

IN THE SUPREME COURT OF PENNSYLVANIA

No. 21 EM 2019

THE PHILADELPHIA COMMUNITY BAIL FUND, *et al.*, Petitioners

v.

ARRAIGNMENT COURT MAGISTRATES OF THE FIRST JUDICIAL
DISTRICT OF THE COMMONWEALTH OF PENNSYLVANIA, Respondents

**BRIEF FOR AMICUS CURIAE AI NOW INSTITUTE IN SUPPORT OF
PETITIONERS**

On Allowance of Objections to the Report of the Special Master
Dated December 16, 2019

Rashida Richardson, Cal. ID No. 282565
Jason M. Schultz, N.Y. ID No. 5376520
AI NOW INSTITUTE
155 Avenue of the Americas, Floor 4
New York, NY 10013
(917) 608-6652
rashida@ainowinstitute.org

Sara E. Jacobson, Pa. ID No. 80965
TEMPLE UNIVERSITY,
BEASLEY SCHOOL OF LAW
1719 North Broad Street, Room 614
Philadelphia, PA 19122
(215) 204-0661
Sara.jacobson@temple.edu

Counsel for Amicus

TABLE OF CONTENTS

Statement of Interest of Amicus Curiae1

Introduction and Summary of Argument2

Argument5

I. Current research shows that risk assessment tools would not improve the Philadelphia bail system and instead undermine the Special Master’s goal in a bail decisional process that is “fair and perceived to be fair.”5

 A. Risk assessment tools reproduce racial disparities prevalent within the criminal justice system because they rely on historic criminal justice data.6

 B. Current technical methods cannot adequately address the deeply skewed and biased criminal justice data used in risk assessment tools.9

 C. There are no satisfactory legal claims available to defendants seeking to challenge biased outputs made by risk assessment tools.14

II. Empirical studies suggest that risk assessment tools’ ability to improve bail-setting practices is at best limited, and at worst counter-productive. ...16

III. The bail system improvements recommended by the Special Master can be made without the adoption of pretrial risk assessment tools.18

Conclusion.....20

TABLE OF AUTHORITIES

Cases

State v. Loomis, 881 N.W.2d 749 (Wis. 2016).....16

Other Authorities

AI Now Institute, *Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems* (2018)13

Alexandra Chouldechova, *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*, 5 Big Data 153 (2017)8, 9

Andrew G. Ferguson, *Policing Predictive Policing*, 94 Wash. U.L. Rev. 1115 (2017).....10

Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 Duke L.J. 1043 (2019)..... 14, 15

Bernard E. Harcourt, *Risk as a Proxy for Race: The Dangers of Risk Assessment*, 27 Federal Sentencing Reporter 237 (2015)13

Brad Heath, *Racial Gap in U.S. Arrest Rates: “Staggering Disparity,”* USA Today (Nov. 18, 2014), <https://perma.cc/V9MY-K2WN>7

Jason Tashea, *Risk-Assessment Algorithms Challenged in Bail, Sentencing and Parole Decisions*, ABA J. (Mar. 1, 2017), <https://bit.ly/38uE7AI>8

John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 Wash. L. Rev. 1725 (2018)..... 12, 18

Jon Kleinberg et al., *Algorithmic Fairness*, 108 AEA Papers and Procs. 22 (2018) 13, 15

Jon Kleinberg et al., *Inherent Trade-offs in the Fair Determination of Risk Scores*, Leibniz Int’l Proc. Informatics, Jan. 2017.....8, 9

Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), <https://bit.ly/38xDwhD>8

Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, Science Advances (Jan. 17, 2018), <https://bit.ly/37gdxuF>17

Juliana Reyes, *Philly Police Will Be First Big City Cops to Use Azavea’s Crime Predicting Software*, Technical.ly Philly (Nov. 7, 2013, 12:30 PM), <https://bit.ly/2u0Z0oe>11

Kristian Lum & William Isaac, *To Predict and Serve?*, 13 Significance 14 (2016)12

Kristian Lum et al., <i>The Impact of Overbooking on a Pre-trial Risk Assessment Tool</i> , Procs. of 2020 Conf. on Fairness, Accountability, & Transparency 482 (2020).....	10
Megan Stevenson, <i>Assessing Risk Assessment in Action</i> , 103 Minn. L. Rev. 303 (2018).....	12, 18
Megan T. Stevenson & Jennifer L. Doleac, <i>Algorithmic Risk Assessment in the Hands of Humans</i> (IZA Inst. of Labor Econ. DP No. 12853, 2019)	18
Pauline T. Kim, <i>Data-Driven Discrimination at Work</i> , 58 Wm. & Mary L. Rev. 857 (2017).....	14
Pl.’s 8th R. to Ct. & Monitor on Stop & Frisk Practices: Fourteenth Amendment Issues, <i>Bailey v. City of Philadelphia</i> , No. 2:10-cv-05952 (E.D. Pa. Jan. 5, 2018), https://bit.ly/2RVvxnU	11
Rashida Richardson et al., <i>Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice</i> , 94 N.Y.U. L. Rev. 192 (2019).....	7, 10, 11
Richard Berk et al., <i>Fairness in Criminal Justice Risk Assessments: The State of the Art</i> , Soc. Methods & Res. (2018)	7, 9
Sam Corbett-Davies et al., <i>Algorithmic Decision Making and the Cost of Fairness</i> , Procs. 23rd Int’l Conf. Knowledge Discovery & Data Mining 797 (2017).....	9
Samantha Melamed, <i>Will Controversy Over Risk Assessments Break Philly’s Touted Criminal-Justice Reform Collaboration?</i> , Phila. Inquirer (May 8, 2019), https://bit.ly/30KDVe0	19
Sandra G. Mayson, <i>Bias In, Bias Out</i> , 128 Yale L.J. 2218 (2019).....	6, 7, 8, 9
Sarah L. Desmarais & Evan M. Lowder, <i>Pretrial Risk Assessment Tools: A Primer for Judges, Prosecutors and Defense Attorneys</i> (2019).....	6, 17
Settlement Agreement, Class Certification, and Consent Decree, <i>Bailey v. City of Philadelphia</i> , No. 2:10-cv-05952 (E.D. Pa. June 21, 2011), https://bit.ly/2GtxmCX	11
Solon Barocas & Andrew D. Selbst, <i>Big Data’s Disparate Impact</i> , 104 Calif. L. Rev. 671 (2016).....	14
U.S. Dep. of Housing & Urb. Dev., <i>Neighborhoods and Violent Crime</i> (2016), https://bit.ly/38LKgs7	13

STATEMENT OF INTEREST OF AMICUS CURIAE

The AI Now Institute (AI Now) is a public interest research institute founded at New York University that performs interdisciplinary research on the social implication of artificial intelligence and other emerging data-driven technologies, including risk assessment tools. AI Now has particular expertise on the challenges and risks created by government use of data-driven technologies in sensitive social domains, like criminal justice, and developing policy interventions to address problems identified in this research.

AI Now seeks to appear as *amicus curiae* in this Response to the Special Master's report which the Court has before it pursuant to its King's Bench Jurisdiction to draw the Court's attention to the disproportionate harms and problems produced by the use of risk assessment in pretrial decisionmaking. We respectfully submit this brief in furtherance of two goals. First, we hope to inform the Court of current research from a variety of disciplines on the efficacy and validity of risk assessment tools. Second, we hope to persuade the Court to deny the Special Master's recommendation to develop and implement a risk assessment tool to be used by Arraignment Court Magistrates for bail determinations in the First Judicial District.

INTRODUCTION AND SUMMARY OF ARGUMENT

In July of 2019, this Court invoked its King’s Bench jurisdiction and appointed a Special Master to report on Plaintiff’s allegations of “systemic failures of the First Judicial District to properly conduct cash-bail matters.” Though the Special Master found that the current bail system was “fundamentally sound,” the Special Master’s report highlighted numerous areas of agreement between representatives of the First Judicial District, the Philadelphia Municipal Court, the Philadelphia District Attorney’s Office, and the Defender Association of Philadelphia on how to improve the bail system. Extraneous to these agreements, the Special Master recommended that the Court consider the development of a pretrial risk assessment tool with the goal of improving the bail system by helping Philadelphia’s Arraignment Court Magistrates (ACMs) take into account “individualized differences” between defendants and sort out defendants who present a “community danger” before they are released.

Though the Special Master’s recommendation may be a mere suggestion, *amicus* AI Now Institute urges the Court to avoid embracing it. Risk assessment tools offer the promise of improving individual assessments for use within the criminal justice system, including for bail setting purposes, but few if any have delivered on that promise. Rather, research has shown copious amounts of evidence of racial biases endemic to risk assessment. Reviving debate over risk

assessment tools at this juncture is an unnecessary distraction from the Court's opportunity to improve on bail. Further, the Special Master's other recommendations can improve the Philadelphia bail system without the use of biased and unfair risk assessment tools.

Despite many risk assessment tools' promises of bringing greater "objectivity" to bail decisions, research from computer science, the social sciences, and the legal field has repeatedly shown that risk assessments produce biased measures of risk that disproportionately harm non-white defendants. These racially biased effects stem from how the tools are designed and their blanket reliance on historical data. Because risk assessment tools evaluate a defendant's likelihood of future recidivism by detecting patterns in historic arrest records, any patterns in the data of past racial biases and disparities in the criminal justice system will be replicated and reinforced in the tool's evaluations. This problem has been known for years, and despite the valiant efforts of hundreds of computer science, social science, and criminal justice researchers working on it, there are still no known technical methods to adequately mitigate these harms. Further, there is no adequate legal recourse available for a defendant whose bail decision may have been influenced by a tool's biased outputs, other than to simply eliminate the use of biased risk assessment tools completely.

Empirical studies also suggest that pretrial risk assessment tools fail to live up to their own promise of improving bail setting practices. Despite hopes that these tools may encourage judges to make less-biased decisions, evidence shows that judges are themselves inconsistent in how they use pretrial risk assessment tools in their bail decisions.

In the past two years, both the District Attorney's Office and the Public Defender withdrew their support for the development and use of a pretrial risk assessment tool in Philadelphia in large part because of the concerns that *amicus* raise today. There is no reason to believe that a renewed effort to develop a risk assessment tool following the Special Master's recommendation would yield a different result. While the Special Master is correct to note that the District Attorney's Office's opposition was not against a pretrial tool *per se*, a new effort to develop a local pretrial risk assessment tool is bound to run into the same obstacles.

More research from the national community of computer scientists, sociologists, and legal practitioners is needed to understand if these fundamental problems of risk assessment tools are surmountable and whether the tools can be made safe and beneficial for use in pretrial settings. In the meantime, the Special Master's proposed improvements to the bail system can be made without pretrial

risk assessment tools and without bogging this process down in further debate over them.

ARGUMENT

I. CURRENT RESEARCH SHOWS THAT RISK ASSESSMENT TOOLS WOULD NOT IMPROVE THE PHILADELPHIA BAIL SYSTEM AND INSTEAD UNDERMINE THE SPECIAL MASTER’S GOAL IN A BAIL DECISIONAL PROCESS THAT IS “FAIR AND PERCEIVED TO BE FAIR.”

The Special Master considers it critical that the Philadelphia bail decisional process be “fair and perceived to be fair.” (Special Master’s Report at 14.)

Research has shown risk assessment tools to be neither because they replicate and reinforce racial biases that have been historically endemic to the criminal justice system. Research also shows the numerous and currently insurmountable challenges preventing tool developers from addressing those biases. In the absence of evidence that these tools can be deployed equitably and constitutionally, the Court should not consider risk assessment tools as plausible solutions when addressing the Special Master’s proposed improvements to the Philadelphia bail system.

A. Risk assessment tools reproduce racial disparities prevalent within the criminal justice system because they rely on historic criminal justice data.

Pretrial risk assessment tools are designed to predict the likelihood that a defendant will either be rearrested if released or will fail to appear in court. Sarah L. Desmarais & Evan M. Lowder, *Pretrial Risk Assessment Tools: A Primer for Judges, Prosecutors and Defense Attorneys* 4–5 (2019), <https://bit.ly/3aVUVme>. This is accomplished through statistical analyses of historical arrest data that evaluate how strongly certain “factors” about defendants—including age, criminal history, employment status, education, and housing status—correlate with rearrest. *Id.* at 3. With that analysis, risk assessment tools weigh the characteristics of new defendants to predict the likelihood that they too will be arrested. For example, if defendants with a history of substance abuse were arrested more often than other defendants in the past, then a risk assessment tool may associate substance abuse with recidivism when evaluating new defendants.

Because risk assessment tools function like a “mirror” that reflects the past through predictions of the future, this process also enables risk assessment tools to produce racially-biased outcomes. Sandra G. Mayson, *Bias In, Bias Out*, 128 *Yale L.J.* 2218, 2251 (2019). In the United States, at least 1,581 police departments arrest Black people at rates at least three times higher than white people. Mayson, *supra*, at 2253 (citing Brad Heath, *Racial Gap in U.S. Arrest Rates: “Staggering*

Disparity,” USA Today (Nov. 18, 2014), <https://perma.cc/V9MY-K2WN>). In at least 70 jurisdictions, Black people are arrested at rates ten times higher than white people. Brad Heath, *Racial Gap in U.S. Arrest Rates: “Staggering Disparity,”* USA Today (Nov. 18, 2014), <https://perma.cc/V9MY-K2WN>. When these arrest statistics are used as data for risk assessment tools, those tools can similarly learn to associate Black defendants with a higher likelihood of rearrest. *See, e.g.,* Mayson, *supra*, at 2224; Richard Berk et al., *Fairness in Criminal Justice Risk Assessments: The State of the Art*, Soc. Methods & Res. 2 (2018). Risk assessment tools’ reliance on historical data means that “a racially unequal past will necessarily produce racially unequal outputs.” Mayson, *supra*, at 2224.

Whether or not the racial disparity in arrest data may be the result of racist policing practices, *see, e.g.,* Rashida Richardson et al., *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. Rev. 192, 197–98 (2019), or racial disparities in the commission of crime, the mere fact that the racial disparity exists can cause risk assessment tools to make unfair and unjust evaluations of Black defendants. Reporting by ProPublica on the popular COMPAS risk assessment tool put this phenomenon in its starkest terms. The COMPAS risk assessment has been used throughout the United States in bail, sentencing, and parole settings. Jason Tashea, *Risk-Assessment Algorithms Challenged in Bail, Sentencing and Parole Decisions*,

ABA J. (Mar. 1, 2017), <https://bit.ly/38uE7AI>. Through statistical analysis, ProPublica researchers determined that the COMPAS assessment was more likely to erroneously label a Black defendant as a “high risk” for recidivism than a white defendant—even if neither defendant went on to recommit a crime. Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), <https://bit.ly/38xDwhD>.

Studies following the ProPublica article have explained that this phenomenon is inevitable because not all metrics of “fairness” can be reconciled. It was impossible for the COMPAS risk assessment tool to evaluate Black and white defendants using the same criteria (a characteristic known as “predictive parity”) and not generate more incorrect predictions about Black defendants (i.e. not subject Black defendants to more false predictions of being high risk). See Alexandra Chouldechova, *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*, 5 *Big Data* 153, 154 (2017); Jon Kleinberg et al., *Inherent Trade-offs in the Fair Determination of Risk Scores*, *Leibniz Int’l Proc. Informatics*, Jan. 2017, at 43:1, 5–6; Mayson, *supra*, at 2238. In other words, some kind of unfairness in risk assessments is inevitable: it may be impossible for COMPAS to evaluate Black defendants and white defendants by the same standard without causing Black defendants to suffer more erroneous detention decisions because of their race.

B. Current technical methods cannot adequately address the deeply skewed and biased criminal justice data used in risk assessment tools.

No technical methodology can overcome the fact that different racial fairness metrics are currently irreconcilable. *See* Chouldechova, *supra*, at 154; Kleinberg, *supra*, at 43:5–6; *see also* Mayson, *supra*, at 2241–47 (defining different measures of fairness). “The goal of complete race . . . neutrality is unachievable.” Berk et al., *supra*, at 18–19. Nonetheless, since the 2016 ProPublica article on racial bias in risk assessment tools, researchers have tried to develop technical methods to mitigate unfairness in risk assessment tools. However, these methods necessarily impose value-laden trade-offs that involve matters of public policy. *See* Mayson, *supra*, at 2249. For example, maximizing the overall accuracy of a risk assessment tool might help attain higher levels of public safety, but doing so may necessarily violate multiple norms of equity and fairness between racial groups. *See, e.g.,* Sam Corbett-Davies et al., *Algorithmic Decision Making and the Cost of Fairness*, Procs. 23rd Int’l Conf. Knowledge Discovery & Data Mining 797, 797, 802 (2017) (suggesting that optimizing for fairness can cause an increase in violent recidivism); Berk, *supra*, at 30–32 (suggesting that different risk assessment processes for Black and white defendants could lead to equal statistical accuracy for both groups but at the cost of treating different races by different criteria).

On an even deeper level, quick-fix technical solutions are inadequate because the data used in risk assessment tools, and processes that generate that data, are fundamentally flawed. Criminal justice data does not necessarily reflect the “true” picture of who commits crimes: instead, arrest data more often reflects policing practices and policies. When policing practices are themselves racially-biased, then the resulting arrest data will reflect that racial bias. *See* Andrew G. Ferguson, *Policing Predictive Policing*, 94 Wash. U.L. Rev. 1115, 1153 (2017); *see e.g.*, Kristian Lum et al., *The Impact of Overbooking on a Pre-trial Risk Assessment Tool*, Procs. of 2020 Conf. on Fairness, Accountability, & Transparency 482, 488 (2020) (finding that non-white defendants receive more serious charges when they are initially booked, and that this practice of “overbooking” can skew pretrial risk assessments even if the defendant is ultimately charged for lesser offenses). In one of our studies, we examined data produced by police departments under scrutiny for unlawful and biased policing practices (e.g. subject to Department of Justice consent decrees or investigations) that subsequently used predictive tools that, in part, relied on historic police data. *See* Richardson et al., *supra*, at 197–98. We found that in at least nine of the thirteen jurisdictions examined, the resulting tools base their evaluations of riskiness on “dirty data,” or data derived from a skewed version of the world created by biased policing practices. Though the tools may be proposed as

“objective” measures to address earlier injustice, they themselves may reproduce injustice.

Our research found that Philadelphia may have been one of the jurisdictions to experience this “dirty data” problem. *See id.* at 232. In 2011, the City of Philadelphia agreed to a consent decree under pressure that the Philadelphia Police Department’s stop and frisk practices were unconstitutional and racially biased. Settlement Agreement, Class Certification, and Consent Decree, *Bailey v. City of Philadelphia*, No. 2:10-cv-05952 (E.D. Pa. June 21, 2011), <https://bit.ly/2GtxmCX>. In 2018, an analysis of the Philadelphia Police Department’s data revealed that unconstitutional stop and frisk practices continued despite the 2011 consent decree. Pl.’s 8th R. to Ct. & Monitor on Stop & Frisk Practices: Fourteenth Amendment Issues, *Bailey v. City of Philadelphia*, No. 2:10-cv-05952 (E.D. Pa. Jan. 5, 2018), <https://bit.ly/2RVvxuU>. Those multiple years of arrests resulting from unconstitutional stops generated “dirty data” that may have been used to build predictive analytics tools to inform the Philadelphia Police Department’s practices. Juliana Reyes, *Philly Police Will Be First Big City Cops to Use Azavea’s Crime Predicting Software*, Technical.ly Philly (Nov. 7, 2013, 12:30 PM), <https://bit.ly/2u0Z0oe>.

Though the predictive tool at issue in Philadelphia focused on policing and not risk assessment *per se*, the same logic applies: in jurisdictions generating “dirty

data,” risk assessment tools might evaluate non-white defendants as higher recidivism risks simply because they were more heavily-targeted by illegal and biased policing. *See also* Kristian Lum & William Isaac, *To Predict and Serve?*, 13 Significance 14 (2016) (finding that a reliance on drug arrest data leads to predictive policing predictions that disproportionately impact communities of color, likely because drug arrest data is not representative of the true distribution of actual drug offenses). Even jurisdictions that do not have demonstrable records of biased policing have cause for concern. Because risk assessment tools are often developed and validated using data from multiple jurisdictions, harmful policing practices in one jurisdiction can affect tools nationwide. *See* Megan Stevenson, *Assessing Risk Assessment in Action*, 103 Minn. L. Rev. 303, 344 (2018) (noting that Kentucky’s pretrial risk assessment was developed using a dataset of pretrial releases from over 300 jurisdictions); *see also* John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 Wash. L. Rev. 1725, 1756 (2018) (“Geographic differences in law enforcement patterns, for example, can undermine tools’ accuracy.”). This challenge is endemic to criminal justice data and presents a major obstacle to technical efforts to develop risk assessments and other predictive tools.

Risk assessment tools also cannot be “fixed” by simply ignoring race as a factor. First, taking a “colorblind” approach to risk assessment tools can still

produce tools that make less accurate *and* less equitable evaluations of risk. *See* Jon Kleinberg et al., *Algorithmic Fairness*, 108 AEA Papers and Procs. 22, 23 (2018). Second, even when risk assessment tools exclude race as an explicit factor, other factors that correlate more closely with race can function as “proxies” for race. For example, because of a history of racial disparity in arrests and imprisonment, Black defendants are more likely than white defendants to have a more extensive criminal history. Thus, risk assessment tools that include “criminal history” as a factor (which is the case for most, if not all popular risk assessment tools) will be more likely to associate Black defendants with a greater risk of recidivism. Bernard E. Harcourt, *Risk as a Proxy for Race: The Dangers of Risk Assessment*, 27 Federal Sentencing Reporter 237, 240 (2015). Similarly, ill-defined factors used in risk assessments like “community disorganization” can also act as proxies for race since predominantly Black neighborhoods experience more violent crime than predominantly white neighborhoods. *See* AI Now Institute, *Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems* 13 (2018) (discussing the use of “community disorganization” in a sentencing risk assessment tool used in Washington, D.C.); U.S. Dep. of Housing & Urb. Dev., *Neighborhoods and Violent Crime* (2016), <https://bit.ly/38LKgs7>. As long as the role of race is salient in American society, whether through policing, housing,

education, or employment, race will play a role in risk assessments whether a developer intends it to or not.

C. There are no satisfactory legal claims available to defendants seeking to challenge biased outputs made by risk assessment tools.

Though discrimination concerns often raise equal protection doctrine questions, “the two main doctrinal touchstones of bad intent and bad classifications provide scant traction for the analysis of algorithmic criminal justice.” Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 Duke L.J. 1043, 1088 (2019); *see also* Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 Calif. L. Rev. 671, 676 (2016) (explaining the difficulty of proving legal discrimination in the analogous context of employment-related assessment tools and Title VII liability); Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 Wm. & Mary L. Rev. 857, 901–25 (2017) (discussing the challenges of fitting existing disparate impact and disparate treatment doctrines under Title VII to discriminatory assessment tools). In the first case, consider the origin of the data used in a risk assessment tool, a “diffuse haze of discrete policing decisions” leading to arrests. Huq, *supra*, at 1092. Even if those arrests originated from unconstitutional policing practices and intentional racial bias, the fact that a risk assessment tool then relied on that data does not necessarily mean that the tool is now infected by the same bad intent. *Id.* at 1093.

Likewise, under an “anticlassification” theory of equal protection, a risk assessment tool that explicitly uses race as a factor may still pass constitutional muster. A court might allow the use of race in risk assessment tools for the same reasons it allows the use of race in criminal suspect descriptions, even if those reasons are not always clear. *See id.* at 1096 (positing that courts might accept racial classifications in suspect descriptions because they believe they are not rooted in “airy suppositions about racial types,” which might similarly lead courts to accept race in predictive tools). Further, just because a risk assessment tool might explicitly include race as an input to its decision does not necessarily mean that a defendant’s race was necessary to reach a particular prediction. Because risk assessments and other predictive tools are often “black boxes” whose logic is hidden from human decisionmakers (and sometimes even the developers themselves), it is difficult to ever conclude that a risk assessment made a decision “*on the basis of*” a racial classification. *See id.* at 1099–1100. Finally, a court is likely to disregard a challenge to racially-aware risk assessment tools for practical reasons. The inclusion of race as an explicit factor can sometimes play a positive role in making predictions more accurate and equitable. Kleinberg et al., *Algorithmic Fairness, supra*, at 23. A court may be reluctant to ban consideration of race if it believes that race can mitigate a tool’s discriminatory effects.

Other legal challenges to risk assessment tools, such as those made on due process grounds, have also so far proven insufficient. In the Wisconsin case *State v. Loomis*, for example, a defendant alleged that the state violated due process by sentencing him on the basis of a risk assessment algorithm for several reasons. *State v. Loomis*, 881 N.W.2d 749, 757 (Wis. 2016). Only one of those reasons involved a claim of discrimination, though not on racial grounds: Loomis claimed that the assessment tool’s consideration of *gender* constituted a due process violation. *Id.*

The Wisconsin Supreme Court decided against Loomis. Most notably, on Loomis’ claim of gender discrimination, the court found that the risk assessment tool’s use of gender “promotes accuracy that ultimately inures to the benefit of the justice system including defendants.” *Id.* at 767. *Loomis* thus supports the notion that courts will ultimately look favorably on arguments in favor of the use of factors that improve the accuracy of risk assessment, even if those factors are in some way constitutionally suspect. Ultimately, current legal doctrine afford defendants no clear way they can challenge the use of a biased risk assessment tool.

II. EMPIRICAL STUDIES SUGGEST THAT RISK ASSESSMENT TOOLS’ ABILITY TO IMPROVE BAIL-SETTING PRACTICES IS AT BEST LIMITED, AND AT WORST COUNTER-PRODUCTIVE.

The Special Master expressed the hope that a risk assessment tool might help arraignment court magistrates “accurately assess whether [a] defendant presents a community danger.” (Special Master’s Report at 18.) He notes that a proper risk assessment tool is an instrument “to be used by an ACM in evaluating the factors which the law requires must be considered.” *Id.* However, studies of the efficacy of risk assessment tools have cast doubt on whether they actually help improve judicial decisionmaking. For example, an empirical study comparing the recidivism predictions of human nonexperts recruited in an online survey to predictions made by the popular COMPAS risk assessment tool found that the nonexperts were as accurate as COMPAS at predicting recidivism. Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, *Science Advances* (Jan. 17, 2018), <https://bit.ly/37gdxuF>. Furthermore, while COMPAS uses 137 factors to come to its prediction of recidivism, the human nonexperts made their decisions with access to only *seven* factors—sex, age, charge, degree-of-crime, the number of non-juvenile priors, the number of juvenile felonies, and the number of juvenile misdemeanors. *Id.*

Very few studies have examined risk assessment tools in the hands of actual judges at a large scale. *See* Desmarais & Lowder, *supra*, at 7. One of the few such studies is on Kentucky’s mandatory statewide use of pretrial risk assessment tool, and the findings are inconclusive on the ultimate value of pretrial risk assessment

tools. *See* Megan Stevenson, *Assessing Risk Assessment in Action*, 103 Minn. L. Rev. 303, 309 (2018). While judges in Kentucky initially released more “low risk” defendants on non-monetary bail conditions, over time judges “returned to their previous bail-setting practices.” *Id.* A study of recidivism risk assessment in Virginia (but used in the felony sentencing context) yielded similar results: there, the use of a new risk assessment declined over time, possibly as judges came to find the tool not useful. Megan T. Stevenson & Jennifer L. Doleac, *Algorithmic Risk Assessment in the Hands of Humans* 31 (IZA Inst. of Labor Econ. DP No. 12853, 2019). These results demonstrate the importance of proceeding cautiously when developing and deploying risk assessment tools.

III. THE BAIL SYSTEM IMPROVEMENTS RECOMMENDED BY THE SPECIAL MASTER CAN BE MADE WITHOUT THE ADOPTION OF PRETRIAL RISK ASSESSMENT TOOLS.

The Special Master identifies eight points of agreement between all parties on how to improve Philadelphia’s bail system, and makes nine additional suggestions of his own. (Special Master’s Report.) Those recommendations can be followed without the implementation of a risk assessment tool. *See* Koepke & Robinson, *supra*, at 1792–1806 (discussing how risk assessment tools are unnecessary to many important bail improvements). Further, it was only in May of 2019 that the concerns of both the District Attorney’s Office and Public Defender

over risk assessments derailed an earlier effort to adopt a risk assessment tool under the MacArthur Safety and Justice Challenge grant. *See* Samantha Melamed, *Will Controversy Over Risk Assessments Break Philly's Touted Criminal-Justice Reform Collaboration?*, Phila. Inquirer (May 8, 2019), <https://bit.ly/30KDVe0>. That earlier tool had been in development for over six years. It is highly likely that a new effort to develop a risk assessment tool in Philadelphia will face the same opposition that risk assessment tools have always faced for the same reasons *amicus* has raised. In the meantime, before the state of the art of risk assessment tools and the like are advanced enough to benefit Philadelphia on an accurate and equitable basis, the Court has the ability to make a range of other independent improvements to Philadelphia's bail system.

CONCLUSION

Because risk assessment tools have been shown repeatedly to have the capacity to discriminate against defendants on the basis of race, we urge the Court to not pursue development of a new pretrial risk assessment tool at this time.

Respectfully submitted,

Date: January 30, 2020

/s/ Rashida Richardson, Cal. ID No. 282565

Jason M. Schultz, N.Y. ID No. 5376520

AI NOW INSTITUTE

155 Avenue of the Americas, Floor 4

New York, NY 10013

(917) 608-6652

rashida@ainowinstitute.org

/s/ Sara E. Jacobson

Sara E. Jacobson, Pa. ID No. 80965

TEMPLE UNIVERSITY,

BEASLEY SCHOOL OF LAW

1719 North Broad Street, Room 614

Philadelphia, PA 19122

(215) 204-0661

Sara.jacobson@temple.edu

Counsel for Amicus

CERTIFICATES

Certification Pursuant to Rule 531

I hereby certify that no person or entity other than staff of the AI Now Institute (AI Now) has: (1) paid in whole or part for the preparation of the *amicus curiae* brief filed by AI Now in this matter, or (2) authored, in whole or in part, the *amicus curiae* brief filed by AI Now in this matter.

Dated: January 30, 2020

/s/ Rashida Richardson
Rashida Richardson

Certification of Word Count

I hereby certify that this brief contains 5242 words, as determined by the word-count feature of Microsoft Word, the word-processing program used to prepare this petition.

Dated: January 30, 2020

/s/ Rashida Richardson
Rashida Richardson

Certificate of Compliance with Pa.R.A.P. 127

I hereby certify, pursuant to Pa.R.A.P. 127, that this filing complies with the provisions of the Public Access Policy of the Unified Judicial System of Pennsylvania: Case Records of the Appellate and Trial Courts that require filing

confidential information and documents differently than non-confidential information and documents.

Dated: January 30, 2020

/s/ Rashida Richardson
Rashida Richardson