of policing resources into those same communities.[146] The biased data diagnosis has produced two responses to this problem. One has been the call to unearth better data about the traits of those who commit crime.[147] The other response has advocated for technical adjustments to the algorithm to blind the algorithm to this reality or to mitigate the effects caused by the algorithm's reliance on such data. The extent to which this problem can be resolved with technical adjustment is discussed in more detail in Part II.C.2.

In the following section, this Article explores in detail three approaches to algorithmic discrimination emanating from the biased data diagnosis. Though these approaches can partially counter the problems in currently used datasets that fuel algorithmic discrimination, these approaches alone do not offer a complete solution, since they do not attend to the data source selection problem.

---

145 Richardson, Schultz & Crawford, *supra* note 17, at 197.

146 *See* Angèle Christin, *Predictive Algorithms and Criminal Sentencing*, *in* THE DECISIONIST IMAGINATION: SOVEREIGNTY, SOCIAL SCIENCE, AND DEMOCRACY IN THE 20TH CENTURY 272, 279–80 (Daniel Bessner & Nicolas Guilhot eds., 2019) ("When predictive algorithms identify 'hot spot' crime zones (usually low-income African American neighborhoods), policemen are more likely to patrol in these neighborhoods and arrest people who will later be convicted. . . . This data will later be entered into the algorithm, thus producing a feedback loop . . . .").

147 *See, e.g.*, Richard Berk, *Accuracy and Fairness for Juvenile Justice Risk Assessments*, 16 J. EMPIRICAL LEGAL STUD. 175, 186–89 (2019) (discussing how predictiveness of certain data depends on socioeconomic status).

C. Problems with Approaches Based on the Biased Data Diagnosis

As consensus has formed around the biased data diagnosis as the main cause of algorithmic discrimination, three common strategies have emerged to redress the problem.[148] The first is the better data approach. The second is the technical adjustments approach. The third is the auditing approach. These approaches are discussed in detail below.

1. *Better Data Approach*

One advocated approach to redressing algorithmic discrimination is the better data approach, which calls for the use of more complete and representative datasets from the knowledge sources currently used in algorithmic construction.[149] The impetus behind this approach is the view that biased algorithmic predictions are caused by the use of incomplete data about racially marginalized and socioeconomically disadvantaged people. If algorithms were constructed with more accurate, complete, and representative data, advocates of this approach claim that risk assessment algorithms would produce more "accurate" predictions.[150] This would reduce the instances of erroneous high-risk designations that are disproportionally ascribed to racially and socioeconomically marginalized people.

There is good reason to focus on the datasets utilized in algorithmic construction. Inaccurate or inappropriate data can be a source of bias that hampers the predictive validity of an algorithm's predictions for pretrial misconduct.[151] Racially marginalized people are routinely unfairly represented in datasets utilized to construct predictive technologies.[152] Not only does this problem affect the predictive validity of these technologies, it also can exacerbate existing inequities by disproportionately subjecting these groups to inaccurate

---

148 It is important to note that some have advocated for strategies to redress algorithmic discrimination that do not neatly fall into the approaches noted in this Article. For instance, Aziz Huq has argued for a race-based adjustment in cases where factoring in race would "achieve substantively accurate policy results" without producing negative spillover costs in the form of increase crime by way of the erroneous release of "high risk" defendants. Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1101–05 (2019).

149 *See* Pasquale, *supra* note 19, at 1939 (noting that incomplete data as one reason for biased algorithmic assessments).

150 *See, e.g.*, Berk, *supra* note 147, at 175 (noting this perspective); Mayson, *supra* note 4, at 2224–25 (same).

151 *See* Pasquale, *supra* note 19, at 1919.

152 *Id.* 1925

forecasts that justify their oversurveillance, over-criminalization, and overincarceration.

Fixing the biases caused by incomplete, inaccurate, or inappropriate datasets has spurred two sets of interventions. One set is directed at knowledge sources, and the other is directed at developers. The first set of interventions involves creating incentives for knowledge sources to produce, maintain, and release only complete and accurate data. In *Policing Criminal Justice Data*, Wayne Logan and Andrew Ferguson advocate for, among other reforms, state and federal mechanisms to impose legal and practical repercussions on criminal legal actors who systematically fail to identify, cure, and remedy data errors.[153] The second set of interventions consists of requiring developers to utilize complete, accurate, and demographically representative datasets. In *Data-Informed Duties in AI Development*, Frank Pasquale advocates for judicial and regulatory intervention to force developers to engage in practices designed to ensure the use of quality datasets.[154] In Canada, judicial intervention has encouraged better data collection practices. In 2018, the Supreme Court of Canada prohibited the use of VRAG-R, a sentencing algorithm, on Indigenous offenders because the tool had not been empirically validated for Indigenous populations—a problem that caused the algorithm to overestimate the recidivism risk among this already overincarcerated group in the Canadian legal system.[155] In the aftermath of this ruling, the Canadian government awarded a fellowship to a research team led by Mark Olver and Maaike Helmus to redesign the VRAG-R algorithm for use on Indigenous people.[156] The redesign effort consisted of improving the data with which the VRAG-R was designed to include data about Indigenous people. To accomplish this task, the team collected the records of more than 1,000 Indigenous incarcerated people from Correctional Services Canada, the Canadian equivalent of the Federal Bureau of Prisons. The team intends to run this data through a statistical model to ascertain traits statistically relevant to

---

[153] Logan & Ferguson, *supra* note 99, at 600–07.

[154] Pasquale, *supra* note 19, at 1917–19.

[155] Ewert v. Canada, [2018] 2 S.C.R. 165, 204–05 (Can).

[156] Federica Giannelli, *Banting Fellowship Leads to Testing of Criminal Risk Tools*, U. SASKATCHEWAN (July 23, 2018), https://news.usask.ca/articles/research/2018/banting-fellowship-leads-to-testing-of-criminal-risk-tools.php [https://perma.cc/YS9B-349S].

recidivism for this group and to adjust VRAG-R's inputs and risk thresholds accordingly.[157]

There is reason to doubt that efforts to develop more complete, accurate, and demographically representative datasets can provide a complete redress for algorithmic discrimination. First, algorithmic discrimination is not merely a problem of using datasets that under-represent certain groups. For instance, the biased predictions produced by pretrial algorithms in relation to Black people are not caused by the group's under-representation in the data used in algorithmic construction. The cause is that Black communities have been and continue to be overpoliced, which leads to the disproportionate arrest, prosecution, and conviction of this group.[158] This racial inequity is embedded in the data and resultingly produces a problematic feedback loop in which risk predictions reflect this racialized reality rather than the actual public safety risk of Black defendants subjected to these systems.[159] Demographically representative datasets cannot correct this problem, since such datasets cannot counter the myriad of ways in which racism at all levels of the criminal legal system produces data destined to maintain that racism.[160]

Second, and more importantly, this approach to address algorithmic discrimination leaves intact the data source selection problem. This is because the data used under this approach tends to be exclusively from carceral knowledge sources. Since the data produced by carceral knowledge sources are inextricably tied to a racially and socioeconomically disparate conceptualization of public safety, algorithms constructed on such data will inevitably produce predictions consistent with the disproportionate pretrial incarceration of racially marginalized and poor communities. These two problems illustrate the limits of addressing algorithmic discrimination by merely including more data about racially marginalized groups. Though inclusion of racially marginalized people in the datasets will counteract the biases

---

157 *Id.*

158 Michael Pinard, *Race Decriminalization and Criminal Legal System Reform*, 95 N.Y.U. L. REV. ONLINE 119, 120 (2020) ("While the criminal legal system particularly infiltrates the lives of poor people, it is singularly relentless and merciless on Black men, women, and children. It is common knowledge that Black communities have borne the brunt of mass incarceration, mass convictions, and every other aspect of the criminal legal system.").

159 Mayson, *supra* note 4, at 2224–25.

160 Pasquale, *supra* note 19, at 1927 (contending that we may not want to use predictive technologies built on unrepresentative data with regard to members of racially marginalized communities).

caused by incomplete data, it does not overcome the racial biases in the knowledge sources themselves. This approach, thus, inadvertently maintains algorithmic discrimination by leaving intact the data source selection problem.

2. *Technical Adjustment Approach*

Another approach to addressing algorithmic discrimination advocates for technically adjusting algorithms to mitigate the racial and socioeconomic harms caused by their use of biased data. There are numerous approaches that fall under this umbrella. The section below considers two such approaches, both of which demonstrate how such technical fixes maintain the data source selection problem.

a. *Colorblindness*

The first is the colorblind approach, which advocates for the removal of data or inputs that are associated with racially or socioeconomically inequitable outcomes.[161] To accomplish this, advocates propose either extracting race and socioeconomic proxies from the data during algorithmic construction or removing them from the final design of the algorithm.[162] This focus on proxies is important, since the inclusion of racial and socioeconomic proxies can cause algorithms to produce inequitable outputs. For example, zip codes can often serve as a proxy for race. This problem caused the Pennsylvania Commission on Sentencing to remove an offender's county of residence as an input in its sentencing algorithm, since "many commentators viewed [it] as a proxy for race."[163] However, many have pointed out that the removal of race, socioeconomic status, and other protected traits from the

---

161 *See* O'NEIL, *supra* note 9, at 87 (suggesting algorithm developers refrain from using racially, socioeconomically, and otherwise disparate inputs in algorithmic systems).

162 Ion Meyn, *Race-Based Remedies in Criminal Law*, 63 WM. & MARY L. REV. 219, 244 (2021) (citing O'NEIL, *supra* note 9, at 210); Bent, *supra* note 20, at 814–16 (recognizing the "general consensus" that "attempting to blind an algorithm to a sensitive characteristic is usually ineffective"). .

163 PA. COMM'N. ON SENT'G, SENTENCE RISK ASSESSMENT INSTRUMENT 5 (2020), https://pennstateoffice365.sharepoint.com/sites/PCSFileshare/Shared%20Documents/Forms/AllItems.aspx?id=%2Fsites%2FPCSFileshare%2FShared%20Documents%2FHome%2FGuidelines%20and%20Statutes%2FRisk%20Assessment%2FSentence%20Risk%20Assessment%2FSentence%20Risk%20Assessment%20Instrument%20%287%2D1%2D2020%29%2Epdf&parent=%2Fsites%2FPCSFileshare%2FShared%20Documents%2FHome%2FGuidelines%20and%20Statutes%2FRisk%20Assessment%2FSentence%20Risk%20Assessment&p=true&ga=1 [https://perma.cc/G6ZQ-6YUH]. The Pennsylvania Sentencing Commission also rejected using past

data or inputs utilized by an algorithm is an ineffective approach to dealing with algorithmic discrimination.[164] Cynthia Dwork and others have pointed out that the removal of such information does not prevent the statistical model utilized in algorithmic construction from inadvertently discovering and relying on racially or socioeconomically biased correlations that will ultimately produce discriminatory outcomes.[165] Moreover, others have argued that the removal of such information will adversely affect the predictive validity of the algorithm and might counterintuitively produce even less accurate predictions that will disproportionately impact racially marginalized groups.[166]

Another critique of the colorblind approach is that it does not attend to the data source selection problem. It is true that certain inputs and data play a unique role in algorithmic discrimination. The exclusion of some of this information could render algorithms less racially disparate and more accurate. For instance, many have argued that the inclusion of arrest data is racially inequitable and adversely affects the predictive validity of algorithmic systems.[167] However, in surgically focusing on specific egregious proxies, the approach misses the way in which the data produced and validated by the entire criminal legal system—not just arrests—is implicated in current racial inequities. By its nature, this data promotes and validates the disproportionate ascription of dangerousness onto the most vulnerable members of our society. This makes it impossible to disentangle racial proxies from the data produced by these institutions. Given this, the colorblind approach may reduce algorithmic discrimination on

---

arrests entirely as an input variable, because arrests had such different predictive significance across racial lines. *Id.*

164 Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 918 (2017) (arguing that "[i]f the goal is to reduce biased outcomes, then a simple prohibition on using data about race or sex could be either wholly ineffective or actually counterproductive due to the existence of class proxies and the risk of omitted variable bias"). Though Kim makes this argument in the employment context, it still applies in the criminal context as well. *See* Zachary C. Lipton, Alexandra Chouldechova, & Julian McAuley, *Does Mitigating ML's Impact Disparity Require Treatment Disparity?*, 32 CONF. ON NEURAL INFO. PROCESSING SYS. 1–2 (2019), https://arxiv.org/pdf/1711.07076.pdf [https://perma.cc/X8VF-EY53] (demonstrating the accuracy cost associated with designing algorithms to be blind to protected traits).

165 *See* Cynthia Dwork, Nicole Immorlica, Adam Tauman Kalai & Max Leiserson, *Decoupled Classifiers for Fair and Efficient Machine Learning*, 81 PROC. MACHINE LEARNING RSCH. 1, 2 (2018), https://proceedings.mlr.press/v81/dwork18a/dwork18a.pdf [https://perma.cc/VQ7D-GTY4].

166 Mayson, *supra* note 4, at 2262–63.

167 Eaglin, *supra* note 4, at 95.

the margins, but it does not fully contend with its production and entrenchment.

b. *Algorithmic Affirmative Action/Race-Aware Approach*

The second category of technical adjustments to algorithms concerns algorithmic affirmative action or race-aware approaches. This category advocates for the development of racially and socioeconomically aware algorithms to mitigate the effects of using biased data.[168] There are a few ways to accomplish this approach. One would be to include racial and socioeconomic information as inputs with an assigned weight that would impact the risk score produced.[169] Another approach, which has been advocated by Deborah Hellman, would be to create a dual track within the algorithm in which race and socioeconomic status would impact the risk factors to which an individual would be subjected to.[170] For instance, studies have shown that homelessness is predictive of pretrial misconduct in white socioeconomically disadvantaged defendants, but not in Black socioeconomically disadvantaged defendants.[171] Under the dual track model, homelessness would be a factor considered for white defendants, but not for Black defendants. Another approach would be to consider race in the estimation step but not in the prediction step of algorithmic construction, as advocated by Crystal Yang and William Dobbie.[172] The final way would be to not consider race and socioeconomic information in the algorithm but to subsequently adjust the risk score produced by the algorithm to reflect the racial or socioeconomic information of individuals subjected to it. For instance, Ion Meyn has argued for a downward discount to be applied to Black defendants in the assessment of their risk score in order to account for the ways in which systemic racism in the criminal legal system has promoted their overincarceration.[173]

The advantage of the affirmative action approach is that it would improve the racial equity of algorithms by explicitly attempting to counter the racial bias in the data in use. Moreover, it would decrease the instances in which racially and

168 Bent, *supra* note 20, at 805, 817–24; Deborah Hellman, *Measuring Algorithmic Fairness*, 106 VA. L. REV. 811, 852–55 (2020).

169 Bent, *supra* note 20, at 804–05, 807, 820.

170 Hellman, *supra* note 168, at 853.

171 *Id.*

172 Yang & Dobbie, *supra* note 37, at 350–52.

173 Meyn, *supra* note 162, at 247–48.

socioeconomically marginalized groups are subjected to risk scores that are more reflective of systemic inequity than the risk that members of these groups pose to public safety. However, there are important drawbacks. First, it may not be legally feasible to engage in algorithmic affirmative action, but there is no consensus on this point. As Jason Bent notes, some forms of algorithmic affirmative action, specifically the forms most likely to redress the racial and socioeconomic harms of these algorithms, may violate the Equal Protection Clause.[174] Sonja Starr suggests that, "[t]here appears to be a general consensus that using race would be unconstitutional."[175] However, Pauline Kim contends that many race-aware strategies would not trigger strict scrutiny.[176] However, even if the Equal Protection Clause were not a barrier to the use of algorithmic affirmative action, there is still a question about how best to actualize the solution. How much weighing to race and class is necessary to counteract the extent to which they feature in the data currently used to construct algorithmic systems? The answer to this question becomes more complicated when viewed in light of the data source selection problem. Since the entirety of the data utilized is inextricably linked to race and class, is it possible to counteract the totality of this bias by including race and class as inputs? By not counteracting the reliance on racist and classist knowledge sources within algorithmic construction, the entire project of algorithmic affirmative action cannot fully redress algorithmic discrimination.

3. *Auditing Approach*

The third approach concerns data auditing practices. The approach calls for the development of laws or industry practices that require developers to audit data for bias before their use in algorithmic systems. There is good reason to focus on auditing practices. One cause of biased algorithmic predictions is connected to the defective and sometimes non-existent processes that most vendors of predictive technologies

---

174 Bent, *supra* note 20, at 845–48.

175 *See* Starr, *supra* note 33, at 812.

176 Pauline T. Kim, *Race-Aware Algorithms: Fairness, Nondiscrimination and Affirmative Action,* 110 CALIF. L. REV. 1539, 1574–1586 (2022); *see also* Yang & Dobbie, *supra* note 37, at 352–55 (contending that an algorithm designed to factor in race might survive constitutional scrutiny in certain cases).

have for identifying and redressing biases in datasets.[177] To compound the problem, auditing efforts by outsiders have been hampered by the opacity that most vendors have around their data collection processes.[178]

The implementation of standardized auditing practices could mitigate some of the racial and socioeconomic harms caused by biased data. For instance, using the example of predictive policing, Rashida Richardson, Jason Schultz, and Kate Crawford have argued that data auditing practices could encourage, or even require, vendors of these technologies to exclude data tainted with racially and socioeconomically inequitable practices from their datasets.[179] Andrew Ferguson has advocated for the implementation of an independent data auditing system designed to root out biases caused by the inclusion of illegal data in datasets.[180] A related practice concerns the use of algorithmic impact assessments. These assessments involve having the developer investigate and inform a jurisdiction about the adverse impacts that an algorithmic system could have if implemented, such as its potential harms to racially marginalized and poor communities. As many have argued, the use of these assessments can assist jurisdictions with identifying algorithmic systems implicated in the continuation or exacerbation of existing inequities.[181]

The benefit of data auditing approaches is that they have the potential to counteract many facets of algorithmic discrimination. First, these protocols can assist in uncovering and addressing the extent to which data derived from

---

177 *See* Pasquale, *supra* note 19, at 1932 (proposing that "agencies [ ] set standards for AI vendors and users to verify the quality and accuracy of the data they use).

178 Richardson, Schultz & Crawford, *supra* note 17, at 21 (noting this problem in the context of predictive policing technologies); Selbst, *supra* note 21, at 190 & n.401 (noting the challenges opacity poses on the data auditing efforts of third parties in regard to predictive policing). For more information about the opacity surrounding algorithmic construction and implementation, see Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1396–1401 (2018) (discussing the impact that trade secret privilege has on efforts to address opacity in algorithmic construction); Natalie Ram, *Innovating Criminal Justice*, 112 NW. U. L. REV. 659, 664–65 (2018) (discussing the lack of public knowledge about the innerworkings of algorithmic technologies).

179 Richardson, Schultz & Crawford, *supra* note 17, at 41, 48.

180 Ferguson, *supra* note 9, at 1167.

181 *See* DILLON REISMAN, JASON SCHULTZ, KATE CRAWFORD & MEREDITH WHITTAKER, ALGORITHMIC IMPACT ASSESSMENTS: A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY ACCOUNTABILITY 15–20 (2018), https://ainowinstitute.org/aiareport2018.pdf [https://perma.cc/53L3-KTJS].

problematic practices and policies affect algorithmic predictions. Second, these approaches can provide jurisdictions with the information needed to reform or to cease using systems discriminatorily impacting marginalized communities on racial and socioeconomic grounds. However, similar to the problem with the better data approach and the technical adjustment approach, data auditing approaches are incapable of fully redressing algorithmic discrimination, since they leave in place the current system of data collection, including its exclusive reliance on carceral knowledge sources. Unless we grapple with the data source selection problem, we cannot audit our way out of algorithmic discrimination.

Exploring the above approaches demonstrates how the biased data diagnosis and solutions emanating from it are partial solutions to the problem of algorithmic discrimination. This is because the use of biased data is but one cause of algorithmic discrimination. The exclusive use of carceral knowledge sources is another. Attacking biased data without dismantling the dominance of carceral knowledge sources cannot change the role that current pretrial algorithms play in sustaining the carceral state.[182] Moreover, the use of solutions derived from the biased data diagnosis entrenches the false narrative that the problem of algorithmic discrimination is an inevitable byproduct of using data-driven technologies in an inequitable society, whilst obscuring the political choices, such as the reliance on carceral knowledge sources, that maintain the problem.[183] To compound this issue, these solutions also can have the perverse effect of fueling the illusion that algorithmic systems produced with these proposed solutions are fairer than their human counterparts.

In this Part, I critiqued existing approaches to solving algorithmic discrimination as simultaneously ignoring and enabling the data source selection problem and its ensuing consequences. Approaches that attend to biased data and not the sources that produce that data are doomed to preserve algorithmic discrimination. In Part III, I explore the potential of

---

182 It is important to note that the limits of these solutions come from the fact that they do not attend to the dominance of carceral knowledge sources in algorithmic construction. However, these solutions combined with a turn away from the dominance of carceral knowledge sources in algorithmic construction could offer a more complete solution to algorithmic discrimination.

183 Jessica M. Eaglin, *When Critical Race Theory Enters the Law & Technology Frame*, 25 MICH. J. RACE & L. 151, 165 (2021) (problematizing how dominant approaches to algorithmic discrimination often frame the problem as the product of performing risk "prediction in a racially unequal world").

using non-carceral knowledge sources to combat the problem of algorithmic discrimination.

III
TOWARD NON-CARCERAL KNOWLEDGE SOURCES

The inability of existing approaches to correct algorithmic discrimination provides fertile ground to rethink the relationship between the algorithm project and racial and socioeconomic justice. The issue has prompted some to advocate for re-purposing algorithms for only non-carceral uses. For instance, Andrew Ferguson has suggested that predictive policing technologies could be subverted to assess police behavior rather than to justify the overpolicing of marginalized communities.[184] And Benchmark Analytics has completed some work on this front.[185] Sandra Mayson has suggested that risk assessment algorithms could be transformed into a tool to identify defendants for treatment and support during the pretrial process.[186] Sean Hill, Sasha Costanza-Chock, and others have invoked this problem in their call to abolish the use of all predictive technologies in the criminal legal system.[187]

If the only knowledge sources available for algorithmic construction are carceral knowledge sources, then re-purposing or abolishing algorithms warrants serious consideration, since these systems will serve solely to justify the disproportionate subjugation of the most vulnerable members of society to the violence of the carceral state.[188]

---

184 *See* Andrew Guthrie Ferguson, *The Exclusionary Rule in the Age of Blue Data*, 72 VAND. L. REV. 561, 627–32 (2019) (contending that predictive technologies should be used to monitor and to check police behavior).

185 *See, e.g.*, Melissa Fassbender, *UChicago Alums, Researchers Develop Evidence-Based Police Force Management and Early Intervention System*, POLSKY (June 8, 2021), https://polsky.uchicago.edu/2021/06/08/benchmarking-police-performance-for-early-intervention-evidence-based-solutions/ [https://perma.cc/L46T-3JKZ].

186 *See* Mayson, *supra* note 4, at 2288–89.

187 Hill, *supra* note 33, at 986–87 (contending that pretrial algorithms could arguably be banned on the basis that they are incompatible with achieving racial justice aims in the pretrial system); SASHA COSTANZA-CHOCK, DESIGN JUSTICE: COMMUNITY-LED PRACTICES TO BUILD THE WORLDS WE NEED 63 (2020) (noting that "[a] prison abolitionist stance does not support allocating additional resources to the development of tools [such as risk assessments] that extend the [Prison Industrial Complex], even to make them 'less biased'"). *But see* Vincent M. Southerland, *The Intersection of Race and Algorithmic Tools in the Criminal Legal System*, 80 MD. L. REV. 487, 561–62 (2021) (proposing that predictive technologies should be used for abolitionist projects).

188 Jessica M. Eaglin, *Population-Based Sentencing*, 106 CORNELL L. REV. 353, 398 (2021) (contending that currently employed risk assessment tools operate to

However, carceral knowledge sources are not the only sources available for algorithmic construction. The problem has been that developers have exclusively relied on such sources, even when attempting to address the issue of algorithmic discrimination. To solve this problem, we must begin by dismantling the dominance of carceral knowledge sources through shifting our focus to knowledge sources that are not intrinsically linked to the maintenance of the carceral state. One such category of knowledge sources that has escaped the attention of developers is the community knowledge sources relied upon by the communities most affected by the criminal legal system.

At the outset, it is important to note that the turn to community knowledge sources may not necessarily provide a complete solution for algorithmic discrimination. As will be discussed in further detail in Part IV, there are powerful counterarguments about the extent to which data produced by such sources is less biased than that of carceral knowledge sources. But the irredeemably racist and classist quality of data from carceral knowledge sources warrant this exploration given that algorithmic systems will remain incompatible with racial and socioeconomic justice unless they can draw from less biased knowledge sources. Moreover, the explosion of social movements and community organizations that focus on the pursuit of racial and economic justice provide grounds for optimism about the anti-racist nature of community knowledge sources.[189] Tapping into community knowledge sources may enable algorithmic systems to produce outcomes around public safety that are not necessarily rooted in racial and class subordination.

Using the example of community knowledge sources, the following sections explore how shifting to non-carceral knowledge sources may offer a path toward addressing algorithmic discrimination. First, it locates the current non-use of community knowledge sources in epistemic oppression,

---

"entrench mass incarceration as a particular mode of governance that operates to manage and control marginalized populations through the carceral state rather than offer support and resources outside it").

189 *See* Amna A. Akbar, *Law's Exposure: The Movement and the Legal Academy*, 65 J. LEGAL EDUC. 352, 357–60 (2015); Akbar, *supra* note 93, at 426–28; K. Sabeel Rahman, *Policymaking as Power-Building*, 27 S. CAL. INTERDISC. L.J. 315, 340–50 (2018) (discussing this growth in the administrative context); K. Sabeel Rahman & Jocelyn Simonson, *The Institutional Design of Community Control*, 108 CALIF. L. REV. 679, 699–719 (2020) (discussing social movements to regain community control).

which allows developers to undervalue the knowledge of subjugated groups. Second, it lays out the contents of the qualitative data about the criminal legal system produced by currently and formerly incarcerated people from most impacted communities, which is a subset of data derived from community knowledge sources. It concludes by considering the potential advantages of turning to this kind of data.

A. Absence of Community Knowledge Sources in Algorithmic Construction

Despite the racial and socioeconomic inequities caused by exclusive reliance on carceral knowledge sources, developers have tended not to draw upon community knowledge sources in algorithmic construction. One explanation for the preference for carceral knowledge sources is that these sources produce large amounts of data that can be quantified by statistical models and resultingly can capture patterns about defendants on pretrial release.[190] Proponents of this explanation claim building algorithmic systems on such data enables these systems to produce predictions about pretrial misconduct free from human bias or subjectivity.[191] Because community knowledge sources produce qualitative data, the argument goes, the inclusion of such data would adversely impact the objectivity and neutrality of algorithmic systems.[192] At first glance, this explanation seems intuitively correct. Unlike qualitative data, quantitative data lends itself to statistical analytics and enables the discovery of patterns and relationships within a dataset that seem inherently meaningful and free from human intervention.[193]

However, there are three deficiencies with this explanation. First, insights derived from quantitative data are neither

---

190 *See generally* Deven R. Desai & Joshua A. Kroll, *Trust but Verify: A Guide to Algorithms and the Law*, 31 HARV. J.L. & TECH. 1, 23–29 (2017) (noting the preference toward algorithmic systems lies in their ability to synthesize and discover correlations in quantitative datasets); Samuel R. Wiseman, *Fixing Bail*, 84 GEO. WASH. L. REV. 417, 439 (2016) (noting that this feature renders risk assessment algorithms superior to individual decision-making in bail).

191 Rob Kitchin, *Big Data, New Epistemologies and Paradigm Shifts*, 4 BIG DATA & SOC'Y 1, 5 (2014) (problematizing this perspective, noting that the insights that can be gleaned from data are rooted in a particular way of understanding that is connected to "culture, politics, policy, governance and capital").

192 *But see* Monica C. Bell, *The Community in Criminal Justice: Subordination, Consumption, Resistance, and Transformation*, 16 DU BOIS REV. 197, 211 (2019) (arguing that recognizing "community members as co-producers of knowledge" generally would allow technocrats to view these communities' "qualitative data [as] on par with quantitative data").

193 *Id.*

inherently neutral nor objective. As Kimani Paul-Emile has explained, these insights occur within a sociopolitical context.[194] Given this, the insights derived from the datasets currently used in algorithmic construction are not neutral and instead reflect the values and assumptions of the developers tasked with data analysis and interpretation.[195] Second, algorithmic systems are not solely built with quantitative data. Because algorithmic construction requires discretionary and value-laden decisions around how the algorithm will interact with existing law and public policy, developers often formally and informally consult institutional actors such as judges, prosecutors, and defense lawyers.[196] Moreover, though community knowledge sources primarily produce qualitative data, they also produce quantitative data. In fact, in recent years, community groups have engaged in quantitative studies around the use of pretrial algorithms in order to organize around the problem of mass pretrial incarceration.[197] The growth in these studies suggests that community knowledge sources could be mined for their quantitative data in addition to their qualitative data.

The discussion above demonstrates that the exclusion of community knowledge sources cannot be justified as a mere preference for quantitative data in algorithmic construction. A more fitting explanation for this exclusion is that the knowledge produced by community knowledge sources is a type of subjugated knowledge. Subjugated knowledge refers to knowledge that is suppressed and largely unacknowledged outside of the epistemic community that created and validated it.[198] Michel Foucault has explained that the subjugated status of this knowledge is not coincidental and instead is rooted in power. Because of the disempowered status of holders of this knowledge, this knowledge is buried and

---

194 Kimani Paul-Emile, *Foreword: Critical Race Theory and Empirical Methods Conference*, 83 FORDHAM L. REV. 2953, 2959 (2015).

195 Eaglin, *supra* note 4, at 88.

196 For example, the Pennsylvania Sentencing Commission interviewed district attorneys and used the qualitative data from these interviews to construct the algorithm. *See Sentencing Risk Assessment Policy Archives*, PA. COMM'N ON SENT'G, http://pcs.la.psu.edu/guidelines/sentence-risk-assessment-instrument/sentence-risk-assessment-policy-archives [https://perma.cc/Z2U9-43WX] (last visited Feb. 26, 2021).

197 *See, e.g.*, *Mapping Pretrial Injustice: A Community-Driven Database*, MOVEMENT ALL. PROJECT, https://pretrialrisk.com/ [https://perma.cc/2UM7-AXCL] (last visited Feb. 26, 2021).

198 MICHEL FOUCAULT, POWER/KNOWLEDGE: SELECTED INTERVIEWS & OTHER WRITINGS 1972–1977, at 8, 30 (Colin Gordon ed., 1980).

masked by more dominant forms of knowledge.[199] Most impacted communities are but one disempowered group that experiences the subjugation of their knowledge.[200]

A byproduct of this subjugation is that holders of such knowledge experience epistemic oppression. Coined by Miranda Fricker, the term "epistemic oppression" captures the harms experienced by epistemic communities that are excluded from or discounted in the knowledge production and validation processes of the powerful groups in society.[201] One harm is that members of these communities are unable to participate in the crafting of law, policy, and practice. Exploring this problem in the context of prisoners, Eve Hanan has argued that epistemic oppression has allowed judges to ignore critical insights possessed by formally incarcerated people about life in prison.[202] She theorizes that consideration of this discarded knowledge could impact sentencing outcomes by revealing to judges the totality of punishment that imprisonment enacts on an offender.[203] Critical race theorists have explored how race and its intersection with other subordinated statuses promote this denial. Bennett Capers' work has evaluated how Black people are routinely disbelieved in the criminal legal system whether as witnesses, complainants, or defendants.[204] In other words, the subjugation of knowledge by disempowered people allows more powerful groups to stereotype individuals from these communities and to refuse to account for their knowledge.

Another harm of this subjugation, as Miranda Fricker has noted, is that it ensures that powerful groups "have some sort of unfair advantage in 'structuring' our *understandings* of the social world."[205] This unfair influence distorts the range of possibilities achievable, because it entrenches the status quo, even if other ways of knowing could generate better outcomes

---

199 *Id.* at 50–52.

200 Richard Jackson, *Unknown Knowns: The Subjugated Knowledge of Terrorism Studies*, 5 CRITICAL STUD. ON TERRORISM 11, 15 (2012).

201 Miranda Fricker, *Epistemic Oppression and Epistemic Privilege*, 25 CAN. J. PHIL. 191, 191–92 (1999) (theorizing about epistemic oppression's impact on knowledge production processes).

202 M. Eve Hanan, *Invisible Prisons*, 54 U.C. DAVIS L. REV. 1185, 1191 (2020).

203 *Id.* at 1204.

204 *See* Bennett Capers, *Evidence Without Rules*, 94 NOTRE DAME L. REV. 867, 885–93 (2018); *see also* Mikah K. Thompson, *Bias on Trial: Toward an Open Discussion of Racial Stereotypes in the Courtroom*, 2018 MICH. ST. L. REV. 1243, 1246 (2018) (discussing the extent and role of juror bias in verdicts).

205 Fricker, *supra* note 201, at 191, 193 (theorizing about epistemic oppression's impact on knowledge production processes) (emphasis in original).

than currently exist.[206] The actualization of this harm has already occurred in the pretrial context. The subjugation of the knowledge of communities most impacted by the criminal legal system has facilitated the production and utilization of a conceptualization of public safety that does not protect the safety of defendants, their families, and their communities.[207] One reason for this is that the currently employed conceptualization of public safety has been exclusively informed by the knowledge produced and validated by carceral knowledge sources. These knowledge sources have failed to provide a realistic assessment about the benefits of pretrial incarceration, because their knowledge does not account for its harms. For instance, community groups representing formally incarcerated people have long noted the adverse impact pretrial incarceration has on public safety.[208] The traumatization experienced by pretrial detainees decreases their capacity to re-integrate into the communities that they come from, thereby increasing their risk for further violence and crime in these communities. As Dorothy Roberts has noted, large scale incarceration in a specific locality can break informal social control mechanisms that are essential for public safety, as social kinship networks are disrupted and overextended.[209] This state of affairs undermines rather than upholds public safety.

Algorithmic construction did not inaugurate the epistemic oppression experienced by most impacted communities. However, it contributes to this oppression in two important ways. First, by exclusively relying on carceral knowledge sources, it promotes the notion that members of these communities are incapable of contributing to the knowledge base from which algorithms are constructed from. This

---

206 Briana Toole, *From Standpoint Epistemology to Epistemic Oppression*, 34 HYPATIA 598, 611 (2019).

207 Collins, *supra* note 32, at 7 (discussing how the kind of public safety pursued by the criminal legal system "excludes the safety of those most directly impacted by the system itself").

208 *See, e.g.*, NAT'L COUNCIL FOR INCARCERATED & FORMALLY INCARCERATED WOMEN & GIRLS [hereinafter NAT'L COUNCIL], https://www.nationalcouncil.us/ [https://perma.cc/BBF2-3QQD] (last visited Feb. 24, 2021) (discussing the personal tolls of incarceration for women and girls); ENVISION FREEDOM FUND, https://envisionfreedom.org [https://perma.cc/NVP2-JTFV] (last visited Oct. 13, 2022) (discussing incarceration harms).

209 Dorothy E. Roberts, *The Social and Moral Cost of Mass Incarceration in African American Communities*, 56 STAN. L. REV. 1271, 1285 (2004). To understand more about the general effects of incarceration, see Elizabeth J. Gifford, *How Incarceration Affects the Health of Communities and Families*, 80 NCMJ 372 (2019).

concept only further entrenches the epistemic oppression that these communities experience in society at large. Second, the refusal thus far to engage with community knowledge sources bolsters the idea that these communities are not epistemic communities and that their knowledge is not credible. Not only does this harm these communities, but it has also created algorithms that have been developed by defective, racist, and classist knowledge sources that are, as a result, unable to produce racially and socioeconomically ameliorative outcomes.

Tapping into community knowledge sources provides a promising alternative. To be clear, this alternative offers more than a tweak on the current system of pretrial incarceration. Turning to community knowledge sources means engaging with a different vision of public safety that includes defendants, their families, and their communities. Such an engagement has the potential to decouple algorithms as well as the pretrial system itself from their implication in the racial and classist status quo. In the following section, this Article will discuss the contents of community knowledge sources.

B. Content of Data Produced by Community Knowledge Sources

Community knowledge sources contain the knowledge produced and validated by the communities most harmed by criminal legal interventions. This knowledge consists of the individual and collective lived experiences of community members.[210] This call to engage with this kind of knowledge has roots in the critical race theory tradition. Critical race theorists have stressed the importance of centering the lived experiences of those most harmed by the current system in law reform efforts.[211] Mari Matsuda has argued that looking to those at the bottom of the social order can shed light on reforms that have the power to transition us toward a more egalitarian society.[212] Francisco Valdes has explored how those subordinated on various axes of discrimination have

---

210 Bell, *supra* note 192, at 209–11.

211 I. Bennett Capers, *Afrofuturism, Critical Race Theory, and Policing in the Year 2044*, 94 N.Y.U. L. REV. 1, 25–28 (2019) (contending that drawing upon the lived experiences of racially marginalized people can offer radical interventions that can produce true change).

212 Mari J. Matsuda, *Looking to the Bottom: Critical Legal Studies and Reparations*, 22 HARV. C.R.-C.L. L. REV. 323, 324–26 (1987) (contending that the lived experiences of people of color should be understood to be an "epistemological source" that can produce conceptualizations of law, and its possibilities are radically different from those generated from the state).

developed specialized knowledge about systems of power that enable them to recognize, to name, and to remedy racial injuries that are invisible to system beneficiaries.[213] Monica Bell has argued that, "as subordinates of the criminal justice system, members of marginalized communities are especially knowledgeable about systemic injustice and thus especially capable of and responsible for rectifying it."[214] Underlying the call to recognize the lived experiences of racially and socioeconomically marginalized communities as knowledge capable of informing law and policy is the understanding that knowledge production practices of powerful constituencies have tended to maintain and to justify existing inequities in the criminal legal system and beyond.[215]

An important feature of this form of knowledge is that, while it is held both individually and by community knowledge sources, it tends to be confined in racialized pools of knowledge inaccessible to outsiders.[216] An example of this is racialized police brutality. Though Black communities have produced and shared knowledge about racialized police violence for centuries, this knowledge has only recently penetrated dominant ways of knowing.[217] As Bennett Capers has explained, "Such acts of violence are part of our 'pools of

---

[213] Francisco Valdes, *Breaking Glass: Identity, Community and Epistemology in Theory, Law and Education*, 47 U.C. DAVIS L. REV. 1065, 1073–74 (2014); *see also* Maria C. Malagon, Lindsay Perez Huber & Veronica N. Velez, *Our Experiences, Our Methods: Using Grounded Theory to Inform a Critical Race Theory Methodology*, 8 SEATTLE J. SOC. JUST. 253, 257 (2009) (discussing the importance of personal narrative as a methodology); EDWARD W. SAID, COVERING ISLAM: HOW THE MEDIA AND THE EXPERTS DETERMINE HOW WE SEE THE REST OF THE WORLD 157 (1997) (defining the concept of antithetical knowledge as "the kind of knowledge produced by people who quite consciously consider themselves to be writing in opposition to the prevailing orthodoxy"); RICHARD DELGADO & JEAN STEFANCIC, CRITICAL RACE THEORY: AN INTRODUCTION 10 (2d ed. 2012) ("Coexisting in somewhat uneasy tension with antiessentialism, the voice-of-color thesis holds that because of their different histories and experiences with oppression, [B]lack, American Indian, Asian, and Latino/a writers and thinkers may be able to communicate to their white counterparts matters that the whites are unlikely to know. Minority status, in other words, brings with it a presumed competence to speak about race and racism.").

[214] Bell, *supra* note 192, at 208.

[215] Valdes, *supra* note 213, at 1078–79.

[216] Russell K. Robinson, *Perceptual Segregation*, 108 COLUM. L. REV. 1093, 1120 (2008) ("In general, [B]lack and white people obtain information through different informational networks, which results in racialized pools of knowledge."); Dhomas Hatta Fudholi, Wenny Rahayu & Eric Pardede, *A Data-Driven Dynamic Ontology*, 41 J. INFO. SCI. 383, 383 (2015) (noting that community knowledge usually remains in that community).

[217] I. Bennett Capers, *Race, Policing, and Technology*, 95 N.C. L. REV. 1241, 1246–47 (2017).

knowledge[,]' talked about in barbershops and hair salons, on church pews and on street corners, and yes, in prisons."[218]

Knowledge around the racial and socioeconomic injustices of the bail system are similarly situated within community pools of knowledge that are accessible by community members and community groups that serve these communities. Focusing on the qualitative data of formerly and currently incarcerated people provides a unique lens into the content of this form of knowledge. Community groups that represent the most impacted communities have often been central to the collection and dissemination of this qualitative data. One reason is that these community groups are active participants in ongoing social movements designed to secure racial and socioeconomic justice by reducing the size of the carceral state.[219] In recent years, their work in the pretrial field has garnered increased scholarly attention, as public outrage grows around cash bail and mass pretrial incarceration.[220]

The work of the National Council for Incarcerated and Formally Incarcerated Women and Girls, JustLeadershipUSA, All of Us or None, and Safe & Just Michigan is illustrative.[221] These groups hold listening sessions and panels and conduct surveys and interviews designed to collect qualitative data from formerly and currently incarcerated people about their experiences in pretrial and posttrial incarceration. Though formerly and currently incarcerated people are not a monolith, the patterns that emerge from their individual anecdotes reveal the individual, familial, and community-based harms of pretrial incarceration. One example is an anecdote from Priscilla Echi, who was incarcerated before trial on an assault charge. In describing the impact of her pretrial incarceration on a JustLeadership USA panel, she recalls:

> It had a tremendous impact on my life and my family's life. My youngest was taken from me. He didn't know who I was when I came home. . . . I didn't get visits with my kids for two

---

218 *Id.* (citing David R. Maines, *Information Pools and Racialized Narrative Structures*, 40 SOC. Q. 317, 219–20 (1999)) (alteration in original).

219 Rahman & Simonson, *supra* note 189, at 693–99 (2020) (discussing how community groups are connected to wider racial and economic movements for justice).

220 Simonson, *supra* note 92, at 589–91.

221 NAT'L COUNCIL, *supra* note 208; JUSTLEADERSHIPUSA, https://jlusa.org/ [https://perma.cc/8GHJ-GC6L] (last visited Feb. 24, 2021); *All of Us or None*, LEGAL SERVS. FOR PRISONERS WITH CHILD., https://prisonerswithchildren.org/about-aouon/ [https://perma.cc/JP4R-LXQZ] (last visited Feb. 24, 2021); SAFE & JUST MICH., https://www.safeandjustmi.org/ [https://perma.cc/B35G-8BA3] (last visited Feb. 24, 2021).

> whole years. . . . I had to take a plea deal because I wanted to come home and start my life again whatever that looked like. . . . I wanted to get back to my kids.[222]

Her story is illustrative of the impact that pretrial incarceration has on mothers.

The data derived from formerly and currently incarcerated people are disseminated in several forms such as podcasts, panels, listening sessions, and video-taped interviews. In performing this data collection and dissemination function, these community groups are engaged in public sensitization and law reform efforts aimed at disrupting taken for granted assumptions around the relationship between public safety and incarceration. The qualitative data produced by these groups counter the notion that pretrial incarceration promotes the public safety for all. Rather, this data shows how the psychologically, financially, and physically destabilizing features of pretrial incarceration undermine it. Given the content of this data, it has the potential to dismantle existing discriminatory ideas and to pave the way for new anti-racist conceptualizations of public safety that can protect defendants, their families, and members of their communities.

The next question is how can we tap into community knowledge sources in algorithmic systems? Though a detailed discussion of operationalizing this proposal is beyond the scope of this Article, the work conducted by Silicon Valley De-Bug (De-Bug), a community organization, is illustrative. Since 2019, De-Bug activists have created "community support identifying" forms that are designed to be completed by community members invested in the outcome of the defendant's bail hearing.[223] These forms ask a series of questions that prompt respondents to provide data about the defendant's standing in the community, caretaking obligations, and the consequences that will occur if they are incarcerated before trial. Respondents are also required to note the obligations that they would undertake to ensure the defendant's future court appearance and non-committance of

---

[222] JustLeadershipUSA, *#FREEnewyork Webinar 2*, YOUTUBE (May 24, 2018), https://www.youtube.com/watch?v=Q3gZFE_onqs [https://perma.cc/XF5N-NRXB].

[223] Raj Jayadev, *The Future of Pretrial Justice Is Not Money Bail or System Supervision—It's Freedom and Community*, SV DE-BUG (Apr. 4, 2019), https://www.siliconvalleydebug.org/stories/the-future-of-pretrial-justice-is-not-money-bail-or-system-supervision-it-s-freedom-and-community [https://perma.cc/6YUH-J4KJ].

pretrial crime if pretrial release was granted.[224] Additionally, the form encourages respondents to contextualize data points, such as convictions, arrests, and past instances of non-appearance—data points that have traditionally been utilized to justify a defendant's incarceration before trial. This contextualization provides bail judges with the information needed to ascertain the weight to afford these data points in a bail determination. The strategy employed by De-Bug could provide a starting point for thinking about how qualitative methods could be harnessed to capture data produced by community knowledge sources for use in algorithmic construction.

A related question is what institutional framework would enable this shift? In prior work, I have proposed the creation of bail commissions composed of members from most impacted communities, which would displace the private system of algorithmic development.[225] Under this model, members from most impacted communities would have complete control or veto power over if and how pretrial algorithms are used within their jurisdiction. An important aspect of this model is that it would involve shifting power over algorithmic construction to members of most impacted communities. The idea of power shifting is not novel. In recent years, critical race theorists, abolitionists, and democratization of criminal law scholars have explored how partially or completely ceding power over criminal legal institutions to most impacted communities has the potential to decouple those institutions from their racist and classist origins.[226] Though power shifting does not guarantee a turn to community knowledge sources, the experience of power shifting in the arena of policing provides a reason for optimism. Jocelyn Simonson's work has discussed how endowing most impacted communities with control over the police can enable policing to be responsive to community knowledge and expertise around public safety.[227]

---

224 *Id.*

225 Okidegbe, *supra* note 10, at 774–76.

226 Benjamin Levin, *Criminal Justice Expertise*, 90 FORDHAM L. REV. 2777, 2777–78, 2779–86 (2022) (documenting the rise in scholarship advocating for power-shifting as a means to deconstruct and reconstitute the notion of expertise in criminal law reform).

227 Jocelyn Simonson, *Police Reform Through a Power Lens*, 130 YALE L.J. 778, 806–10 (2021) (discussing how shifting power over policing to communities means allowing for communities to draw on their own knowledge to reform policing).

C. Implications of Turning to Community Knowledge Sources

There are important limitations to using community knowledge sources in algorithmic construction. The turn to community knowledge sources does not guarantee a particular outcome. Like carceral knowledge sources, the data from community knowledge sources is neither objective nor neutral, meaning that its utilization raises issues around incompleteness, bias, and inaccuracy. However, the turn to these knowledge sources offers the chance to engage with data that is not necessarily tied to the functioning of the current racial and class caste system, meaning that algorithms built with such data might produce different and perhaps more equitable outcomes. With this in mind, there are two different sets of implications that may accrue from drawing on community knowledge sources. The first pertains to the construction of future pretrial algorithms, and the second concerns pretrial policy.

1. *Pretrial Algorithms*

The turn toward community knowledge sources could facilitate the creation of algorithms that can produce more accurate risk predictions. Since currently employed algorithms are constructed solely with data from carceral knowledge sources, they produce risk scores that are not predictively accurate regarding members from racially and socioeconomically marginalized communities. Data from community knowledge sources might increase the predictive accuracy of algorithmic systems, since community knowledge around which factors are probative for ascertaining a defendant's public safety risk could ensure that algorithmic inputs track a defendant's dangerousness rather than their experiences arising from being at the intersection of race and class disadvantages.[228] Utilizing this qualitative data could mean eliminating the blanket use of certain inputs, such as prior convictions, as risk factors in algorithmic systems.

Moreover, tapping into community knowledge sources could affect the problem formation stage of algorithmic construction. Rather than defining the problem that the algorithm aims to address solely in terms of whether a defendant will be arrested for pretrial crime or fail to appear at court, developers engaged with community knowledge would

228 Okidegbe, *supra* note 10, at 777–79.

have the ability to define the problem in terms of how the community understands public safety. One possibility is that it could enable the development of algorithms that aim to address the various risks at play during a pretrial release eligibility determination. Such algorithms might produce a holistic risk prediction that accounts for the risks that incarceration poses to a defendant, their family, and their community alongside the risks that pretrial release poses. Seriously contending with this knowledge could lead to the development of algorithms that consider factors that should count against pretrial incarceration, since incarcerating defendants with such factors would produce more harms than benefits. One such factor that might count against pretrial incarceration is the fact that the defendant is the primary custodian of a minor child. Such possibilities indicate that turning to community knowledge sources could produce more just outcomes in the pretrial arena for racially marginalized defendants.

2. *Pretrial Policy*

This Article takes no position as to whether the algorithms developed under its prescription should be mandatory or advisory. However, even if these algorithms are advisory, the turn to community knowledge sources in pretrial algorithmic construction could have implications on pretrial policy. For example, it could spur legislative reform around the law governing bail. As it currently stands, bail judges are not statutorily required to consider the public safety risks that may accrue from a defendant's incarceration before trial. The use of the algorithms envisioned by this Article might push jurisdictions to revise bail statutes to mandate that bail judges consider these risks. Doing so would allow bail judges to conduct a realistic cost-benefit analysis about the public safety risks and harms implicated in bail determinations.[229]

The implementation of algorithms built with community data could alter the practices of institutional actors. One view is that bail judges might ignore the predictions produced by such algorithms, a problem that has already begun to

---

229 It is important to note that empirical work has shown that bail judges do not engage in any kind of "straightforward cost-benefit analysis" when determining a defendant's pretrial release eligibility. Stevenson & Mayson, *supra* note 77, at 770–71. Instead, "bail magistrates seem to be engaged in a mental and moral calculus outside of a technical evaluation of risk." *Id.* at 771.

materialize with currently employed algorithms.[230] The turn toward community knowledge sources in algorithmic construction might hasten this trend or avert it, depending on how bail judges view the predictions produced by such algorithms. Bail judges might adhere to the predictions produced by these algorithms on the theory that these algorithms represent community preference and resultingly are more democratically legitimate than currently employed algorithms. In that case, such algorithms would not only affect pretrial release determinations, they could also affect prosecutorial practices in the pretrial arena. For instance, community data may indicate that those with certain kinds of criminal convictions pose little to no threat to the public safety of their community. Such convictions would resultingly not be a factor within the algorithm. Under these circumstances, prosecutors may choose not to contest the pretrial release of defendants with these criminal convictions. This outcome would radically change how the pretrial system would operate in communities that have been most harmed by the criminal legal system.

The use of community knowledge sources would also impact the strategies that most impacted communities might undertake in their effort to reform or dismantle the pretrial system. It is important to note that most impacted communities hold mixed views about the relationship between the criminal legal system and public safety. Monica Bell's work has shown that impacted communities have divergent views about abolition and the continuation of the criminal legal system.[231] In prior scholarship, I have suggested that the paradigm governing algorithms should endow communities with complete or substantial control over if and on what basis algorithmic reforms are used in order to accommodate the divergent perspectives held towards the criminal legal system and the algorithm project. For communities open to the algorithm project under a community control model, the turn toward community knowledge sources could enable algorithms to be built on a different and potentially anti-racist

230 Megan T. Stevenson & Jennifer L. Doleac, *Algorithmic Risk Assessment in the Hands of Humans* 30–31 (IZA Inst. of Labor Econ. Working Paper No. 12853), https://www.econstor.eu/bitstream/10419/215249/1/dp12853.pdf [https://perma.cc/22QK-T6SD] (noting that the influence that a defendant's risk score has on judicial discretion decreases the longer the risk assessment tool is used in the jurisdiction).

231 Monica C. Bell, *Anti-Segregation Policing*, 95 N.Y.U. L. REV. 650, 732–734, 760–65 (2020).

conceptualization of public safety.[232] Since currently used algorithmic systems only utilize data from carceral knowledge sources, they are built on a notion of public safety that only considers pretrial crime and pretrial flight. This singular focus is reflected in the risk factors that are common inputs in algorithmic systems.[233] The current conceptualization of public safety has harmed members of most impacted communities by rendering them particularly vulnerable to pretrial incarceration. Due to the overpolicing, over-criminalization, and overincarceration of their communities, they are more likely than members of under-policed communities to accrue arrests, convictions, and other "high risk" factors that will cause the pretrial system to label them a threat to public safety.[234] The qualitative data of community groups has already shown that the conceptualization of public safety pursued by the pretrial system will continue to be underinclusive unless it is expanded to incorporate the harms of incarceration. For example, the quantitative data from Priscilla Echi and other formerly incarcerated mothers has illustrated the harms of parental incarceration. Those harms are particularly acute in most impacted communities where the high rate of parental incarceration weakens the social norms and controls that operate to uphold the wellbeing and safety of these communities.[235] Engagement with community knowledge sources could enable the creation of algorithms that pursue a conceptualization of public safety, which realistically accounts for the public safety risks and harms implicated in bail determinations. While this different conceptualization of public safety may not be a complete solution to the problem of racialized mass pretrial incarceration, it may trigger more equitable and just outcomes in the pretrial arena. In sum, these potentialities indicate that turning to community knowledge sources in algorithmic construction could provide a

232 Okidegbe, *supra* note 10, at 784.

233 The risk factors used by PSA only relate to pretrial flight and crime risk. For this reason, the PSA system only utilizes the following factors: pending charges, prior convictions, prior failures to appear, age, and prior sentences to incarceration. The PSA system does not consider factors that would gauge the public safety risks associated with detaining a defendant before trial. *See About the Public Safety Assessment: How It Works*, ADVANCING PRETRIAL POL'Y & RCH., https://advancingpretrial.org/psa/factors/ [https://perma.cc/3EPC-FC8J] (last visited Feb. 24, 2021).

234 Eaglin, *supra* note 15, at 487; Mayson, *supra* note 4, at 2229–30, 2234.

235 Lewis, *supra* note 209, at 1221 (noting how public safety cannot be protected if law, policy, and court decisions around incarceration undermine the norms and dynamics that facilitate community well-being and cohesion in an area).

path toward orienting the pretrial system away from its carceral, classist, and racist tendencies.

IV
IS A TURN TO NON-CARCERAL KNOWLEDGE SOURCES IN ALGORITHMIC CONSTRUCTION POSSIBLE?

This Article has argued that the data source selection problem has fueled algorithmic discrimination. It has explained how any effort to reform algorithmic discrimination that does not attend to the dominance of carceral knowledge sources in algorithmic construction will fail. It has identified how a turn toward non-carceral knowledge sources would provide a promising basis from which to construct algorithms that have the potential to become tools for racial and socioeconomic justice.

However, one critical question remains: Is a turn to non-carceral knowledge sources possible? The turn to non-carceral knowledge sources, such as community knowledge sources, is far afield from the current paradigm governing algorithmic construction. The proposal also runs counter to traditional justifications given by algorithm reformers for the use of algorithms in the pretrial system, in particular, and in the criminal legal system generally.[236] Additionally, its operationalization is subject to a public will to recognize and rectify the barrier that carceral data plays in order to redress racial and socioeconomic inequities in the pretrial system and beyond. If community knowledge sources are to be used, this Article's proposal further requires the trust, consent, and buy-

236 In the criminal legal context, one reason in support of evidence-based practices like algorithms comes from the fact that these tools are developed, implemented, overseen, and used by traditional experts as opposed to by members of the public. For these proponents, because these tools are not the product of populism or the "public," they are insulated from the kind of punitive populism that has fueled mass incarceration, and they can therefore improve the criminal legal system. For an example of scholarship on this point, see generally RACHEL ELISE BARKOW, PRISONERS OF POLITICS: BREAKING THE CYCLE OF MASS INCARCERATION 15 (2019) (advocating for evidence-based strategies to redress mass incarceration). Yet it is important to note that Barkow has also taken the position that lived experience as a source of expertise is important to improving criminal law outcomes. *See generally* Rachel E. Barkow & Mark Osler, *Designed to Fail: The President's Deference to the Department of Justice in Advancing Criminal Justice Reform*, 59 WM. & MARY L. REV. 387, 459 (2017) (contending that the lived experiences of formerly incarcerated people should be utilized in a presidential criminal justice advisory commission). On this point, it is also important to note that Erin Collins's work highlights how the neoliberal paradigm and epistemology governing evidence-based strategies hamper the capacity of such strategies to meaningfully transform the criminal legal system. *See* Collins, *supra* note 32, at 50.

in from communities that have been ravaged by the carceral state and, as a result, might prefer abolition over expending their limited resources to improve these tools, especially if jurisdictions continue to mandate that the tools be used solely for carceral purposes. It also requires the creation of new legal and institutional regimes around the development, implementation, and oversight of such algorithms. If any of the above barriers cannot be overcome, then this Article's proposal is unrealizable. Acknowledging this caveat, this section addresses other objections that may be posed even if the above issues were resolved.

A. Subjectivity Objection

Some may claim that the qualitative data produced by community knowledge sources would be too subjective to be translated into algorithmic inputs that will ultimately impact the bail determinations of an entire community or even a jurisdiction. It is true that, by definition, lived experience varies on an individual basis, even if an individual is a member of a community highly impacted by the criminal legal system. The fact that there are various perspectives about the pretrial system may complicate the process of translating these perspectives into algorithmic inputs. However, there are two flaws with this objection. First, though this Article focuses on the use of qualitative data, there is no reason why quantitative data from community knowledge sources could not be used as well. Second, there are common threads within individual accounts of the pretrial system that can be drawn upon to develop algorithmic inputs. For instance, many community groups that have operated as community knowledge sources have produced qualitative data demonstrating the specific harm of parental incarceration, a harm that could serve as an algorithmic input.[237] Moreover, deriving algorithmic inputs from qualitative data is not that different from deriving algorithmic inputs from quantitative data. Though statistical modeling assists developers in determining which factors are most statistically correlated to alleged pretrial misconduct, developers have to make normative judgments about which factors to include in the final algorithmic model.[238] Often, the factors chosen are not necessarily the ones most statistically correlated to pretrial misconduct.[239] There is no reason why

237 *See infra* subpart III.B.
238 Eaglin, *supra* note 4, at 87–94.
239 *Id.* at 83–84.

the institutional structure charged with developing the algorithms envisioned by this Article could not make similar normative judgments about which inputs to include in the final algorithmic system when confronted with conflicting qualitative data.

B. Accuracy Objection[240]

Another objection may be that algorithms derived from community knowledge sources will not provide accurate predictions. A version of this objection has been raised by Sandra Mayson, who warns that current attempts to redress algorithmic discrimination through either algorithmic affirmative action or regulating race-correlated input variables risks creating algorithms that produce predictively inaccurate assessments of a defendant's flight or public safety risk.[241] This critique is compelling only if we continue to define "accurate" algorithms as ones tethered to the current conceptualization of public safety that only accounts for non-appearance and risk of pretrial crime. Turning to community knowledge sources means engaging with a different conceptualization of accuracy, one rooted in a different way of knowing about public safety that may provide a realistic assessment of the costs and benefits of pretrial incarceration. Under this notion of accuracy, community knowledge sources could enhance the accuracy of algorithms by enabling them to predict whether a defendant's incarceration is likely to be in line with protecting public safety given that community's own cost/benefit assessment.[242] Given this, the notion of accuracy should be subject to the knowledge of communities most harmed by the current system.

C. Democratization Objection

Some may critique this Article's proposal for the unwarranted faith that it places in the power of community

240 This Article uses the term "accuracy" in reference to whether the algorithm reliably predicts the likelihood of pretrial misconduct. This issue is generally referred to in computer science literature as the "validity" of the tool, but this Article uses "accuracy" since such aligns with the common use of the word.

241 Mayson, *supra* note 4, at 2262, 2266–67, 2272.

242 It is important to note that these predictions could be more accurate in the sense that the algorithm could be built with community knowledge that factors in the public safety risks of pretrial incarceration. However, the turn to community knowledge sources may not be a complete solution, particularly if the state continues to divest from and govern these marginalized communities through carceral systems. See Dorothy E. Roberts, *Foreword: Abolition Constitutionalism*, 133 HARV. L. REV. 1, 15 (2019).

knowledge sources to transform algorithmic systems. This critique is connected to recent criminal law scholarship that is highly critical of moves to increase public participation in the criminal legal system. A recent articulation of this objection was offered by John Rappaport, who argues that increased public participation in the criminal law process risks increasing rather than decreasing the punitiveness of the criminal legal system.[243] These concerns have pushed Rappaport and others to advocate for technocratic control over criminal legal reform.[244] The idea behind this model is to empower technocrats to pursue evidence-based approaches "consistent with democratic values" in order to yield a fairer criminal justice system.[245]

This Article's proposal does not advocate for the form of democratization at which Rappaport's critique is directed. Nonetheless, his critique remains applicable, since my approach endorses community knowledge sources, which necessitates a turn away from the carceral knowledge sources at the heart of evidence-based approaches. One response to his critique is that there is reason to question whether evidence-based approaches can produce a fairer criminal legal system.[246] This is because using carceral knowledge sources imposes constraints on our understanding of the relationship between public safety and incarceration. The very fact that currently employed pretrial algorithms produce assessments that omit considerations of the harms of incarceration represents a particular view about the purpose of the pretrial system and how it should protect public safety. Departing from this view is a prerequisite to unlocking true change in the pretrial system, but this change cannot be achieved without shifting to different knowledge sources.

CONCLUSION

Algorithmic discrimination presents a troubling challenge to racial and socioeconomic justice. It has caused today's algorithms to maintain the very same racial and class-based

---

243 John Rappaport, *Some Doubts About "Democratizing" Criminal Justice*, 87 U. CHI. L. REV. 711, 764–65, 808–09 (2020).

244 *Id.* at 720, 809-13; *see also* Levin, *supra* note 226, at 2836–37 (documenting the rise of the technocratic model of criminal legal reform and the incompatibility of it with abolitionist and community driven visions of law reform).

245 Rappaport, *supra* note 243, at 809-13.

246 Collins, *supra* note 32 (contending that the use of the evidence-based model as currently employed replicates the racial and carceral logics of mass incarceration and thus supports rather than dismantles the status quo).

inequities that their use was designed to eliminate. By so doing, pretrial algorithms operate to uphold the structural harms that undermine public safety for members of vulnerable, poor, and racially marginalized communities in this country. As the debate around algorithmic discrimination continues, increased attention must be paid to the data source selection problem. By shedding light on this problem, the hope is that this Article adds an epistemological dimension to the ongoing conversations about algorithmic discrimination. An approach that only critiques the biases in the data currently used in algorithmic systems without attending to the knowledge sources that have produced this data cannot solve algorithmic discrimination. Moreover, such approaches threaten to perversely impede the ability to shape future algorithms toward racial justice by tethering them to knowledge sources implicated in racial and class hierarchy. Resolving algorithmic discrimination is potentially realizable, but to do so requires a radical paradigmatic shift away from carceral knowledge sources toward non-carceral knowledge sources. With this shift, latent possibilities open up for the development of algorithms that could produce different outcomes. Perhaps, algorithms offer a path to the radical reorientation of the pretrial system that would decrease the harms that it enacts on racially and socioeconomically marginalized communities. But unlocking that path requires us to dismantle the knowledge sources that brought us here in the first place.