Layered Opacity: Criminal Legal Technology Exacerbates Disparate Impact Cycles and Prevents Trust

Ben Winters*

Introduction

Disparate impacts based on race, gender, socioeconomic status, and ethnicity from algorithmic Statistical Analysis Tools ("tools") manifest in, among others, day-to-day law *enforcement* (police forces patrolling, stopping people, making arrests, and being present) and Risk Assessment Tools employed throughout the broader criminal legal cycle in the United States.

In law enforcement, tools such as Predictive Policing and ShotSpotters inform the allocation of resources in patterns that perpetuate rather than confront bias. Tools such as Facial Recognition, Automated License Plate Readers, and both public and private surveillance systems exacerbate those effects. Individuals in over-policed areas continue to be stigmatized, targeted, and arrested more as a result. Predictive policing tools obfuscate the distinction between arrest data and offense data and create a multi-faceted self-fulfilling prophecy. The results of Risk Assessment Tools are inextricably linked to this trend because certain inputs into these tools, which can lead to harsher treatment at every stage for an accused

[*] Ben Winters is Counsel at the Electronic Privacy Information Center (EPIC). Research supporting this piece was largely supported by the Philip M. Stern Foundation through an Equal Justice Works Fellowship completed at EPIC from 2019-2021. This piece reflects developments in the law through 2020.

individual, include neighborhood arrest data and socioeconomic figures, in addition to other proxies for protected classes.

The developers of these tools conceal the inner-workings of their programs, often embracing over-broad trade secret protections and the culture of opacity in technology and government. The "black box" remains hidden not only from the public but often from the agencies employing them.[1] This opacity diminishes accountability, transparency, trust, and the exercise of a complete criminal defense, to the particular detriment of defendants in protected classes. It also embraces and encodes, rather than confronts, the reasons these biased effects exist. Advocates work to achieve the most basic levels of transparency regarding systems used in a given jurisdiction to some success, but the burden should fall on the users and purchasers of the technologies to articulate and publish the purpose of these tools, make them transparent, and evaluate the propriety of their use.

The specific concerns and solutions to ameliorate negative effects vary from tool to tool, but they both operate in, reflect, and support the same biased system. Both predictive policing tools and risk assessment tools effectively criminalize poverty as a result of the factors they use to make the determination they aim to make.[2]

As research using both types of tools continue to show these disparate impacts, scholarship on algorithmic accountability and governance has ballooned, forcing

[1] *See, e.g.*, State v. Loomis, 881 N.W.2d 749, 761 (Wis. 2016) ("Northpointe, Inc., the developer of COMPAS, considers COMPAS a proprietary instrument and a trade secret. Accordingly, it does not disclose how the risk scores are determined or how the factors are weighed. . .. Thus, to the extent that Loomis's risk assessment is based upon his answers to questions and publicly available data about his criminal history, Loomis had the opportunity to verify that the questions and answers listed on the COMPAS report were accurate. Additionally, this is not a situation in which portions of a PSI are considered by the circuit court, but not released to the defendant. The circuit court and Loomis had access to the same copy of the risk assessment.").

[2] This piece uses predictive policing systems and risk assessment tools as two exemplars used by government actors to illustrate the degree of interconnectedness. Many of the same principles discussed here are transferrable and applicable to other technologies that vary greatly in who uses them, who develops them, their purpose, and use procedures. *See also* TAWANA PETTY, MARIELLA SABA, TAMIKA LEWIS, SEETA PEñA GANGADHARAN & VIRGINIA EUBANKS, OUR DATA BODIES: RECLAIMING OUR DATA INTERIM REPORT (2018). In *Our Data Bodies*, the authors feature excerpts from interviews describing the "cycle of injustice":

> *The collection, storage, sharing, and analysis of data as part of a looping cycle of injustice that results in diversion from shared public resources, surveillance of families and communities, and violations of basic human rights. Connected to the experience of power and powerlessness, the theme of "set-up" concerns how data collection and data-driven systems often purport to help but neglect and fail Angelinos. Interviewees described these set-ups as "traps" or moments in their lives of being forced or cornered into making decisions where human rights and needs are on a chopping board. When using social services to meet basic needs or expecting that a 9-1-1 call in an emergency will bring health and/or safety support into their homes or communities, our interviewees spoke about systems that confuse, stigmatize, divert, repel, or harm. These systems—or the data they require—give people the impression of helping, but they achieve the opposite. They ask or collect, but rarely give, and that leads to mistrust, disengagement, or avoidance. Furthermore, systems perpetuate violent cycles when they are designed to harm, criminalize, maintain forced engagement. Id. at 20.*

the question of whether tools like these can be used equitably in any form. The competing ideas of what "ethical" use of automated decision-making systems means in any sector must be more focused and applied to public sector use, especially in Criminal Justice where fundamental rights are directly at stake. Although these tools are rarely the complex machine learning algorithms one might think of when discussing AI, the principles coming from this field are directly applicable to the types of automated decision-making tools used in and around the criminal justice system.

This piece will connect findings of Criminal Justice researchers to research about algorithmic harm generally in order to demonstrate a key relationship between predictive policing tools and risk assessment tools. The interrelatedness of disparate automated decision-making systems is central to understanding how the ecosystem of criminal justice technology harms poorer poople and people of color disproportionately. Part I lays out the two main tools discussed in this piece, with case studies laying out examples and explorations of how trust is eroded by the singular use of them: predictive policing tools and risk assessment tools. Part II discusses the opacity issues common to these tools and similarly situated tools, and how trust problems with the police are exacerbated by their use, even if they were entirely transparent. Part III will articulate the negative feedback loop between enforcement supported by the policing tools and risk assessments. Part IV will lay out a small survey of applicable algorithmic ethics principles and explore how they have and should be combined to increase public trust in the institutions that comprise the criminal justice system.

I. THE TOOLS

Many police departments, courts, and prisons use statistical analysis tools such as Predictive Policing Systems and Risk Assessments to make critical decisions. Some of these are more complicated than others, ranging from Machine Learning Algorithms to a series of inputs that are tallied together. One constant is that statistical analysis, inferences, political decisions and assumptions about a given community dictate the inputs, weight of each input, the logic of the system, and the outputs of the tools.

In most cases, both governments employing these systems and the contractors that create the tools go to great lengths to keep details about their tools opaque. Governments do their part in ensuring this opacity in awarding contracts, allowing restrictive contract terms, withholding public records to the greatest extent allowable under state law – and while many of these would ideally be changed by legislative *requirements*, it should not be understated how much power entities contracting these tools hold in procurement decisions. Customers of criminal justice technologies are police departments and other government entities, not the people who feel the negative effects, which creates a distance from control and understanding about the important details about those systems.

The details most often shielded from meaningful public scrutiny include the factors considered in the tools, the research supporting the use of those factors, and the weight of those factors. There is how government entities within the criminal justice cycle are *procuring* technology; how they're *using* it; what the purpose of adopting this technology is and how the adoption is justified; how data collection, management, and sharing is carried out within an agency; how is the effectiveness of the tool evaluated and more. The opacity in the decision-making process around these tools is layered and exacerbated, and at each point, public trust and public safety are compromised. This diminishes accountability, transparency, trust, and the exercise of a complete criminal defense, to the particular detriment of defendants in protected classes and their counsel. These tools are attempting to automate inherently difficult and dangerous aspects of decision-making processes. In doing this, it is assigning values and making political decisions about who is treated as dangerous.

A. Predictive Policing

Predictive Policing tools such as PredPol, RAND, or Police One serve to perpetuate, rather than confront racial and ethnic bias in enforcement patterns. Certain tools used in policing, such as Facial Recognition Tools[3] and Automated License Plate Readers,[4] add the direct risk of disparate impacts on racial and gender minorities. Over-policed areas continue to be over-policed and treated as higher risk – which creates a multi-faceted self-fulfilling prophecy.

Predictive Policing is "any policing strategy or tactic that develops and uses information and advanced analysis to inform forward-thinking crime prevention."[5] Predictive policing comes in two main forms: location-based and person-based. Location-based predictive policing works by identifying places of repeated property crime combined with other data and trying to predict *where* they would occur next, while person-based predictive policing aims to pinpoint *who* might be committing a crime – trying to measure the risk that a given individual will commit crimes. Both are used in different jurisdictions and use past policing data as the main driver of these predictions, necessarily creating a cycle of arresting resources. The Bureau of Justice Assistance is a top provider of large grants to Police departments around the country to create and pilot these programs.[6]

Beyond predictive policing per se, police use a variety of algorithmic tools that help the police "assess risk" in patrolling and make other operating decisions.

[3] In particular, women of color had the greatest risk for false positives. *NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software*, NAT'L. INST. OF STANDARDS & TECH. (Dec. 19, 2019), https://perma.cc/F3HQ-NE36.

[4] Kansas v. Glover, 140 S. Ct. 1183 (2020); Brief for Electronic Privacy Information Center et al. as Amici Curiae Supporting Respondents at 15, Kansas v. Glover, 140 S. Ct. 1183 (2020) (No. 18-556).

[5] CRAIG D. UCHIDA, A NATIONAL DISCUSSION ON PREDICTIVE POLICING: DEFINING OUR TERMS AND MAPPING SUCCESSFUL IMPLEMENTATION STRATEGIES 1 (National Institute of Justice ed., 2009) ("Predictive policing refers to any policing strategy or tactic that develops and uses information and advanced analysis to inform forward-thinking crime prevention.").

[6] *EPIC v. DOJ (Criminal Justice Algorithms)*, ELEC. PRIV. INFO. CTR., https://perma.cc/WLM5-HL6H.

These include facial recognition matching systems such as Clearview AI,[7] ShotSpotter, live feeds from Ring doorbell cameras,[8] and private data bought from third parties.[9]

In a 2014 report from the Department of Justice, received through a Freedom of Information Act lawsuit,[10] the agency cautioned President Obama that "individual liberty is at stake" with the use of statistical analysis throughout the Criminal Justice System. The DOJ warned that static factors should not be used in these tools nor risk assessment tools, stating specifically that "these factors [immutable characteristics unrelated to the criminal conduct at issue, such as a defendant's education level, socioeconomic status, or neighborhood of residence] may unintentionally exacerbate unjust disparities in our criminal justice system."[11] However, these factors are routinely used in some Risk assessment tools and Predictive Policing 3.0.[12] In that report from 2014, the DOJ framed the following issues as essential for considering predictive analytics:

- Does the use of a given predictive analytics tool lead to an improvement in public safety outcomes when compared to existing law enforcement methods?
- Does the use of a given technique result in a greater or lesser disparate impact on marginalized communities than the use of existing law enforcement methods?
- Can any given technique be modified to further minimize any disparate impact without compromising its predictive value?

[7] They claimed to stop selling to non-law enforcement entities – doubling down on the use of it in enforcement. Nick Statt, *Clearview AI to Stop Selling Controversial Facial Recognition App to Private Companies*, THE VERGE (May 7, 2020, 8:29 PM), https://perma.cc/R4YN-EPSZ.

[8] Jane Wakefield, *Ring doorbells to send live video to Mississippi police*, BBC NEWS (Nov. 5, 2020), https://perma.cc/VG5Z-HQJM.

[9] *See e.g.*, Joseph Cox, *Police Are Buying Access to Hacked Website Data*, VICE (July 8, 2020; 9:29 AM), https://perma.cc/NG8A-9N9K.

[10] U.S. DEP'T OF JUST., PREDICTIVE ANALYTICS IN LAW ENFORCEMENT: A REPORT BY THE DEPARTMENT OF JUSTICE 23 (2014), https://perma.cc/KH4B-FHA6 [hereinafter DOJ Predictive Analysis].

[11] *Id.* **Static factors** are historical factors that generally do not require an interview by a trained professional. The data most commonly associated with this type of factor are past criminal convictions, arrest history, and more. **Dynamic factors** are factors that require interviews and consistently change. They can include factors such as employment, social network, drug use, residence, cell phone ownership, and mental health. A prominent group of criminal defense lawyers expressed that "in order to reduce unnecessary detention and help to eliminate racial and ethnic bias in the outcome of the tool." *Id.*

[12] As named by the author in Andrew G. Ferguson, *Policing Predictive Policing*, 94 94 WASH. UNIV. L. REV. 1109, 1137 (2017) ("Predictive Policing 3.0 rests on the insight that negative social networks, like environmental vulnerabilities, can encourage criminal activity. Also, it involves utilizing big data capabilities to develop predictive profiles of individuals based on past criminal activity, current associations, and other factors that correlate with criminal propensity.").

- Are there certain factors that should not be relied upon by predictive analytics models? Does the answer depend on how the results of the model are used?
- How should techniques be evaluated on each of these questions over time?
- Are there training guidelines, best practices, or other protections that law enforcement agencies can adopt to ensure that predictive analytics are being used in a manner that ensures civil rights protections?

Six years later, most of these questions remain unanswered. As of the memo written in 2014, the National Institute of Justice ("NIJ") alone had "funded more than a dozen law enforcement agencies, researchers, and other entities to develop and implement advance place-based techniques."[13] Millions of dollars have been granted to police departments around the country by the Department of Justice's Bureau of Justice Assistance.[14]

1. Case Study: Chicago Police Department

One example of a project funded by NIJ was one deployed by the Chicago Police Department, who spent over 3.8 million dollars and ten years developing risk models known as the Strategic Subject List (SSL) and Crime and Victimization Risk Model (CVRM).[15]

The SSL, or "heat list," had a primary goal of "rank[ing] individuals with a criminal record according to their probability of being involved in a shooting or murder, either as a victim or an alleged offender, known as a 'Party to Violence' (PTV)."[16] This ranking was calculated and made into a list. Variables included:

- The number of times an individual was a victim of a shooting;
- An individual's age during their most recent arrest;
- The number of times an individual was the victim of an aggravated battery orassault;
- The number of prior arrests an individual had for violent offenses;
- The individual's number of prior narcotics arrests;
- The number of prior arrests an individual had for unlawful use of a weapon;

[13] DOJ Predictive Analysis, *supra* note 10, at 3.

[14] *See* generally Bureau of Justice Assistance Awards, Department of Justice, https://perma.cc/553Y-SN7L (last visited Dec. 9, 2021)

[15] CHI. POLICE DEP'T, SPECIAL ORDER S09-11, SUBJECT ASSESSMENT AND INFORMATION DASHBOARD (SAID) (2019), https://perma.cc/U5GP-E3LQ.

[16] CHI. POLICE DEP'T, SPECIAL ORDER S09-1, STRATEGIC SUBJECT LIST (SSL) DASHBOARD (2016), https://perma.cc/YJ72-YJEG.

- An individual's trend in recent criminal activity;
- An individual's gang affiliation.;
- Before 2017, the criminality of social networks (as defined by co-arrests, wereincluded). [17]

The CVRM was a "statistical model built into the Chicago Subject Assessment and Information Dashboard [SAID] that estimates an individual's risk of becoming a victim or a possible offender in a shooting or homicide in the next 18 months based on risk factors in a person's recent criminal or victimization history."[18]

Their systems were developed with an NIJ grant, which was later continued through the Bureau of Justice Assistance. The police department worked with the RAND Corporation, which the Bureau contracts to evaluate Statistical Analysis tools used throughout the Criminal Justice system. Throughout six variations of the risk modeling system, RAND's evaluations determined that "the SSL did not effectively predict an individual's propensity for gun violence criminality or victimization" and that "the CVRM risk modeling was not operationally effective to assist CPDs crime prevention strategies."[19] Following these results, and the use of a further refined risk algorithm that primarily used prior violent and gun-related charges, a tool like this was still determined not to hold significant utility that justified its use.[20]

On January 23, 2020, the City of Chicago Office of Inspector General released their "Advisory Concerning the Chicago Police Department's Predictive Risk Models," referring to CVRM and SSL. It explained that the Chicago Police used these risk models to, among other things, target and prosecute individuals "with the high propensity toward violent, gang-related crime"[21] and implement the Gang Violence Reduction Strategy which included linking individuals through "network mapping, dissemination of intelligence, information gathering, and analysis" that included SSL. These scores also started to be used as justification for the arrest and included in the narrative sections of arrest reports. The Chicago Inspector General's report voiced concern that there were insufficient access controls, employees tasked with gathering information that the algorithms used were improperly trained, that despite the grant and infrastructure to review the models systematically, "Neither CPD nor the RAND corporation evaluated Versions 2

[17] Brianna Posadas, *How Strategic is Chicago's "Strategic Subjects List"? Upturn Investigates.*, MEDIUM: EQUAL FUTURE (June 22, 2017), https://perma.cc/ZB78-7D69?type=image.

[18] *Id.*

[19] Letter from Dean O'Malley, Gen. Couns. Off. of the Superintendent Chi. Police Dep't., to Joseph M. Ferguson, Inspector Gen. City of Chicago Off. of Inspector Gen., Jan. 7, 2020, https://perma.cc/U8PE-PK62.

[20] *Id.*

[21] CHI. INSPECTOR GEN.'S OFF., ADVISORY CONCERNING THE CHICAGO POLICE DEPARTMENTS PREDICTIVE RISK MODELS 5 (2020), https://perma.cc/6449-KSY4.

through 5 of CPD's PTV risk models."[22] In 2019, the Chicago Police began to phase out the use of these risk models.

Any tool that predicts future crime based on past and current arrest data will be necessarily flawed and biased.[23] Nationwide, police have been shown to arrest black people at a higher rate than white people,[24] stop more black and Hispanic men than white men,[25] and use force on black men at a significantly higher rate than other demographics.[26]

These are the practices that have long been in place in police throughout America, and they are the data points that populate these systems. These systems are the first layer in which a false predisposition and measuring of "risk" become a real risk for these people that have been historically overpoliced (and more violently policed), who are now being continually overpoliced, but under the guise of scientific rigor and the veneer of "analysis" and "data."[27]

Alone, the use of predictive policing erodes the public trust of law enforcement by (1) reinforcing policing patterns in certain neighborhoods that are being overpoliced to start with, (2) limits an ability to address the failure of the state that led to these neighborhoods to be considered and manifested as "higher-crime," (3) lends a false veneer of objectivity and cover for policing behaviors and (4) further removes policing as a function led by people sworn to protect people, rather than a semi-automated state aimed at bringing people into the criminal justice system. Predicting, and then fulfilling that prediction, where arrests will be made, further dehumanizes people seen as threats by law enforcement.[28]

[22] *Id.*

[23] Rashida Richardson, Jason 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 192, 205 (2019), https://perma.cc/2BQP-GXF8.

[24] Radley Balko, *There's Overwhelming Evidence that the Criminal Justice System is Racist. Here's the Proof.,* WASH. POST (JUNE 10, 2020), https://perma.cc/6DJ5-LNB8.

[25] Emma Pierson, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, Phoebe Barghouty, Cheryl Phillips, Ravi Shroff & Sharad Goel, *A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States*, 4 NATURE HUM. BEHAV. 736–745, 736 (2020), https://perma.cc/56TN-3KKA.

[26] *See, e.g*., Frank Edwards, Hedwig Lee & Michael Esposito, *Risk of Being Killed by Police Use-of-Force in the U.S. by Age, Race/Ethnicity, and Sex*, 116 PROC. OF THE NAT'L ACAD. OF SCIS., No. 34, 16793–16798 (2019), https://perma.cc/NP5E-TZ5K; Bethany Bruner & Bill Bush, *Columbus Police Use Force Disproportionately Against Minorities, Study Finds*, COLUMBUS DISPATCH (Aug. 21, 2019, 7: 47 AM), https://perma.cc/AD8V-P4JQ; Richard A. Oppel, Jr. & Lazaro Gamio, *Minneapolis Police Use Force Against Black People at 7 Times the Rate of Whites*, N.Y. TIMES (June 3, 2020), https://perma.cc/5ZX7-MFAC.

[27] Will Douglas Heaven, *Predictive Policing Algorithms Are Racist. They Need to be Dismantled.*, MIT TECH. REV. (July 17, 2020), https://perma.cc/DS5L-JQRD ("But there is an obvious problem. The arrest data used to train predictive tools does not give an accurate picture of criminal activity. Arrest data is used because it is what police departments record. But arrests do not necessarily lead to convictions.. . . arrest data encode patterns of racist policing behavior. As a result, they're more likely to predict a high potential for crime in minority neighborhoods or among minority people. Even when arrest and crime data match up, there is a myriad of socioeconomic reasons why certain populations and certain neighborhoods have higher historical crime rates than others. Feeding this data into predictive tools allows the past to shape the future.").

[28] *See generally,* Laura Moy, *A Taxonomy of Police Technology's Racial Inequity Problems*, 2021 U. ILL. L. REV. 139 (2021), https://perma.cc/SAP3-K37U (providing fuller explanation of the different types of harm in policing technology).

B. Risk Assessment Tools

All predictive policing tools and risk assessment tools purport to assess risk. The tools vary but pretrial risk assessment tools commonly purport to estimate using "actuarial assessments" (1) the likelihood that the defendant will re-offend before trial ("recidivism risk") and (2) the likelihood the defendant will fail to appear at trial ("FTA").[29] Those tools contribute to decisions around detention and bail. Risk assessment tools are also used outside of pre-trial, which are often treated by proprietary techniques are used to determine sentencing, how individuals are treated in prison, parole, and contribute to determinations about guilt or innocence.[30]

There are myriad variables that differ from tool to tool, which include but are not limited to:

- inputs included in the algorithm;
- the weight of each input;
- how these inputs are gathered (only through static means, which would not require an interview, or would it include information that is more subjective and gathered via interview);
- how those individuals are trained;
- how the decision that the tool comes to is used;
- the security of the system;
- the data retention, sharing, and maintenance policies of the system.

Overbroad trade secret protections and assertions as well as inadequate public information and procurement laws about the jurisdictions using these tools exacerbate the public trust problem and heighten the risk for everyone, in particular those who are already somewhere in the criminal justice cycle.

1. Case Study: Idaho Department of Corrections

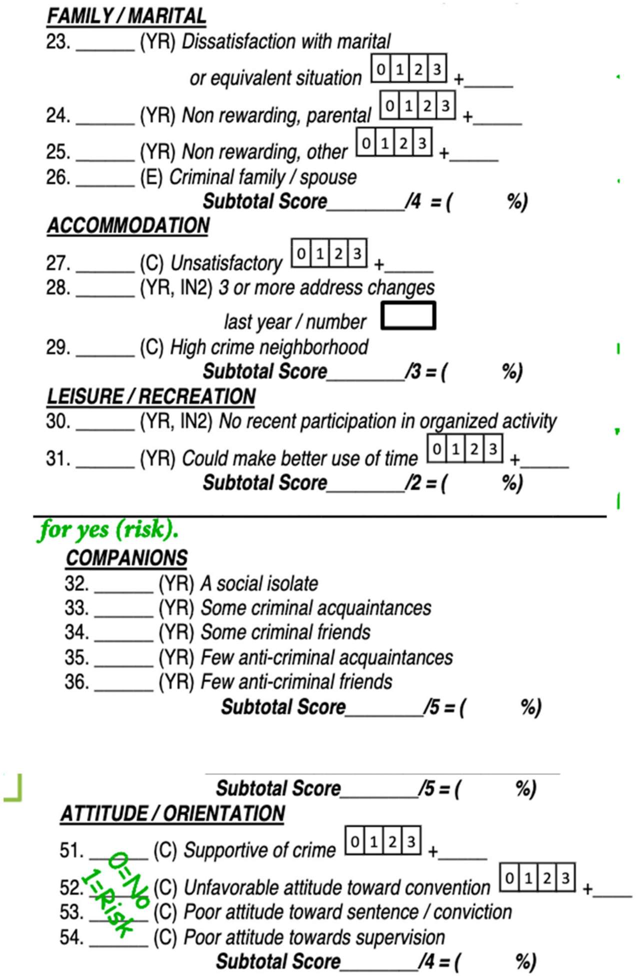
Certain inputs into risk assessment tools, which can lead to harsher treatment at every stage for an accused offender, are based on zip code, personal arrest data, and neighborhood arrest data. Many risk assessment tools go farther, though. One Idaho Department of Corrections risk assessment uses age, sex, geography, family background, employment status, and highly subjective categories such as alleged "criminality" in social network and perceived "attitude towards authority."[31]

[29] *AI and Human Rights: Criminal Justice System*, ELEC. PRIV. INFO. CTR., https://perma.cc/3G5Y-76XN.

[30] State v. Loomis, 881 N.W.2d 749, 761 (Wis. 2016); *See generally* Alyssa M. Carlson, Note, *The Need for Transparency in the Age of Predictive Sentencing Algorithms*, 103, https://perma.cc/5RQU-9NPD

[31] *Idaho Department of Corrections LSI-R Annotated Scoresheet*, ELEC. PRIV. INFO. CTR., (Nov. 21, 2019), https://perma.cc/6TB6-LZJW.

This Level of Service Inventory-Revised tool is used widely throughout the U.S.[32] and Canada.[33] Generally, their training is contracted out but is their goal to have 80% "inter-rater" reliability (i.e. consistency between the interviewers) and teach the workers to get 54 yes/no questions on incredibly nuanced questions within the one-hour slot they are allotted. The scoring guide and training materials illuminate this further.



FAMILY / MARITAL

23. ______ (YR) *Dissatisfaction with marital or equivalent situation* 0 1 2 3 +_____

24. ______ (YR) *Non rewarding, parental* 0 1 2 3 +_____

25. ______ (YR) *Non rewarding, other* 0 1 2 3 +_____

26. ______ (E) *Criminal family / spouse*

Subtotal Score_______/4 = (%)

ACCOMMODATION

27. ______ (C) *Unsatisfactory* 0 1 2 3 +_____

28. ______ (YR, IN2) *3 or more address changes last year / number* □

29. ______ (C) *High crime neighborhood*

Subtotal Score_______/3 = (%)

LEISURE / RECREATION

30. ______ (YR, IN2) *No recent participation in organized activity*

31. ______ (YR) *Could make better use of time* 0 1 2 3 +_____

Subtotal Score_______/2 = (%)

for yes (risk).

COMPANIONS

32. ______ (YR) *A social isolate*

33. ______ (YR) *Some criminal acquaintances*

34. ______ (YR) *Some criminal friends*

35. ______ (YR) *Few anti-criminal acquaintances*

36. ______ (YR) *Few anti-criminal friends*

Subtotal Score_______/5 = (%)

Subtotal Score_______/5 = (%)

ATTITUDE / ORIENTATION

51. ______ (C) *Supportive of crime* 0 1 2 3 +_____

52. ______ (C) *Unfavorable attitude toward convention* 0 1 2 3 +____

53. ______ (C) *Poor attitude toward sentence / conviction*

54. ______ (C) *Poor attitude towards supervision*

Subtotal Score_______/4 = (%)

0=No
1=Risk

Screenshots of a portion of LSI-R Scoresheet

[32] *See, e.g.* Christopher T. Lowenkamp & Kristin Bechtel, *The Predictive Validity of the LSI-R on a Sample of Offenders Drawn from the Records of the Iowa Department of Corrections Data Management System*, 71 FED. PROB. 3 (2001), https://perma.cc/7RHV-2F3M.

[33] KELLEY BLANCHETTE, CLASSIFYING FEMALE OFFENDERS FOR CORRECTIONAL INTERVENTIONS, 9 FORUM ON CORRECTIONS RESEARCH NO. 1 (1997), https://perma.cc/Q4PC-EGGW.

C. Issues Common to All of These Tools

1. Bias

Predictive Policing and Risk Assessments categorize an individual's perceived risk based on immutable characteristics such as race, gender, socioeconomic status, age, and ethnicity. This perception of risk is then given false legitimacy by its label of "data-driven" analysis and use by government entities.

The relationship between policing and arrest data necessarily informs the calculations that these tools reach. Using data from historically racially disparate policing patterns turns racially imbalanced arrest data into "offense" data, reflecting who is more likely to be stopped by a policeman rather than *commit* a crime – which is an inherently unascertainable statistic. Statistical analysis tools like these and others which the DOJ had reservations about in 2014 yet continually fund and exacerbate these issues. This tension is complicated and cannot be easily quantified. The DOJ wrote that "the length of a defendant's prison term should not be adjusted simply because a statistical analysis has suggested that other offenders with similar demographic profiles will likely commit a future crime."[34] Still, a vast majority of tools that remain successfully opaque and behind trade secret protections to particular defendants and other similar commercial protections to all people via open government laws contribute to the length of a defendant's prison term and are based on these points.

The use of this historical data is just one way in which bias can occur in these systems. One report articulates six ways that risk assessment tools can negatively affect black defendants: bias in the data, bias in a given predictive model, bias from differential censoring, bias introduced when selecting thresholds for risk categories, and bias introduced by relying on factors that override risk-assessment scores.[35]

Without diving deep into the findings of each piece of research in this paper, there is significant empirical and anecdotal data supporting allegations of bias in both types of tools explored in this paper.[36]

[34] DEP'T. OF JUST., PREDICTIVE ANALYTICS IN LAW ENFORCEMENT: A REPORT BY THE DEPARTMENT OF JUSTICE 23 (2014), https://perma.cc/WK8T-AEA9. DOJ Predictive Analysis, *supra* note 10, at 23.

[35] Kristin Porter, Cindy Redcross & Luke Miratrix, *Balancing Promise and Caution in Pretrial Risk Assessments*, MEDIA & DEMOCRACY RES. CTR. (May 2020), https://perma.cc/42SA-JVYS.

[36] *See, e.g.,* Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks.,* PROPUBLICA (May 23, 2016), https://perma.cc/8X36-QMKM; Massimo Calabresi, *Exclusive: Attorney General Eric Holder to Oppose Data-Driven Sentencing,* Time (July 31, 2014, 10:35 AM), https://perma.cc/VUH7-R5FU; Megan Stevenson, *Assessing Risk Assessment in Action,* 103 MINN. L. REV. 303 (2018); Melissa Hamilton, *The Biased Algorithm: Evidence of Disparate Impact on Hispanics,* 56 AM. CRIM. L. REV. 1553 (2019); Megan T. Stevenson & Christopher Slobogin, *Algorithmic Risk Assessments and the Double-Edged Sword of Youth,* 96 WASH. U. L. REV. (2018); Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification,* 81 PROC. OF MACH. LEARNING RESEARCH 77 (2018); Songül Tolan, Marius Miron, Emilia Gómez & Carlos Castillo, *Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia,* 17 PROC. OF THE SEVENTEENTH INT'L CONF. ON A.I. AND L. 83 (2019); Will Douglas Heaven,

In 2020, organizations that had been at the forefront of promoting the tools began to voice caution against their use. The Pretrial Justice Institute explained their change in position, saying they "can no longer be a part of our solution for building equitable pretrial justice systems. Regardless of their science, brand, or age, these tools are derived from data reflecting structural racism and institutional inequity that impact our court and law enforcement policies and practices. Use of that data then deepens the inequity."

2. Accuracy, Transparency & Training Concerns

In e-mails between the Nebraska Department of Corrections Executive Officer and the developer of their STRONG-R RAT, serious flaws with these programs were detailed. Several concerns about the validity and reliability of the STRONG-R assessment results have been raised by NDCS staff members. "Unresolved issues that I have personal involvement with surround the validity of the tool, itself:

- There are errors in how the "Severity Index" of specific crimes is coded in the Vant4ge software. These errors affect the final risk and need score calculations produced by the assessment.
- Some offenses are not mapped to the appropriate questions. For example, a prior criminal conviction for "Arson 1st Degree" scores on a question that relates to prior assault convictions, not prior arson convictions. Some offenses are mapped to the appropriate questions but do not have the appropriate index score/weight assigned. (Legal/ Legislative).
- These issues can only be fixed by a comprehensive review of all state statutes to determine whether (a) the substantive language of the law matches the crime description associated with each severity index score and (b) the offenses are mapped to the appropriate question in the STRONG-R. (Legal/Legislative).
- The Criminal Conviction Record (CCR) software includes only state statutes and does not allow staff to select any city ordinance violations. Because of this, staff have either entered these convictions or have used a state statute that they deem to be the best proxy.
- In a number of cases, staff have entered "Official Misconduct" in the CCR because "Disorderly Conduct" is not an available option. However, "Official Misconduct" is qualitatively different from the convicted offense because it refers to malfeasance by a public official within his or her job capacity.

Predictive Policing Algorithms Are Racist. They Need to Be Dismantled., MIT TECH. REV. (July 17, 2020), https://perma.cc/S3A3-GQN2.

- Both the missing offense codes and the substitution of proxy offense codes have an unknown effect on the calculated risk and needs scores produced by the tool.
- There has not been consistency in how the STRONG-R training is delivered, either among NDCS trainers or between NDCS and Vant4ge trainers. However, we are resolving this issue by revising new user training for all users across NDCS and Parole." (emphasis added).[37]

This illuminates the simple known fact that technology used by governments in the Criminal Justice system is vulnerable to the same risks, downsides, bugs, and errors that any other application or device is. This is expected, and perhaps unavoidable. However, the regulatory environment in most jurisdictions does not adequately account for the impact. The stakes are too high – rather than being unable to access an app on a phone as a result of a bug, constitutionally protected freedoms are risked.

Not only individuals subject to these tools are left under-informed. Even some jurisdictions adopting these tools themselves face the opacity problems.[38] This result stems from procurement policies that are insufficient to address the magnitude of government-contracted systems that directly impact an individual's liberty, an adversarial criminal justice system, and a lack of regulation around government automated decision-making.[39] An important caveat is that these two categories of tools are not the only tools that these discussions apply to, but rather probabilistic genotyping, surveillance infrastructure maintained by both corporate and government actors, gunshot detectors, gang databases, and more.

II. THE OPACITY PROBLEMS COMMON TO THE TOOLS, AND WHY THEY ERODE TRUST

A majority of tools used in the U.S. are not created by the government entity themselves, but are contracted out. This model of providing government services via contractor is pervasive and varied.[40] In some states, the State Supreme Court recommends and contracts with a developer.[41] Other times, an individual at a

[37] State v. Loomis, 881 N.W.2d 749, 761 (Wis. 2016).

[38] *Id.*

[39] The term automated decision-making, for this paper, refers to a system that helps facilitate a decision. There is still usually a requirement of human action to carry out the recommendation of most of these systems or to use them in certain nefarious ways.

[40] *See, e.g.,* U.S. v. Curry, 965 F.3d 313, 338 (4th Cir. 2020); Table of *Risk Assessment Tools State-by-State*, ELEC. PRIV. INFO. CTR., https://perma.cc/BME6-WVSJ.

[41] *See*, *e.g*, *Order Adopting Statewqide Use of the Nevada Pretrial Risk Assessment*, Supreme Court of the State of Nevada (2019), available at https://perma.cc/LD5S-WSB6

Department of Corrections or Public Safety contracts with a developer via an Request For Proposal process.[42] For policing tools, the software has been frequently purchased or developed with grants from the Bureau of Justice Assistance.[43] In some cases, it is legislatively required and adopted state-wide – mostly, though, these tools are adopted piecemeal in a given jurisdiction. All of this is to say, the path towards procurement is not a monolith, but has continually grown and diversified over time.

The ways advocates, defense counsel, and the public finds out about the harms behind the use of these tools is often a news story publicizing a particular case of harm caused by automated decision-making, advocacy organizations using public records requests to try to ascertain the details around a given system, or a person is subject to these tools and has the resources both emotional and financial to challenge them for the first time. Regarding a person being subject to it, unless they are explicitly and negatively impacted by a tool, many automated decision-making tools are "invisible technologies."[44] Predictive policing tools are a clear example of an invisible technology – akin to Medicare benefits and face analysis for employers during job interviews.

Invisibility is not created equal, though. Although automated decision-making tools that are both kinds of invisible can be harmful, they are worth distinguishing. Firstly, there is invisibility where an individual does not know at all that an automated system either made or contributed to a decision that affected them directly. This is seen in many tools in assigning health benefits, predictive policing, or something like an Airbnb rental request. Second, there are degrees of a softer version of invisibility, which can give the illusion of transparency. This can be where you either know, have been aware that similar tools have been used in other jurisdictions, you have a general knowledge that it might be used, or there is a surveillance infrastructure like cameras installed where it is unclear if facial recognition software will be used using it and if other systems are connected to it. This last one has increased with increased awareness of these tools and can be dizzying for residents.

For most of these tools, especially in the Criminal Justice context, but also in the government benefits and employment context, the individual has little or no power of choice over whether the automated decision-making tools will be used, or power over the accuracy of the data collected and analyzed.[45] Transparency would allow for public analysis and advocacy around the tools used in their

[42] *See*, e.g., Joseph O'Sullivan, *State won't renew Corrections contract with company criticized by GOP*, Seattle Times (June 15, 2016), https://perma.cc/Q9U5-W575

[43] *See, e.g.*, BUREAU OF JUST. ASSISTANCE, U.S. DEP'T OF JUST., BJA-2020-17273, FY 2020 GULF STATES LAW ENFORCEMENT TECHNOLOGY INITIATIVE (2020); *see generally* BUREAU OF JUST. ASSISTANCE, U.S. DEP'T OF JUST., AWARDS (2021), https://perma.cc/C4J3-9ZRE.

[44] *See, e.g., Invisible Technologies Are Making Critical Decisions About Us. Here's How to Identify Them.* ACLU OF WASH. (May 26, 2020), https://perma.cc/75T5-DVZE.

[45] *See, e.g. Id.* Some jurisdictions, such as Vermont, give individuals a choice over whether to have their risk score assessed in the bail context. *13 V.S.A. § 7554c*

communities, and it has the opportunity to increase the public trust in these systems. It also would lead to an increased sense of choice – a sense that is diminishing along with the increase of private surveillance systems used in public contexts,[46] yet materializing when given a choice automated decision-making tools being adopted.[47]

Opacity in this context can be categorized into coming from three government sources: procurement regulations, open government laws, and trade secret privileges in evidentiary practices.[48]

Procurement laws define how a given levl of government contracts services, and they differ greatly from state to state. As of now, without complementary regulation of automated decision-making procurement, more transparent and responsible procurement policies can meet the current need for transparency and oversight. It is a field of law that is ripe for updating given the unceasing automation of the administrative state at every level. It is through the procurement process that agreements with contractors that give this level of deference are accepted, where hundreds of thousands of dollars are spent on a given tool, and where changes can be made to ensure transparency and other forms of public oversight. One good example of how regulation aimed at curbing harm caused by automated decision-making used procurement to increase transparency and oversight was proposed in Washington state in 2020. The proposed legislation introduced in both the House and Senate included the responsibility of a "public agency intending to develop, procure, or use an automated decision system [which] must produce an algorithmic accountability report for that system," which must then publish and consider public comment for a minimum of thirty days.[49] The bill goes on to explain that this report must include "clear and understandable statements of," among other things, the vendor; the inputs; how data is collected and processed; a statement of purpose; what decisions it will make or aid; what it'is intended benefits are; data security and management policies; and any potential violations of civil liberties or cause of disparate impacts, coupled with a mitigation plan.[50] Still, depending on the tool and the jurisdiction, it is unclear if these tools will satisfy the definitions of an automated decision-making system. However, regulations like these should aim to explicitly include recommendation systems used in the criminal justice systems through legislative language.

Concerning open government laws, the federal Freedom of Information Act includes an exemption for trade secrets and other commercial IP protections, as

[46] *See, e.g.*, Joseph Cox, *CBP Refuses to Tell Congress How it is Tracking Americans Without a Warrant,* VICE (Oct. 23, 2020, 11:03 AM), https://perma.cc/6DJV-CMGS.

[47] *See, e.g.,* Ellina Abovian, *California rejects Prop 25, keeping cash bail system in place*, KTLA (Nov. 4, 2020, 7:15 PM PST), https://perma.cc/TZC6-8G3P.

[48] *See* Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018) (discussing a more fleshed out explanation of this topic).

[49] S.B. 5116, 66th Leg., Reg. Ses. (Wash. 2020).

[50] *Id.*

well as deliberative process privilege.[51] In a 2019 Supreme Court case, the trade secret protection was expanded, reversing decades of requiring a showing of competitive harm if the "trade secret" was released under the Freedom of Information Act.[52] Now the entity must either prove that they treat the information confidential, or, if it is the type of information that is usually kept confidential and there is express or implied assurance by the government, that they will maintain confidentiality.[53] In Department of Justice guidance after the decision, they explain that if the government has not made any "express or implied indications at the time the information was submitted that the government would publicly disclose this information," then there is a presumption of valid trade secrecy if the entity customarily held the information as private.[54] Open government laws rely on the concept that transparency in government is a virtue embraced beyond how it affects one person. To varying extents, State Open Government laws across the country have similar commercial protections over trade secrets. For these two causes, traditional justifications for trade secret protection should be weighed against the interest at stake: preservation of commercial viability and promoting innovation for policing tools and tools that directly affect bail and sentencing decisions, it quite literally risks people's liberty.

Transparency "silos" reflect what often happens in evidentiary practices.[55] One of the more common ways that defendants gain access to usually nonpublic information about a Criminal Justice technology is through an agreement with a specific defendant fighting to be able to defend themselves, and the company allowing access to information along with a protective order. This functionally recognizes the need for the use of these tools for adequate representation, while refusing to change the general agreement with the public, who has an ever-present risk of being subject to tools used by their government that risk basic liberties.

Opacity is multifaceted and can be fixed with both proactive measures by those using and developing these tools *and* legislative requirements. It weakens public trust and accountability.

III. THE CYCLE: HOW THESE TWO MAIN TYPES OF TOOLS ARE LAYERED AND INTERCONNECTED, AND WHY THEY MUST BE TREATED THIS WAY

Although functionally distinct, these predictive policing tools and pre-trial RATs are inextricably linked. They yield similar concerns, use some of the same inputs, and suffer from similar opacity at both the public records and evidentiary

[51] Freedom of Information Act, 5 U.S.C. § 552(b)(4)-(5).

[52] *Food Marketing Institute v. Argus Leader* Media, 139 S. Ct. 2356 (2019).

[53] *Step-by-Step Guide for Determining if Commercial or Financial Information Obtained From a Person is Confidential Under Exemption 4 of the FOIA*, U.S. DEP'T OF JUST. (Oct. 19, 2019), https://perma.cc/S5U4-KVKL.

[54] *Id.*

[55] Hannah Bloch-Wehba, *Access to Algorithms,* 88 FORDHAM L. REV. 1265 (2020). Bloch-Wehba names the increasingly accepted arrangement between an individual defendant seeking access to a tool used about them specifically and the developer of a tool to have limited access, with a protective order, while still preventing this information from being more broadly available. *Id.*

stages. Moreover, they reflect the greater cycle of incarceration, and criminalization of poverty, among other factors.[56] Predictive policing tools are generally adopted by police departments, while risk assessment tools, used in pre-trial and sentencing, are adopted by the Department of Corrections, the Supreme Court of a state, local courts, or the Attorney General's office. They reflect the same concerns that have pervaded policing, pretrial detention, and incarceration. The more someone is arrested, charged, or policed, the "riskier" they will be labeled as when inputting those data points into the pretrial assessment tool. Critically, even if people around a person has more interactions with the Criminal Justice system, an individual will be treated as riskier. Consequently, for location-based or person-based predictive policing tool, the more someone has been arrested, or the more an area an individual lives in has been policed, the more arrest data will become offense data in a self-fulfilling cycle of predictive policing tools. Analyses of policing data reported to the FBI find black people were arrested at a rate five times more than white people in 2018,[57] and in many cities ten times more.[58] Examples of this are in no short supply but are borne out in examples of cities across the country such as San Francisco, California – where 5.2% of the population is black, but they make up 37.8% of the arrests, or Albany, New York which is 29.9% black with a 70.6% black share of the overall arrest rate.[59]

Even the most limited risk assessment tools that use only static factors and regular independent validation, such as one developed by the University of Alaska and the Alaska Department of Corrections' Pretrial Enforcement Division,[60] or the Public Safety Assessment, developed by Arnold Ventures, uses age, arrests, convictions, and sentencing data from throughout one's life in creating a score aimed at assessing an individual's "risk" of offending (being arrested) again.

The cycle worth identifying here is that predictive policing tools dictate some of the inputs used in different risk assessment tools and exacerbate the biases in both policing *and* determinations made throughout the criminal justice system that is reflected in risk assessment models. The predictive policing tool continues to drive more arrests to communities that have been overpoliced and providing some sort of mythical "data-driven, objective" justification for doing so. And in even the most stripped-down risk assessment tools, which at different stages dictate the treatment someone in the criminal justice system receives, being either a former entrant into the criminal justice system or being "associated with" people who have or living in a certain neighborhood does count against you. The

[56] Press Release, U.S. Dep't of Just., Fact Sheet on White House and Justice Department Convening–A Cycle of Incarceration: Prison, Debt and Bail Practices (Dec. 3, 2015), https://perma.cc/A2W8-BMT2.

[57] Pierre Thomas, John Kelly & Tonya Simpson, *ABC News Analysis of Police Arrests Nationwide Reveals Stark Racial Disparity*, ABC NEWS (June 11, 2020, 5:04 AM), https://perma.cc/8DJC-N8AC.

[58] John Kelly, *Analysis of Police Arrests Reveals Stark Racial Disparity in NY, NJ, and CT*, ABC NEWS (June 10, 2012), https://perma.cc/QU48-47X3.

[59] *See* THOMAS, KELLY & SIMPSON, *supra* note 57; KELLY, *supra* note 58.

[60] Pamela Cravez, *Pretrial Risk Assessment Tool Developed for Alaska,* UAA JUST. CTR. AT THE UNIV. OF ALASKA-ANCHORAGE (2018), https://perma.cc/3Y5H-ASDW.

treatment leads to more difficulty getting jobs, keeping an apartment, owning a phone, maintaining relationships, and treating mental health. Those factors lead to more stringent parole decisions, which bring people closer to being thrown back into the cycle. With few exceptions, the widespread adoption of risk assessment tools and predictive policing tools encodes the system that exists in America, and with more "data," which cannot and should not be treated as a neutral word, can guarantee and accelerate it. The use of these tools is antithetical to the concept of restorative justice and is incompatible with public trust. The permanence of the scores is another concern, with widespread data sharing, data breaches, and the buying and selling of sensitive data between data processors and data holders.

The shortcomings at one point in the criminal justice cycle filled with automated decision-making systems undoubtedly affect the rest of the cycle negatively. In doing so, it makes those impacts harder to identify and remedy.

IV. AUTOMATED DECISION-MAKING AND GOVERNANCE OF THESE TOOLS

The web of non-governmental ethics frameworks and the proliferation of government-endorsed statements of principles regarding the regulation of automated decision-making can be dizzying.[61] Even deciding what word to use when describing the regulation is an important, but difficult decision. Most of the tools discussed in this paper are not what comes to mind when the word algorithm or artificial intelligence is used, but they vary greatly in complexity. As of now, risk assessment tools can be described as an algorithm in that they are a set of inputs that lead to an output. A more accurate moniker for many of these tools that recognize their power and demystify their technological sophistication would be automated decision-making. However, the conversation around regulating artificial intelligence, algorithms, and automated decision-making are complementary. Eight principles came out as predominant in an analysis of frameworks by a variety of stakeholders when focusing the priorities of how these tools should be regulated: privacy; accountability; safety and security; transparency and explainability; fairness and non-discrimination; human control of technology; responsibility; and "promotion of human values."[62]

In this section, I will outline the basic tenets in the field of accuracy, explainability, transparency, and accountability of automated decision-making systems to the extent that they relate to regulating the tools discussed throughout this piece. These are largely interrelated, and actions towards improving one would complement others. The dearth of trust exacerbated by these tools can be ameliorated by addressing these four categories. These four categories, if fulfilled, would not provide a panacea to the social ills the systems they reflect. However,

[61] *See* JESSICA FJELD, NELE ACHTEN, HANNAH HILLIGOSS, ADAM CHRISTOPHER NAGY & MADHULIKA SRIKUMAR, PRINCIPLED ARTIFICIAL INTELLIGENCE: MAPPING CONSENSUS IN ETHICAL AND RIGHTS-BASED APPROACHES TO PRINCIPLES FOR AI (2020).

[62] *Id.* at 4-5.

it would improve oversight, public trust, and allow these tools to be publicly analyzed, litigated, and organized around.

Accuracy in automated-decision systems is important and is usually referenced when describing that a given tool works as intended. Applying to risk assessment tools, this is generally measured by a validation study. Validation studies are used to prove that a model designed to do x can do x. Here, x is that the individuals a model predicts will bring more risk to a community have *in that community, over the time the tool has been used*. One survey of validation of pretrial risk assessments shows that only 45% of jurisdictions surveyed had performed validation studies – and this does not mean they are done by independent entities, done with data points specifically in the locality it is being used, or is being done regularly.[63] Another survey, when interviewing jurisdictions using risk assessment tools, found that 21% of the jurisdictions validated 5-10 years ago, 21% were validated using nonlocal data, 9% used validation studies from over 10 years ago, and only 28% used validation studies using local data within 5 years.[64] For many validation studies, it is also key to point out that better than 50% accurate means accurate – a worrisome bar.[65] Using this as an exemplar, accuracy cannot be taken for granted – it has to mean something, be done independently and regulatory standards of what accuracy means would be an important step forward.

A more holistic concept of accuracy is complementary to accountability and transparency and has to do with what the corresponding purpose of adopting the tool is, and whether the use of these tools has led to success in achieving that goal. One piece points out that implementation of the tools can be the problem: "research demonstrating predictive validity does not equate with research demonstrating implementation success. Indeed, even a well-validated tool may not produce the intended results of more accurate, decarceral, and racially and ethnically equitable decisions relative to practice as usual for many reasons."[66] Measures that would be required, for example, in the proposed Washington statute mentioned above, would require this statement of purpose and method of evaluation that would be equally beneficial to those administering it and the public holding those entities accountable.[67]

[63] *Scan of Pretrial Practices,* PRETRIAL JUST. INST. (Sept. 28, 2019), https://perma.cc/74S3-9X95.

[64] *Validation*, MAPPING PRETRIAL INJUSTICE (Nov. 12, 2020), https://perma.cc/RVV2-K254.

[65] *See* Email from Zachary Hamilton, Assoc. Professor, Wash. State Univ., to Doug Koebernick, Inspector Gen. of Corrs. for the Neb. Legislature (Nov. 2019), (obtained through FOIA Production), https://perma.cc/PWJ8-M62S, ("The primary criterion for creating a validated tool to improve the prediction of recidivism beyond random chance (i.e. a coin flip) . . . one should not simply be concerned that the tool improves beyond random chance but that its prediction is more accurate than any other tool under consideration. Again, I cannot argue that the YLS/CMI has been identified to provide a better prediction than random chance in more places than any other tool. However, we attempted to create the STRONG-R to be more accurate than the YLS/CMI and to customize the prediction for the specific population it is being used to assess.").

[66] SARAH L. DESMERAIS & EVAN M. LOWDER, PRETRIAL RISK ASSESSMENT TOOLS, (Safety + Just. Challenge 2019), https://perma.cc/33UQ-HPSE.

[67] S.B. 5116, 66th Leg., Reg. Ses. (Wash. 2020).

Explainability is essential to trust for automated decision-making – the desire to know why a system made a certain recommendation or decision. Minds differ as to what this means in practice, but the National Institute of Science and Technology ("NIST") recently published a draft white-paper on principles of Explainable AI,[68] in which they provide four helpful aspects of explainability that users of an automated decision-making system, and more importantly regulators, can use as a checklist of sorts to increase meaningful explainability, and consequently increase trust and accountability:

> That the system produce[s] an explanation;
> That the explanation be meaningful to humans;[69]
> That the explanation reflects the system's process accurately;
> That the system expresses its knowledge limits.

These four should be legislatively required and made public clearly and concisely for all automated decision-making systems used by the government, but especially ones that make an impact in the Criminal Justice System. One aspect in which explainability beyond the system itself can be fulfilled is that the way it is implemented is also explainable. For example, requiring a judge, probation officer, or police department to have public policies about how they will use an automated decision-making system, including a weighing of risks with corresponding mitigation efforts.

Accountability can be a vague concept but it encompasses many of the same concepts discussed above. Creating actionable rights for people affected by automated-decision making systems and obligations for entities using them are essential. It should be noted that, as Frank Pasquale articulated, there is a second wave of algorithmic accountability scholars that considers not just how to improve existing systems, but "whether they should be used at all - and, if so, who gets to govern them."[70] Recommendations when coming to automated decision-making systems in the criminal justice systems include that there should be (1) a clear, public publication of the: developer, a narrowly tailored stated purpose of the tool, input data, logic, decision-making matrixes, and data sharing and retention policies; (2) regular, local, independent evaluation that includes specific studies on both efficacy and bias for all protected classes, as well as evaluating the propriety of the tools' use in light of the stated purpose and a requirement of

[68] P. Jonathon Phillips, Carina A. Hahn, Peter C. Fontana, David A. Broniatowski & Mark A. Przybocki, *Four Principles of Explainable Artificial Intelligence* (Nat'l Inst. of Standards & Tech., Working Paper No. 8312, 2020) https://perma.cc/MDH3-D5P6.

[69] *Id* at 2-3 ("A system fulfills the Meaningful principle if the recipient understands the system's explanations. Generally, this principle is fulfilled if a user can understand the explanation, and/or it is useful to complete a task. This principle does not imply that the explanation is one size fits all. Multiple groups of users for a system may require different explanations. The Meaningful principle allows for explanations that are tailored to each of the user groups.").

[70] Frank Pasquale, *The Second Wave of Algorithmic Accountability*, L. & POL. ECON. (LPE) PROJECT (Nov. 25, 2019) https://perma.cc/4EWB-QD3R.

reauthorization for continued use based on certain findings; and (3) the jurisdictions must have minimum technological standards, principles, and policies that include uniform data minimization, deletion, and disclosure policies all oriented to minimize improper reuse of data or data exposure to outside entities. Transparency has been the subject of substantial news coverage and research, as well as Part II of this paper, and is important to consider at both the Open Government stage (to the general public) and the evidentiary stage (to a specifically given defendant that is being subject to the tools). A law in Idaho passed in March 2019 requires "all documents, data, records, and information used by the builder to build or validate the pretrial risk assessment tool and ongoing documents, data, records, and written policies outlining the usage and validation of the pretrial risk assessment tool" to be publicly available.[71] This law allows a party in a criminal case to review the calculations and data underlying their risk score, and precluding trade secret or other intellectual property defenses in discovery requests regarding the development and testing of the tool.[72] Legislation like this is a productive development in light of the case law and the consistent battles for basic levels of transparency in a criminal case, and the satisfaction of basic constitutional principles.[73] There are also efforts to use statewide Artificial Intelligence commissions as a vehicle to obtain centralized information about different automated decision-making systems government-wide in a given jurisdiction. Examples of these are in New York, Vermont, Alabama, and more.[74] However, the value of the transparency aspects of a given commission comes with the power they have to gather that information and mandates to make it public.[75]

In regulating and approaching the procurement of automated decision-making tools in the criminal justice system, the principles focused on the broader AI regulation field can offer guidance. Mandated and culturally promoted transparency, accountability, explainability, and accuracy in these systems would increase public trust in this field of law.

[71] IDAHO CODE. tit. 19 § 1910(1)(b)-(c).

[72] *Id.*

[73] Natalie Ram, *Innovating Criminal Justice*, 112 NW. U. L. REV. 659, 692 (2018) ("Trade secret assertion in the context of criminal justice tools also raises constitutional concerns. The secrecy surrounding the existence, use, and function of criminal justice tools interfere with defendants' and courts' efforts to ensure that the government does not engage in unreasonable searches. Such secrecy is also at least in tension with, if not in violation of, defendants' ability to vindicate their due process interests throughout the criminal justice process, as well as their confrontation rights at trial.").

[74] *See* N.Y.C., N.Y., LOCAL LAW 49 §1(b)(2) (2018); N.Y. ACT, S3971B (2020); VT., ACT 137 (2018); ALA., ACT 269 (2019). *See generally*, *State Artificial Intelligence Policy*, ELEC. PRIV. INFO. CTR., https://perma.cc/MV86-ZA4T.

[75] *See, e.g.*, RASHIDA RICHARDSON, CONFRONTING BLACK BOXES: A SHADOW REPORT OF THE NEW YORK CITY AUTOMATED DECISION SYSTEM TASK FORCE (A.I. Now Inst. Dec. 4, 2019), https://perma.cc/42TV-PB3Z.

CONCLUSION

Considerations of whether these tools are incurable ills or options that can be tweaked to reduce bias, create records for after-the-fact accountability, and proactive changes in teaching and training are incredibly worthwhile. While not the subject of this piece, it is an important question and is inextricable from the questions surrounding what movements towards defunding police and meaningful criminal justice reforms should yield. If in the most basic sense, the carceral state is continually endorsed, with people treated as targets to profile and jail, these tools can be painted as helpful within that system. However, as of now, they should be treated rather as a mirror about those systems of power, the judgments we make as a whole society, and where we can provide support to communities or types of people years of evidence show that are more likely to be put in the hamster wheel of the Criminal Justice system.

Predictive policing tools and risk assessment tools used pre-trial and throughout the criminal justice cycle carry centuries of racial, ethnic, socioeconomic, and age bias with them, in addition to serious transparency and oversight concerns. These deficiencies in each of these tools compound in their interrelatedness, as outputs of one dictate the input of the other. When addressing the regulation of automated decision-making in the criminal justice system and the system that operates within and around it, their interrelatedness has to be reflected in regulation.