HOW TO ARGUE WITH AN ALGORITHM: LESSONS FROM THE COMPAS-PROPUBLICA DEBATE

ANNE L. WASHINGTON, PHD*

The United States optimizes the efficiency of its growing criminal justice system with algorithms. However, legal scholars have overlooked how to frame courtroom debates about algorithmic predictions. In State v. Loomis*, the defense argued that the court's consideration of risk assessments during sentencing was a violation of due process because the accuracy of the algorithmic prediction could not be verified. The Wisconsin Supreme Court upheld the consideration of predictive risk at sentencing because the assessment was disclosed and the defendant could challenge the prediction by verifying the accuracy of data fed into the algorithm.*

Was the court correct about how to argue with an algorithm?

The Loomis *court ignored the computational procedures that processed the data within the algorithm. How algorithms calculate data is equally as important as the quality of the data calculated. The arguments in* Loomis *revealed a need for new forms of reasoning to justify the logic of evidence-based tools. A "data science reasoning" could provide ways to dispute the integrity of predictive algorithms with arguments grounded in how the technology works.*

This article's contribution is a series of arguments that could support due process claims concerning predictive algorithms, specifically the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) risk assessment. As a comprehensive treatment, this article outlines the due process arguments in Loomis*, analyzes arguments in an ongoing academic debate about COMPAS, and proposes alternative arguments based on the algorithm's organizational context.*

* Assistant Professor of Data Policy, Department of Applied Statistics, Social Science, and Humanities. Steinhardt School, New York University. This work draws on my 2016–17 fellowship at the Data & Society Research Institute in New York where I developed the concept of data science reasoning. Dr. David C. Morar, PhD, faithfully tracked the publication of new articles in 2016–17. I am grateful to my fellow D&S fellows Andrew Selbst, Julia Powles, and Rebecca Wexler. Many thanks to Jennifer Eaglin, Frank A. Pasquale, Siona Robin Listokin, David G. Robinson, Paul Ohm, Margaret Hu, the diligent law journal students at Colorado, as well as my colleagues at NYU and Data & Society.

Risk assessment has dominated one of the first wide-ranging academic debates within the emerging field of data science. ProPublica investigative journalists claimed that the COMPAS algorithm is biased and released their findings as open data sets. The ProPublica data started a prolific and mathematically-specific conversation about risk assessment as well as a broader conversation on the social impact of algorithms. The ProPublica-COMPAS debate repeatedly considered three main themes: mathematical definitions of fairness, explainable interpretation of models, and the importance of population comparison groups.

While the Loomis *decision addressed permissible use for a risk assessment at sentencing, a deeper understanding of daily practice within the organization could extend debates about algorithms to questions about procurement, implementation, or training. The criminal justice organization that purchased the risk assessment is in the best position to justify how one individual's assessment matches the algorithm designed for its administrative needs. People subject to a risk assessment cannot conjecture how the algorithm ranked them without knowing why they were classified within a certain group and what criteria control the rankings. The controversy over risk assessment algorithms hints at whether procedural due process is the cost of automating a criminal justice system that is operating at administrative capacity.*

INTRODUCTION

How do you argue with an algorithm?

This question was at the center of a 2016 case that considered whether states could use the predictions of risk assessment algorithms in sentencing. *State of Wisconsin v. Loomis*[1] considered whether an individual could reasonably dispute predictions made by an algorithm[2] that is designed to serve the operations of the criminal justice system. While algorithms are essential to any contemporary organization,[3] it is not clear whether courts are equipped to explain judicial reasoning that is influenced by algorithmic predictions.

The sentencing of Mr. Eric L. Loomis of Wisconsin was based in part on a predictive algorithm that classified him as a high-risk defendant.[4] The algorithm at issue, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), is a proprietary algorithm sold by equivant, a private company that was doing business as Northpointe before 2017.[5] The Wisconsin circuit court sentenced Mr. Loomis to the maximum penalty on two counts after reviewing the predictions derived from the COMPAS risk-assessment algorithm, despite the defendant's claims that using a proprietary predictive risk assessment in sentencing violated his due process rights.[6] The Wisconsin Supreme Court dismissed the due process claims because (1) identical COMPAS reports were available to both the defendant and to the State,[7] and (2) the

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

2. I use "algorithm" to refer to computer procedures that take input information, process it using formalized logic, and produce a result. For a legal introduction to algorithms, see David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine* Learning, 51 U.C. DAVIS L. REV. 653 (2017).

3. Both private and public sector organizations increasingly rely on algorithms and data-driven management. Computer modernization has transformed public sector operations and predictive algorithms now influence many aspects of governance. For a discussion on government organizations, see Marijn Janssen & George Kuk, *The Challenges and Limits of Big Data Algorithms in Technocratic Governance*, 33 GOV'T INFO. Q. 371, (2016). For a more general discussion on the turn towards a data-driven economy, see Steve Lavalle et al., *Big Data, Analytics and the Path from Insights to Value*, 52 MIT SLOAN MGMT. REV. 21 (2011); *Data is Giving Rise to a New Economy*, ECONOMIST (May 16, 2017), https://www.economist.com/news/briefing/21721634-how-it-shaping-up-data-giving-rise-new-economy [http://perma.cc/H5J2-G879].

4. *Loomis*, 881 N.W.2d at 755.

5. Northpointe, Inc., CourtView Justice Solutions, Inc., and Constellation Justice Systems, Inc. consolidated into a single rebranded entity called equivant on January 9, 2017. They retained the product name COMPAS. *CourtView, Constellation, & Northpointe Re-Brand to equivant*, EQUIVANT, http://www.equivant.com/blog/we-have-rebranded-to-equivant [http://perma.cc/BS2L-ZSZH].

6. *Loomis*, 881 N.W.2d at 756.

7. *Id.* at 761–62 ("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. Loomis had an

defendant had the opportunity to correct any inaccurate responses to the questions used to calculate the risk assessment.[8] On appeal, the Wisconsin Supreme Court let the sentence stand, effectively affirming the use of predictive assessments in sentencing decisions.[9]

According to the *Loomis* court, the way to argue with a prediction from a COMPAS algorithm is to question the accuracy of the input data.[10] The court reasoned that by correcting any inaccurate information that went into the algorithm, the defense would be challenging the output of the algorithm.[11] The *Loomis* court narrowed the scope for challenging a COMPAS risk assessment to a single data quality point[12]: the accuracy of the defendant's responses used in the algorithm.[13]

> Was the court in *Loomis* right about how to argue with an algorithm?

An ongoing academic debate about the COMPAS algorithm suggests that the court's reasoning was flawed.[14] By focusing solely on data accuracy, the *Loomis* court ignored the computational procedures that processed the input data. The court dismissed an essential aspect of how algorithms function and overlooked the possibility that accurate data could produce an inaccurate prediction. While concerns about data quality are necessary, they are not sufficient to challenge, defend, nor improve the results of predictive algorithms. How algorithms calculate data is equally worthy of scrutiny as the quality of the data themselves. The arguments in *Loomis* revealed a need for the legal scholars to be better connected to the cutting-edge reasoning used by data science practitioners.[15]

---

opportunity to challenge his risk scores by arguing that other factors or information demonstrate their inaccuracy.").

8. *Id.* at 761 ("Although Loomis cannot review and challenge how the COMPAS algorithm calculates risk, he can at least review and challenge the resulting risk scores set forth in the report attached to the PSI.").

9. *Id.* at 772, *cert. denied*, 137 S. Ct. 2290 (2017).

10. *Id.* at 761–62.

11. *Id.* ("Loomis had the opportunity to verify that the questions and answers listed on the COMPAS report were accurate.").

12. Data quality has multiple dimensions beyond accuracy, including completeness, consistency, timeliness, representativeness, unambiguousness, meaning, precision, and reliability. *See* Yair Wand & Richard Y. Wang, *Anchoring Data Quality Dimensions in Ontological Foundations*, COMMS. ACM, Nov. 1996, at 86, 93–94.

13. *Loomis*, 881 N.W.2d at 763.

14. For a review of the initial academic impact of a May 2016 ProPublica investigative journalism article, see Julia Angwin & Jeff Larson, *Bias in Criminal Risk Scores is Mathematically Inevitable, Researchers Say*, PROPUBLICA, (Dec. 30, 2016), https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say [http://perma.cc/Y7VC-66TG].

15. In a concurring opinion to *Loomis*, Justice Shirley S. Abrahamson noted, "this court's lack of understanding of COMPAS was a significant problem in the instant case.

An academic debate on risk assessment began in May 2016 when investigative journalists at ProPublica published "Machine Bias."[16] The authors Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner accused COMPAS of systematically giving advantages to people identifying as white.[17] Northpointe, the company that owned COMPAS in 2016, disputed ProPublica's claims with their own analysis[18] and ProPublica replied.[19] Within 18 months, over 200 academic papers cited the ProPublica article.[20] The scholarly debate was only possible because ProPublica openly released its files on a popular site computer scientists use to share data and software code.[21] Scholars very quickly reproduced the ProPublica results, employed alternative methods to produce different results, and published findings. Data scientists, statisticians, criminal justice professionals, and journalists jumped into a public and mathematically-specific conversation about risk assessment. The breadth and intensity of the ProPublica-COMPAS debate underscore the many subjective considerations of producing algorithms.[22]

A legal question raised in *Loomis* was one of due process.[23] Specifically, is it a violation of due process when courts use algorithmically derived predictions to support a sentencing decision?[24] The ability to give a reason for a decision is essential to

---

At oral argument, the court repeatedly questioned both the State's and the defendant's counsel about how COMPAS works. Few answers were available." *Loomis*, 881 N.W.2d at 774 (Abrahamson, J., concurring).

16. Julia Angwin et al., *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [http://perma.cc/3M9F-LFDM].

17. *Id.*

18. WILLIAM DIETERICH ET AL., COMPAS RISK SCALES: DEMONSTRATING ACCURACY EQUITY AND PREDICTIVE PARITY (2016), http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf [http://perma.cc/L7VU-T4BT].

19. Jeff Larson & Julia Angwin, *ProPublica Responds to Company's Critique of Machine Bias Story*, PROPUBLICA (July 29, 2016), www.propublica.org/article/propublica-responds-to-companys-critique-of-machine-bias-story [http://perma.cc/FM9V-W5EU].

20. On January 8, 2019 we ran a Google Scholar search for articles that cited the URL or the title of the ProPublica article in 2016–17 and found 248 English-language results. There were another 330 results in 2018. Citations to "Machine Bias" by ProPublica, GOOGLE SCHOLAR, https://scholar.google.com/ (type "'machine bias" OR "www.propublica.org/article/machine-bias"' into Google Scholar).

21. *Data and Analysis for 'Machine Bias',* GITHUB, https://github.com/propublica/compas-analysis/ [http://perma.cc/6UEP-24YS] [hereinafter *Data and Analysis*]; *see* Jeff Larson et al., *How We Analyzed the COMPAS Recidivism Algorithm*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm [http://perma.cc/QD4F-3VBR].

22. For an in-depth discussion of the social construction of actuarial risk in risk assessment, see Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59 (2017).

23. State v. Loomis, 881 N.W.2d 749, 753 (Wis. 2016).

24. *Id.*

legal practice.[25] Legal decisions are documented so they can be interpreted. Documentation is an extension of how the administrative state is held accountable by the public for what they do and why they do it.[26] The implementation of the European Union's General Data Protection Regulation (GDPR) has sparked a legal interest in the explainability of algorithms.[27] Reasoning about predictive algorithms in the future is likely to be closely tied to developments in GDPR regulation.

Algorithms supporting public sector operations raise concerns about the visibility of administrative decisions and fair procedures, specifically what Danielle K. Citron called "technological due process."[28] Arguments about due process must have a better grasp on how predictive algorithms function and how they are implemented in criminal justice organizations.

This article offers legal scholars an analysis of the essential arguments made by data scientists in the ProPublica-COMPAS debate with the goal of resolving uncertainty about what information is useful to evaluate a COMPAS assessment. This discussion is intended to better inform future legal considerations of procedural due process when information is derived from predictive algorithms. This comprehensive treatment could potentially serve as a basis for future courtroom arguments on the integrity of algorithmically derived predictions.

Section I of this article outlines the *Loomis* case and the points made by both sides about challenging the COMPAS algorithm. This Section also explores the legal question of due process considered in the case.

---

25. Thanks to Julia Powles and Andrew Selbst for this point. *See* Frederick Schauer, *Giving Reasons*, 47 STAN. L. REV. 633, 633 (1995); Martin Shapiro, *The Giving Reasons Requirement*, 1992 U. CHI. LEGAL F. 179, 180 (1992).

26. Legal and public administration scholars have written extensively on the importance of transparent procedures. *See, e.g.*, *Loomis*, 881 N.W.2d at 774 (Abrahamson, J., concurring) ("First, I conclude that in considering COMPAS (or other risk assessment tools) in sentencing, a circuit court must set forth on the record a meaningful process of reasoning addressing the relevance, strengths, and weaknesses of the risk assessment tool."); CHRISTOPHER HOOD & DAVID HEALD, TRANSPARENCY: THE KEY TO BETTER GOVERNANCE? (2006); Rónán Kennedy, *Algorithms and the Rule of Law*, 17 LEGAL INFO. MGMT. 170, 170–72 (2017); Shapiro, *supra* note 25.

27. The 2017 GDPR regulates the online exchange of data in the European Union and calls for more accountability than contemporary laws in other jurisdictions. *See* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085 (2018); Andrew D. Selbst & Julia Powles, *Meaningful Information and the Right to Explanation*, 7 INT'L DATA PRIVACY L. 233, 234 (2017).

28. Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1254 (2008) ("The opacity of automated systems shields them from scrutiny. Citizens cannot see or debate these new rules. In turn, the transparency, accuracy, and political accountability of administrative rulemaking are lost.") Legal scholars warned about the possibility that technology could hide procedures. For a discussion on procedures hidden in predictive scoring algorithms, see FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015).

Section II provides background on risk assessments, especially scoring algorithms like those used to predict the risk of criminal defendants.[29]

Section III presents ProPublica's claims about COMPAS, identifies the conditions that laid the foundation for the academic debate, and delves into the scholarly concerns prompted by the ProPublica series. This Section translates the main findings from scholars who analyzed the ProPublica data set. The claims repeatedly employed by data science scholars were mathematical definitions of fairness, data model simplicity, and population comparisons.

Section IV proposes alternative claims about risk assessments using *Loomis* as an example. The proposals seek to establish professional norms for practice, reveal algorithm provenance, or share sample data. A criminal justice organization could proactively follow these proposals, especially during procurement of third-party products. Defendants could request this information to challenge a risk assessment. Data science reasoning[30] is put forward as one way to conceptualize the logic behind predictive analytics, such as risk assessment scores that provide probabilistic insights.

The use of risk assessment algorithms in *Loomis* points to a broader social concern about how to appropriately use algorithmically-derived information in the public sector. As a computer scientist and scholar of digital government, I view predictive risk assessments as a special case of informatics[31] and innovation in the public sector. This article positions risk assessment algorithms as a commercial software product sold to support the operations of the courts. This position opens up new avenues for understanding prediction in judicial decision-making.

29. Criminal justice has a history of risk assessments even before the use of computational algorithms. *See, e.g.*, Charles W. Dean & Thomas J. Duggan, *Problems in Parole Prediction: A Historical Analysis*, 15 SOC. PROBS. 450, 457 (1968); Michael Hakeem, *The Validity of the Burgess Method of Parole Prediction*, 53 AM. J. SOC. 376, 379 (1948).

30. "Data Science Reasoning" was the title of my 2016–17 Fellowship at the Data & Society Research Institute where I considered how to improve data science education and data literacy in the public sector. *See Data Science Reasoning*, DATA & SOC'Y, https://datasociety.net/initiatives/additional-projects/datareasoning/ [http://perma.cc/L85T-URUG].

31. Public sector informatics considers the institutional and social contexts of the texts created by government. *See generally* Kevin P. Jones, *Informatics*, 261 NATURE 370 (1976) (defining informatics as the study of structure within large collections of text); Rob Kling, *What Is Social Informatics and Why Does It Matter?*, 5 D-LIB MAG., Jan. 1999, http://www.dlib.org/dlib/january99/kling/01kling.html [http://perma.cc/M897-5JKA].

I. THE CASE: *WISCONSIN V. LOOMIS*

Wisconsin v. Loomis addressed the use of risk assessments generated by an algorithm in sentencing.[32] The case drew on state and federal case law about pre-sentencing disclosures.[33] The Wisconsin Supreme Court dismissed claims that the defendant was denied due process.[34] This Section reviews the *Loomis* arguments on pre-sentence information, sentencing, and due process.

A. Pre-Sentence Information

The concerns in *Loomis* revolved around the contents of a Pre-Sentence Investigation (PSI) report. The Wisconsin circuit court ordered a PSI report on the defendant in *Loomis,* which included a risk assessment generated by the COMPAS algorithm.[35] PSI reports support the internal operational efficiency of the court.[36] The Wisconsin Supreme Court cited the *State v. Skaff* decision, which determined that a defendant was in the best position to refute, explain, or supplement incorrect or incomplete information in the PSI.[37]

The PSI in *Loomis* included a COMPAS risk assessment score, a graph chart showing the placement of the score, and twenty-one related questions and answers.[38] COMPAS scores range from 1 to 10, with 10 representing the strongest prediction of risk.[39] The predictive risk scores are then grouped into classifications: 1–4 Low Risk, 5–7 Medium Risk, and 8–10 High Risk.[40]

The COMPAS risk assessment is derived, in part, from responses to a series of questions.[41] The sources that go into the COMPAS algorithms differ by jurisdiction and predictive

32. State v. Loomis, 881 N.W.2d 749, 772 (Wis. 2016).
33. *Id.* at 760–64.
34. *Id.* at 753.
35. *Id.* at 754.
36. The Wisconsin court system regularly commissions studies on the efficiency of the courts. A 2012 report considered how to improve PSI reports. *See* SUZANNE TALLARICO ET AL., EFFECTIVE JUSTICE STRATEGIES IN WISCONSIN: A REPORT OF FINDINGS AND RECOMMENDATIONS 156 (2012), https://www.wicourts.gov/courts/programs/docs/ejsreport.pdf [https://perma.cc/9C69-DSTS].
37. *Loomis*, 881 N.W.2d at 760 (citing *State v. Skaff,* 152 Wis. 2d 48, 58 (Ct. App. 1989)).
38. *Id.* at 761.
39. *Id.* at 754.
40. Glimpses in published reports and legal cases are the only way to guess how the proprietary COMPAS risk assessment algorithms function. *Practitioner's Guide to COMPAS Core*, NORTHPOINTE 1, 8 (Mar. 19, 2015), http://www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-_031915.pdf [http://perma.cc/4GA6-7QPZ].
41. In affirming the circuit court's decision, the Supreme Court of Wisconsin cited three questions and answers to justify the high-risk assessment score: "[1] How many times has this person been returned to custody while on parole? 5+ [2] How many times has this person had a new charge/arrest while on probation? 4 [3] How many times has this person been arrested before as an adult or juvenile (criminal arrest only)? 12." *Loomis*, 881 N.W.2d at 761.

assessment product. The literature on risk assessment scores generally states that they are based on administrative data, public records, self-reporting, and interviews.[42] The questions might cover substance abuse, employment, education, criminal history, residential stability, family criminality, and social environment, according to reports.[43]

B. Sentencing

In *Loomis*, the defendant denied involvement in the crime but waived his right to trial by agreeing to a plea deal.[44] The plea deal left the actual sentence to the discretion of the Wisconsin circuit court judge.[45] The judge accepted the guilty plea from the defendant and ordered a risk assessment as part of the PSI.[46] The COMPAS risk assessment predicted that the defendant had high pre-trial risk, high risk of recidivism, and high risk of violent recidivism.[47] Instead of one year in county jail with probation, which the prosecution and defense had agreed upon, the circuit court sentenced the defendant to "seven years with four years initial confinement" for operating a motor vehicle without the owner's consent.[48] For attempting to flee an officer, the circuit court sentenced him to four years with two years of initial confinement to be served consecutively in state prison.[49] Both charges were repeat offenses.[50]

The defendant filed a motion requesting a new sentencing hearing arguing that "the circuit court erroneously exercised its discretion" by referring to a high-risk assessment score when imposing the maximum sentence.[51] At the sentencing hearing, the circuit court referenced the high COMPAS risk classification given to the defendant, specifically stating that his PSI shows "a high risk

42. *See id.* at 754, 761.

43. TIM BRENNAN ET AL., ENHANCING PRISON CLASSIFICATION SYSTEMS: THE EMERGING ROLE OF MANAGEMENT INFORMATION SYSTEMS 48 (Northpointe Inst. for Pub. Mgmt., Inc. ed., 2004) [hereinafter BRENNAN ET AL., ENHANCING PRISON CLASSIFICATION SYSTEMS], https://permanent.access.gpo.gov/lps56481/019687.pdf [http://perma.cc/4NEE-T8S5]; DIETERICH ET AL., *supra* note 18, at 5–6; Tim Brennan, et al., *Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System*, 36 CRIM. JUST. & BEHAVIOR 21, 25 (2009) [hereinafter Brennan et al., *Evaluating the Predictive Validity*].

44. *Loomis*, 881 N.W.2d. at 754.

45. *Id.*

46. *Id.*

47. *Id.* at 755.

48. *Id.* at 756 n.18.

49. *Id.*

50. *Id.* at 754.

51. *Id.* at 756.

to the community."[52] The defendant's motion for a new sentencing hearing was denied.[53]

C. *Due Process Claims*

The defendant in *Loomis* appealed, claiming that the sentence denied him procedural due process.[54] The Wisconsin Supreme Court heard the case.[55] In the petition, the defendant stated two concerns about due process: access to pre-sentencing disclosures and the right to fair sentencing with accurate data.[56]

First, the defendant argued that the sentence violated his constitutional due process rights to pre-sentencing information disclosures.[57] In *Gardner v. Florida*, the United States Supreme Court considered how information is shared before sentencing.[58] The defendant in *Gardner* was sentenced to death, partly due to confidential PSI information that was not disclosed to defense counsel.[59] The Supreme Court in *Gardner* determined that a denial to release the information used in sentencing was a denial of due process.[60]

The defendant in *Loomis* claimed that the confidentiality of PSI information in *Gardner* was similar to the proprietary aspects of the COMPAS algorithm.[61] The use of a proprietary COMPAS algorithm, therefore, was a failure of disclosure. The information requested by the defendant in *Loomis* included access to the software code and to the algorithmic weighting. Both requests were denied because the COMPAS algorithm is proprietary and is protected by trade secret laws. The Wisconsin Supreme Court determined that the State did not rely on information withheld from the defendant because both parties had the COMPAS risk assessment report.[62]

---

52. *Id.* at 755 ("You're identified, through the COMPAS assessment, as an individual who is at high risk to the community. In terms of weighing the various factors, I'm ruling out probation because of the seriousness of the crime and because your history, your history on supervision, and the risk assessment tools that have been utilized, suggest that you're extremely high risk to re-offend.").

53. *Id.* at 756–57.

54. *Id.* at 757.

55. *Id.* at 757 ("The court of appeals certified the specific question of whether the use of a COMPAS risk assessment at sentencing 'violates a defendant's right to due process, either because the proprietary nature of COMPAS prevents defendants from challenging the COMPAS assessment's scientific validity, or because COMPAS assessments take gender into account.'").

56. *See id.* at 760–61 ("[I]t violates a defendant's right to be sentenced based upon accurate information, in part because the proprietary nature of COMPAS prevents him from assessing its accuracy.").

57. *Id.*

58. 430 U.S. 349, 351 (1977).

59. *Id.*

60. *Id.*

61. *Loomis*, 881 N.W.2d at 761.

62. *Id.*

Second, the defendant in *Loomis* drew on Wisconsin case law regarding the accuracy of PSI reports.[63] *State v. Skaff* held that a criminal defendant has the right to challenge pre-sentencing information.[64] Following the decision in *Skaff*, a criminal defendant has the right to check for inaccuracies as well as "refute, explain, or supplement" information that might affect the sentence.[65] The defendant in *Loomis* argued that it was impossible to challenge a risk assessment without sufficient information about how COMPAS functions, such as how risk is determined and how factors are weighed to calculate the assessment.[66] Because the defendant could correct responses to questions, the court determined that he had the ability to determine the accuracy of his risk assessment.[67]

The focus on data quality by the *Loomis* court overemphasized a sense of determinism from the selected responses.[68] Current scholarship in data studies has distanced itself from the idea that data are "raw" and instead considers data as highly contextualized observations.[69]

The Wisconsin Supreme Court did not find the precedents about pre-sentencing disclosure compelling and let the sentencing decision stand.[70] The Court affirmed the use of risk assessments by narrowly specifying permissible use for COMPAS at sentencing to avoid violation of due process.[71] The defendant submitted a petition to the United States Supreme Court after his appeal was denied.

---

63. *Id.*

64. 152 Wis. 2d 48, 53 (Ct. App. 1989).

65. *Id.* at 57. The *Loomis* opinion stated multiple times that the defendant had the ability "to refute, explain, or supplement the [pre-sentencing report]." *Loomis*, 881 N.W.2d at 760.

66. *Loomis*, 881 N.W.2d at 761.

67. *Id.*

68. Computer scientists debate over whether data structures or operations are more influential in determining the outcome of algorithms. Moshe Vardi compares the problem to physicists arguing about whether light is a particle or a wave. Moshe Y. Vardi, *What Is an Algorithm?,* COMMS. ACM, Mar. 2012, at 5, 5.

69. *See generally* "RAW DATA" IS AN OXYMORON (Lisa Gitelman ed., 2013) (arguing that data are anything but "raw" and that data should be viewed as a cultural resource that needs to be generated, protected, and interpreted).

70. *Loomis*, 881 N.W.2d at 764.

71. *Id.* at 757, 763–64 ("Although we ultimately conclude that a COMPAS risk assessment can be used at sentencing, we do so by circumscribing its use. Importantly, we address how it can be used and what limitations and cautions a circuit court must observe in order to avoid potential due process violations . . . . Specifically, any PSI containing a COMPAS risk assessment must inform the sentencing court about the following cautions regarding a COMPAS risk assessment's accuracy: (1) the proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are to be determined; (2) risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed; (3) some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having a higher risk of recidivism; and (4) risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations.").

The question presented for review by the Court was whether the proprietary nature of the COMPAS violated a defendant's constitutional right to due process because a defendant cannot challenge the algorithm's accuracy or scientific validity.[72] The United States Supreme Court declined to hear the case, allowing the Wisconsin Supreme Court's ruling to stand.[73]

In a concurring opinion in the Wisconsin *Loomis* case, Justice Abrahamson called for ways that courts could keep up to date with developments in evidence-based decision-making, noting that "[t]he court needed all the help it could get."[74] This paper attempts to provide some of that help with arguments about algorithms from data science scholarship.

II. THE ALGORITHM: RISK ASSESSMENT

The risk assessments in *Loomis* were derived from an algorithm. The ability to argue with an algorithm requires confronting the base assumption that an algorithmically-derived assessment is objectively true, distant, and fixed. This article challenges the premise that risk assessment scores reflect a single objective reality. Risk assessments are actively constructed and are subject to a variety of subjective influences.[75] This Section provides background on algorithms, information systems management, and data quality.

A. Why Assess Risk with an Algorithm?

An algorithm is a method for solving and refining the performance of finite procedures usually implemented on a computer.[76] To computer scientists, algorithms are modular programs that can sort, search, count, and classify.[77] Courts, along with other government agencies, are modernizing and algorithms are part of the process of joining the data economy and expanding operational capacity to large populations.[78]

Risk assessments attempt to maintain public safety by identifying those who repeatedly commit crimes and are likely to be a threat to society if not incarcerated.[79] Assessing the risk of

72. Loomis v. Wisconsin, 137 S. Ct. 2290 (2017) ("Petition for writ of certiorari to the Supreme Court of Wisconsin denied.").

73. *Id.*

74. *Loomis,* 881 N.W.2d at 774 (Abrahamson, J., concurring).

75. Eaglin, *supra* note 22.

76. ROBERT SEDGEWICK & KEVIN DANIEL WAYNE, ALGORITHMS 4 (4th ed. 2011).

77. *Id.*

78. *See generally* Amanda Clarke & Helen Margetts, *Governments and Citizens Getting to Know Each Other? Open, Closed, and Big Data in Public Management Reform*, 6 POL'Y & INTERNET 393 (2014) (discussing the use of big data analysis by governments).

79. Risk assessments are commonly offered along with needs assessments. The needs of defendants entering the system are assessed to identify low-risk offenders who, if certain criteria are met, can be supervised in outside rehabilitation. COMPAS provides

criminal defendants requires balancing the costs of managing prisons with the benefits of maintaining public safety. Although risk assessment was done before computing,[80] risk assessment algorithms are popular because of the increased availability of data sources and technology that can quickly calculate thousands of attributes into a single predictive score.[81]

Risk assessment algorithms attempt to find offenders who might commit more crimes if not placed in confinement. Algorithms, like COMPAS, operationalize risk of recidivism[82] as predicting those who are likely to be arrested again, which rarely considers geographic structural conditions.[83] Predicting a misdemeanor arrest or felony arrest is analytically different from predicting misdemeanor conviction or felony conviction. Those who object to predictive assessments suggest that the best way to reduce the incarcerated population is to contain policing behavior that leads to over monitoring neighborhoods through frequent arrests for minor incidents.[84] However, changing police behavior is often not the goal of risk assessment policies.[85] Given uneven police arrest behavior, arrest as an outcome has been challenged as the wrong measurement.[86]

---

both of needs-risk assessment products available. *See* CHRIS BAIRD ET AL., A COMPARISON OF RISK ASSESSMENT INSTRUMENTS IN JUVENILE JUSTICE, at i–ii (2013), https://www.ncjrs.gov/pdffiles1/ojjdp/grants/244477.pdf [http://perma.cc/5D3H-9S5L]; Sheldon X. Zhang et al., *An Analysis of Prisoner Reentry and Parole Risk Using COMPAS and Traditional Criminal History Measures*, 60 CRIME & DELINQUENCY 167, 168 (2014). *But see* Tracy L. Fass et al., *The LSI-R and the COMPAS: Validation Data on Two Risk-Needs Tools*, 35 CRIM. JUST. & BEHAV. 1095, 1106 (2008).

80. *See* Ernest W. Burgess, *Protecting the Public by Parole and by Parole Prediction*, 27 J. CRIM. L. & CRIMINOLOGY 491, 498–501 (1936).

81. Risk scoring is part of a larger trend of predictive evaluation of populations that Citron and Pasquale refer to as the "Scored Society." For example, consumer credit scores in the United States estimate the relative credit worthiness of potential consumers. Marketing scores extend the same model by determining the likelihood that someone will make a purchase. *See* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1413 (2014) (discussing the increasing use of big data to rank individuals through predictive algorithms).

82. *See* DIETERICH ET AL., *supra* note 18, at 15.

83. Risk assessments rarely consider historical, societal, and structural problems that reproduce crime and arrest patterns. *See* Chelsea Barabas et al., *Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment*, 81 PROC. MACHINE LEARNING RES. 1, 6 (2017) ("We posit that machine learning should not be used for prediction, but rather to surface covariates that are fed into a causal model for understanding the social, structural and psychological drivers of crime. We propose an alternative application of machine learning and causal inference away from predicting risk scores to risk mitigation.").

84. *See generally* BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE (2007) (questioning the use of predictive assessments).

85. *See* Mirko Bagaric et al., *Bringing Sentencing into the 21st Century: Closing the Gap Between Practice and Knowledge by Introducing Expertise into Sentencing Law*, 45 HOFSTRA L. REV. 785, 826–29 (2017); Wayne A. Logan & Andrew Guthrie Ferguson, *Policing Criminal Justice Data*, 101 MINN. L. REV. 541, 543–44 (2016).

86. Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 133–34 (2017).

The tension between identifying risk and minimizing harm[87] was the focus of the academic debate over COMPAS risk assessment. Is it possible to minimize harm to those who might be misidentified as a high risk? Is it possible to minimize harm to the public if someone is misidentified as a low risk? These are age-old questions of public policy. Discussing similar predictions in 1936, Burgess writes: "Parole, and in fact our whole system of criminal justice, must constantly be prepared to face trial in the court of public opinion."[88]

B. Why Do Information Systems Matter?

Risk assessment algorithms promise efficiency and fairness to organizations that are struggling to manage increasing prison and jail populations. From police stops and arrest to incarceration, the numbers keep growing. In New York City, police conducted 4.4 million stops from 2004 through 2012.[89] Arrests nationwide numbered over 10 million in 2016, according to the FBI's Uniform Crime Reporting.[90] The incarcerated population in the United States over the last few decades grew beyond the capacity of the organizations charged with maintaining public safety. The incarcerated population in 1983 was 438,830, while in 2014 it was over 1.5 million.[91]

Algorithms and data-driven technology help to ease the administrative burdens of these growing systems.[92] In some cases, courts pay external vendors to produce information the court needs, such as classifying people with predictive assessments of

87. *See generally* Rachel Courtland, *Bias Detectives: The Researchers Striving to Make Algorithms Fair*, 558 NATURE 357, 358–59 (2018); Geoff Pleiss et al., *On Fairness and Calibration*, 31 CONF. ON NEURAL INFO. PROCESSING SYS. 2904 (2017), https://papers.nips.cc/paper/7151-on-fairness-and-calibration.pdf [http://perma.cc/K7EM-5B8B] (summarizing machine learning research on fairness and bias).

88. Burgess, *supra* note 80, at 491.

89. The stops were not equally distributed throughout the population. 83% of the stops involved a person who was identified as black or Hispanic. Only 6% of these stops resulted in an arrest. *See* N.Y. Times Editorial Bd., *Racial Discrimination in Stop-and-Frisk*, N.Y. TIMES (Aug. 12, 2013), https://nyti.ms/15x3ngU [https://perma.cc/K5AX-MXK9].

90. *Table 18: Estimated Number of Arrests United States, United States 2016*, FBI UNIFORM CRIME REPORTING, https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/tables/table-18 [http://perma.cc/K5RZ-N2FS].

91. E. ANN CARSON, BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, PRISONERS IN 2014, at 1 (2015), https://www.bjs.gov/content/pub/pdf/p14.pdf [http://perma.cc/M4Y5-UGNK]; BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, PRISONERS IN 1983, at 1 (1984), https://www.bjs.gov/content/pub/pdf/p83.pdf [http://perma.cc/827X-8L8D].

92. Teppo Felin et al., *The Law and Big Data*, CORNELL J.L. & PUB. POL'Y 357, 359 (2017). Courts began to modernize using technology along with other parts of government. An important point in the United States was the Federal E-Government Act of 2002. *See* E-Government Act of 2002, Pub. L. No. 107-347, 116 Stat. 2899 (2002) (codified as amended across sections of the U.S. Code).

recidivism.[93] COMPAS is a brand of risk-need assessment tools designed to provide decisional support through classification.[94] COMPAS is a product sold to support the operations of criminal justice organizations and serves as an extension of existing judicial information systems.[95]

The Wisconsin Department of Corrections uses COMPAS risk-needs algorithms to make "placement decisions, [manage] offenders, and [plan] treatment."[96] However, it can be difficult to assess the validity of information generated through proprietary algorithms because vendors often claim that their algorithms are trade secrets that cannot be shared.[97] Trade secret claims are complicated when proprietary algorithms are sold to public sector organizations that are expected to meet standards of transparency, accountability, and rule of law.[98]

Organizations, in general, have three choices when modernizing: build an internal system, purchase a retail system, or create a system in alliance with others who have the capacity.[99] The State of Wisconsin had developed its own assessment system in the late 1970s.[100] Why did they choose to abandon that project and buy a commercial vendor? What were the differences in cost? If the defendant in *Loomis* had been arrested multiple times, is it possible that there were other risk assessments done using a different algorithm?

---

93. BRENNAN ET AL., ENHANCING PRISON CLASSIFICATION SYSTEMS, *supra* note 43, at 21; *COMPAS Classification*, EQUIVANT, http://www.equivant.com/solutions/inmate-classification [http://perma.cc/AK27-PJ7L].

94. Ed Yong, *A Popular Algorithm Is No Better at Predicting Crimes than Random People*, ATLANTIC (Jan. 17, 2018) https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646 [http://perma.cc/RPF9-8R9S]; *Practitioner's Guide to COMPAS Core*, *supra* note 40, at 2. Dressel and Farid conducted a study that compared COMPAS assessments with non-expert human assessments. *See* Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, SCI. ADVANCES, Jan. 17, 2018, at 1, 4

95. E-justice systems in the judicial branch of government developed alongside e-government systems in executive branch. In both cases, the systems were designed to meet public sector statutory goals more efficiently. *See generally* BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, REPORT OF THE NATIONAL TASK FORCE ON COURT AUTOMATION AND INTEGRATION (1999), http://www.ncjrs.gov/pdffiles1/177601.pdf [http://perma.cc/9DB4-JLVR]; João Rosa et al., *Risk Factors in E-Justice Information Systems*, 30 GOV'T INFO. Q. 241 (2013).

96. State v. Loomis, 881 N.W.2d 749, 754 (Wis. 2016).

97. Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1421–22 (2018).

98. *See generally* Janssen & Kuk, *supra* note 3, at 373; Kennedy, *supra* note 26, at 170; Frank Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEO. WASH. L. REV 1, 6 (2019).

99. Abhishek Borah & Gerard J. Tellis, *Make, Buy, or Ally? Choice of and Payoff from Announcements of Alternate Strategies for Innovations*, 33 MARKETING SCI. 114, 114 (2014).

100. MIKE EISENBERG ET AL., THE COUNCIL OF STATE GOV'TS JUSTICE CTR., VALIDATION OF THE WISCONSIN DEPARTMENT OF CORRECTIONS RISK ASSESSMENT INSTRUMENT 1 (2009), https://csgjusticecenter.org/wp-content/uploads/2012/12/WIRiskValidationFinalJuly2009.pdf [http://perma.cc/TD7X-LFYV].

Risk assessment is designed and marketed to support court employees,[101] yet defendants need a meaningful way to express their objection to predictive classifications. Unlike data-driven tools sold to private organizations, the public sector must meet a higher standard for explainability. Organizations that rely on risk assessments should be able to confirm that vendors[102] are providing appropriate information that meets the organization's statutory obligations[103] and expectations of public-sector transparency.

C. Why Is Data Quality Alone Insufficient?

Under *Loomis*, errors in the underlying data are a threshold requirement for disputing a risk assessment score.[104] Algorithms require input data and the quality of that data is indeed an essential aspect of evaluating the results of an algorithm.[105] Data quality alone, however, is not sufficient to dispute a risk assessment because it does not account the essential procedures for processing data. A risk assessment algorithm processes data by combining sources, weighting variables, establishing ranks, and setting category boundaries.

Not all data that goes into a predictive score are equal. Few people could review their bank transactions to understand their credit score without knowing that a utility payment might matter more than a grocery bill. The same is true for risk assessments. Prior convictions might be considered more important than marital status in a risk assessment. Algorithms balance the relative importance of each data element to create weighted measures. The design requirements of the predictive algorithm would specify what the weighted values are for each data element. Without any indication of how the responses were evaluated, it would not be possible to challenge the overall predictive score by reviewing question responses.

---

101. The COMPAS product is marketed to specific roles in judicial organizations on a series of pages directed towards, for instance, a court administrator or public defender. *See Court Administrator*, EQUIVANT, http://www.equivant.com/roles/Court-Administrator [http://perma.cc/Z3T6-Q7PF]; *Public Defenders*, EQUIVANT, http://www.equivant.com/roles/public-defender [http://perma.cc/9WR3-T99V].

102. For a discussion of vendors and public values, see Bram Klievink et al., *The Collaborative Realization of Public Values and Business Goals: Governance and Infrastructure of Public–Private Information Platforms*, 33 GOV'T INFO. Q. 67 (2016) and Foster Provost & Tom Fawcett, *Data Science and Its Relationship to Big Data and Data-Driven Decision Making*, 1 BIG DATA 51, 51 (2013).

103. For a discussion of translating statutory obligations into algorithms and software, see Kennedy, *supra* note 26, at 170–72.

104. State v. Loomis, 881 N.W.2d 749, 760–64 (Wis. 2016).

105. Kenneth C. Laudon, *Data Quality and Due Process in Large Interorganizational Record Systems*, COMMS. ACM, Jan. 1986, at 4, 4; Yang W. Lee & Diane M. Strong, *Knowing-Why About Data Processes and Data Quality*, 20 J. MGMT. INFO. SYS. 13, 15 (2003); Wand & Wang, *supra* note 12.

Another fundamental flaw in the court's logic is its failure to understand that risk scores are a probabilistic ranking mechanism. Risk assessments need to be presented as conditional classifications.[106] Data science is a science of probability.[107] Predictions are based on extrapolations of trends assuming that some factors remain stable across time. Data analytics is often seen as objective because there is a distance between those who collect the information and those who use it. However, studies of databases and techniques in practice reveal that data are interpretive objects that carry the meaning that aligns with their production.[108] Classifications can go wrong and turn into a cycle of harms that equate to a blacklisting effect that restricts individual liberty.[109]

Of particular concern in algorithms are the differences across population groups, also known as base rates, which can create uneven impacts between groups.[110] In *Loomis*, the risk assessment algorithm put the defendant in the high-risk category.[111] ProPublica claims that the COMPAS algorithm may be more likely to classify population members incorrectly due to base-rate differences.[112]

While risk assessment might be consistently produced, the category thresholds can be implemented differently. A score of 4 might be high in Boise but low in Portland. Each jurisdiction might interpret the score threshold differently. In addition, differences in population base-rates might influence differences in risk assessment thresholds across jurisdictions.

In order to argue with a risk assessment algorithm, it is necessary to understand something about the ranking mechanism, the weighting, or the community in which the individual is placed. People subject to a risk assessment cannot second guess how the algorithm ranked them without knowing why they were classified

106. Muhammad Bilal Zafar et al., *Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification Without Disparate Mistreatment*, 26 PROC. INT'L WORLD WIDE WEB CONF. 1171 (2017), *cf.* CHRISTOPHER SLOBOGIN, PROVING THE UNPROVABLE: THE ROLE OF LAW, SCIENCE, AND SPECULATION IN ADJUDICATING CULPABILITY AND DANGEROUSNESS (2007).

107. *See* Provost & Fawcett, *supra* note 102, at 56.

108. Lev Manovich, *Database as Symbolic Form*, 5 CONVERGENCE 80, 84 (1999).

109. Margaret Hu, *Big Data Blacklisting*, 67 FLA. L. REV. 1735, 1738–40 (2015).

110. DIETERICH ET AL., *supra* note 18, at 1 ("ProPublica focused on classification statistics that did not take into account the different base rates of recidivism for blacks and whites."); Sam Corbett-Davies et al., *Algorithmic Decision Making and The Cost of Fairness*, 23 ACM SIGKDD INT'L CONF. ON KNOWLEDGE DISCOVERY & DATA MINING 797, 797 (2017) ("These algorithms do not explicitly use race as an input. Nevertheless, an analysis of defendants in Broward County, Florida revealed that black defendants are substantially more likely to be classified as high risk. Further, among defendants who ultimately did not reoffend, blacks were more than twice as likely as whites to be labeled as risky."); Tracy L. Fass et al., *The LSI-R and the COMPAS: Validation Data on Two Risk-Needs Tools*, 35 CRIM. JUST. & BEHAV. 1095, 1106 (2008) (suggesting "that there is predictive inaccuracy driven by racial/ethnic status").

111. State v. Loomis, 881 N.W.2d 749, 755 (Wis. 2016).

112. *See* Angwin et al., *supra* note 16.

within a certain group, what criteria dominants the rankings, and which groups they are being compared to.

III. THE DEBATE: PROPUBLICA AND COMPAS

The COMPAS risk assessment algorithm became prominent during the summer of 2016—when several concurrent events put a spotlight on algorithmically-derived risk assessment scores. In May 2016, ProPublica and Northpointe began a written public dispute.[113] In June 2016, the 114th U.S. Congress considered legislation to require risk assessment scores in federal prisons.[114] And in July 2016, the Wisconsin Supreme Court made a decision in *Loomis* that set a standard for using COMPAS scores in sentencing.[115] These events initiated a flurry of press and academic attention.

The publications that considered the ProPublica-COMPAS issue formed one of the first scholarly conversations about data science.[116] At least 578 scholarly articles cited either the ProPublica "Machine Bias" article between May 2016 and December 2017. For this analysis, I reviewed publications in computer science, data science, and statistics that used the ProPublica data set.[117]

This Section presents a taxonomy of the main concerns in the ProPublica-COMPAS debate. The recurring topics in the debate, discussed below, were definitions of fairness, algorithm explanation, and population base rates.

A. ProPublica Claims Bias

A few weeks before the *Loomis* decision, investigative journalists at ProPublica published a controversial article claiming that COMPAS risk assessment was biased.[118] Although COMPAS and other risk assessments scores had been accused of gender

113. ProPublica published their article in May 2016 and Northpointe replied with a report disputing their claims in July 2016. *See* DIETERICH ET AL., *supra* note 18; Angwin et al., *supra* note 16; Larson & Angwin, *supra* note 19.

114. CORRECTIONS Act, S. 467, 114th Cong. (2015); Sentencing Reform and Corrections Act, S. 2123, 114th Cong. (2015); Recidivism Risk Reduction Act, H.R. 759, 114th Cong. (2015); Sensenbrenner-Scott SAFE (Safe, Accountable, Fair, Effective) Justice Reinvestment Act, H.R. 2944, 114th Cong. (2015).

115. *Loomis*, 881 N.W.2d at 749.

116. DIETERICH ET AL., *supra* note 18; Angwin et al., *supra* note 16; Larson & Angwin, *supra* note 19;

117. *Data and Analysis*, *supra* note 21.

118. ProPublica published the "Machine Bias" Article on May 23, 2016. Angwin et al., *supra* note 16. The *Loomis* decision was released on July 13, 2016. *Loomis*, 881 N.W.2d 749.

bias,[119] ProPublica presented evidence that COMPAS was racially biased.[120]

Using public records laws, ProPublica requested the COMPAS recidivism risk assessment scores from the Sheriff's Office in Broward County, Florida. Broward County is subject to Florida's open record laws.[121] ProPublica analyzed whether defendants who had a predictive risk actually entered the criminal justice system again. ProPublica claimed that the pattern of incorrect COMPAS predictions, false positives,[122] uniformly landed on one racial group more than another.[123] Northpointe denied the accusation of racial bias and denounced the statistical choices ProPublica made.[124]

Because ProPublica made its COMPAS data freely available, a vigorous debate followed which involved academics, criminal justice professionals, journalists, statisticians, political scientists, and machine learning experts who each employed different types of

119. Shaina Massie, *Orange is the New Equal Protection Violation: How Evidence-Based Sentencing Harms Male Offenders*, 24 WM. & MARY BILL RTS. J. 521 (2015) (illustrating how some states give different threshold cutoffs or tailor actuarial instruments to reflect differences in people labeled as male or female); *see* John Lightbourne, *Damned Lies & Criminal Sentencing: Using Evidence-Based Tools*, 15 DUKE L. & TECH. REV. 327 (2017).

120. ProPublica profiled two shoplifting arrests as an illustration. A teenage African-American girl who had never been arrested before was rated as a medium risk by COMPAS after being charged with burglary for attempting to steal a bike. A 54-year-old man of European heritage had been arrested twice, had a criminal record, and had drugs in his car, but he was rated as low risk by COMPAS after being arrested for shoplifting. These individual examples represented the statistical problem that inaccurate predictions, or false positives, were not uniformly applied. Angwin et al., *supra* note 16.

121. *Id.*

122. For an excellent visual depiction of the differences in false positives between the groups, see Sam Corbett-Davies et al., *A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It's Actually Not that Clear.*, WASH. POST (Oct. 17, 2016), https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/ [https://perma.cc/9DC3-3CFH] [hereinafter Corbett-Davies et al., *A Computer Program Used for Bail*] ("ProPublica points out that among defendants who ultimately did not reoffend, blacks were more than twice as likely as whites to be classified as medium or high risk (42 percent vs. 22 percent). Even though these defendants did not go on to commit a crime, they are nonetheless subjected to harsher treatment by the courts. ProPublica argues that a fair algorithm cannot make these serious errors more frequently for one race group than for another . . . . Black defendants who don't reoffend are predicted to be riskier than white defendants who don't reoffend; this is ProPublica's criticism of the algorithm.").

123. Angwin et al., *supra* note 16; *see* Jon Kleinberg et al., *Inherent Trade-Offs in the Fair Determination of Risk Scores*, 2017 PROC. OF INNOVATIONS THEORETICAL COMPUTER SCI. (2017), https://arxiv.org/pdf/1609.05807.pdf [http://perma.cc/E3NX-QJWX] (noting the ProPublica point as "[o]ne of their main contentions was that the tool's errors were asymmetric: African-American defendants were more likely to be incorrectly labeled as higher-risk than they actually were, while white defendants were more likely to be incorrectly labeled as lower-risk than they actually were").

124. Responding to the ProPublica article, Northpointe stated: "Based on our examination of the work of Angwin et al. and on results of our analysis of their data, we strongly reject the conclusion that the COMPAS risk scales are racially biased against blacks. ProPublica focused on classification statistics that did not take into account the different base rates of recidivism for blacks and whites." DIETERICH ET AL., *supra* note 18, at 1.

reasoning to argue for or against the ProPublica findings. The ProPublica controversy over risk assessment scores reveals how the same evidence[125] can support a wide range of differing arguments.

The swiftness and completeness of these publications were based on an environment that privileged open access to information. It started with ProPublica publishing a methods paper along with the article and subsequently releasing their data so others could replicate their results.[126] Scholars who attempted to replicate the ProPublica findings made their designs and evaluation models widely available for peer review and public scrutiny. How did they choose which observations to exclude? How did they handle observations where race was not a binary black/white? What statistical tests were appropriate for their stated intentions? How did they interpret fairness and with what mathematical model? These are some of the questions that were addressed in the scholarly debate yet were absent in the *Loomis* case.

The central aspect of the debate was evaluating the difference between ProPublica and Northpointe's definition of bias. Supporters argue that scores introduce additional objective analysis that is better than current human-biased systems.[127] Detractors argue that scores unfairly limit individual evaluation and are unnecessarily opaque, making it hard for defendants to argue against the results.[128] Critics also say that scores are an inappropriate attempt to predict criminal behavior by fusing poor sources of data that might be incorrect or give an incomplete picture.[129]

B. Fairness in Predictive Algorithms

The ProPublica-COMPAS debate questioned what fairness means and how each definition could be mathematically specified. Fairness could be defined as treating everyone the same or it could be defined as giving everyone similar outcomes. Similar outcomes may require that statistical treatments vary. Variation by race or

125. *See* Alexandra Chouldechova, *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*, 5 BIG DATA 153 (2017).

126. *See Data and Analysis*, *supra* note 21.

127. *See* Kiel Brennan-Marquez, *"Plausible Cause": Explanatory Standards in the Age of Powerful Machines*, 70 VAND. L. REV. 1249, 1265–73 (2017); Anthony W. Flores et al., *False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks.,"* 80 FED. PROBATION 38, 38 (2016).

128. *See* Katherine Freeman, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in* State v. Loomis, 18 N.C. J.L. & TECH. 75, 83–86 (2016); Jeffrey Johnson, *The Question of Information Justice*, COMMS. ACM, Mar. 2016, at 27, 27–29.

129. Selbst, *supra* note 86.

gender may improve statistical outcomes but raises equal protection concerns.[130]

The research underscores the multiple ways that it is possible to describe treatment of people within sets and call it fair. The research led by Jon Kleinberg presents three conditions that could denote fairness: (1) calibration;[131] (2) balancing negative impact; and (3) balancing positive impact.[132] Kleinberg includes mathematical proofs that show that it is not possible to simultaneously have all three conditions at once.[133] The research led by Sam Corbett-Davis considers what fairness means by running tests that avoid race-specific characteristics or including them.[134] They also discuss the problem of giving special treatment in the database to protected classes. Alexandra Chouldechova gives a well-argued comprehensive view of ways to define fairness mathematically, providing more alternatives than Kleinberg et al. does.[135] Chouldechova provides the proofs along with citations to a wide range of literature that discusses each idea further. The team lead by Sarah Tan developed techniques to detect bias by evaluating statistical differences in outcome variables.[136]

There is no single mathematical definition of fairness. The people developing a "fair" algorithm must decide on the uniformity or variation that is necessary for a functioning system.[137] Data science experts conclude that the people who control the algorithms define fairness.

C. Explainable Data Science

Data-driven organizations, including governments, thrive on finding unusual data sources and complex algorithms to create predictions.[138] Algorithms in public service, however, have a special need to be understood by the general public through models with

130. Logan & Ferguson, *supra* note 85; Pasquale, *supra* note 98.

131. *See generally* Pleiss et al., *supra* note 87 (defining calibration).

132. Kleinberg et al., *supra* note 123, at 2–4.

133. Kleinberg does introduce one hypothetical condition where it is possible to meet all three conditions. The trade-offs disappear if all populations have equal base rates. This means the groups are essentially identical in distribution and behavior. Only the label changes. In national risk assessment data, the individuals in the black and white sets have different base rates of recidivism. *Id.* at 5–6, 17.

134. Corbett-Davies et al., *A Computer Program Used for Bail*, *supra* note 122.

135. Chouldechova, *supra* note 125.

136. *See generally* Sarah Tan et al., *Detecting Bias in Black-Box Models Using Transparent Model Distillation*, 2018 AAI/ACM CONF. ON ARTIFICIAL INTELLIGENCE, ETHICS, & SOC'Y 96 (2018).

137. Since 2015, the annual Fairness Accountability and Transparency conferences investigate new concerns machine learning and algorithms. *See ACM Conference on Fairness, Accountability, and Transparency (ACM FAT)*, ACM FAT CONF., http://fatconference.org [https://perma.cc/89GL-RG7N].

138. *See generally* Judie Attard et al., *Value Creation on Open Government Data*, 49 HAW. INT'L CONF. ON SYS. SCI. 2605 (2016).

clear dependencies.[139] Scholars proved that clarity and simplicity could achieve comparable predictive results. The ProPublica-COMPAS debate advocated for risk assessments that could be explained to people who are not trained in data science. The solution proposed by these scholars was to explain the outcomes by revealing the relationships between essential factors.

Scholars demonstrated that accuracy could be maintained with only a handful of factors. Fewer factors achieved the same results but with a model that could easily be interpreted by non-experts. Elaine Angelino and her team argued against black box proprietary models because good results can be determined with simply a few factors.[140] James E. Johndrow & Kristian Lum found good predictive power with only seven essential characteristics.[141] Chris Baird, in a comparison of ten state juvenile assessment systems, found that simpler models outperformed more complex ones.[142] It is difficult to ascertain exactly how many factors the COMPAS model considers because different product versions were placed in different jurisdictions at different times. COMPAS makes use of significantly more factors than the academic studies.[143]

Data science experts prioritized simplicity. An algorithm that was explainable could achieve equal results to an algorithm that was complicated and not explainable.

D. Comparing Populations

The ProPublica-COMPAS debate emphasized the best statistical practices for comparing populations in models. Models guide how data that is input into an algorithm is processed into an output.[144] Models are often expressed as equations showing relationships between concepts. A risk assessment score is built on a model that abstracts behavioral data about past populations.[145]

The base rate is a vital indicator because it reflects actual population trends within the data set. Baird expressed concern about algorithms being used across jurisdictions because of changes in population base rates.[146] Predicting the likelihood of arrest is

139. *See* Jim Dwyer, *Showing the Algorithms Behind New York City Services*, N.Y TIMES, Aug. 24, 2017, at A18.

140. Elaine Angelino et al., *Learning Certifiably Optimal Rule Lists for Categorical Data*, 18 J. MACHINE LEARNING RES. 234 (2018).

141. James E. Johndrow & Kristian Lum, *An Algorithm for Removing Sensitive Information: Application to Race-Independent Recidivism Prediction*, ANNALS APPLIED STAT. (forthcoming 2019) (manuscript at 5), https://arxiv.org/pdf/1703.04957.pdf [http://perma.cc/KE2D-YGPP].

142. *See* BAIRD ET AL., *supra* note 79, at 134.

143. Dressel & Farid, *supra* note 94, at 1–2.

144. *See* Lehr & Ohm, *supra* note 2, at 671.

145. Lightbourne, *supra* note 119, at 329; Logan & Ferguson, *supra* note 85, at 554–56; Provost & Fawcett, *supra* note 102, at 52.

146. BAIRD ET AL., *supra* note 79, at 11. For a further discussion of base rates as a test of validity, see Dr. David Thompson's expert testimony on base rates as a test of

determined by the arrest rates for the population in the jurisdiction. Policing behavior will impact who is likely to be arrested and placed in the database for comparative analysis. For instance, the Anchorage Police in Alaska may have very different arrest patterns than the Boston Police Department in Massachusetts. Furthermore, populations in each place are different and will be reflected in different base rates. For example, it is unlikely that there are as many indigenous peoples in Boston, so the impact of certain weightings on that population will vary. The *Loomis* case discussed the potential problem of not validating[147] the COMPAS instruments with data from the state or jurisdiction.[148]

Data science scholars recognize the mathematical significance of base rates. Richard Berk and his team considered how to handle base rates of legally protected groups.[149] Chouldechova expands on this point by considering error rate balance.[150] Writing for Northpointe, William Dieterich and his team center their argument against ProPublica on predictive parity of population groups.[151] Flores strongly criticizes ProPublica for not following standard procedures known in the criminal justice community and provide a credible series of arguments about how the ProPublica researchers did not understand risk assessment populations.[152]

A risk assessment prediction is based, in part, on population base rates. Knowing the population distribution is essential to understanding how a risk assessment algorithm produces a predictive ranking. Data science experts agreed on the importance of comparing populations, keeping in mind base rate differences.

---

validity: "The Court does not know how the COMPAS compares that individual's history with the population that it's comparing them with. The Court doesn't even know whether that population is a Wisconsin population, a New York population, a California population . . . . There's all kinds of information that the court doesn't have, and what we're doing is we're mis-informing the court when we put these graphs in front of them and let them use it for sentence." State v. Loomis, 881 N.W.2d 749, 756–57 (Wis. 2016).

147. For a discussion on validation in data science, see Galen Panger, *Reassessing the Facebook Experiment: Critical Thinking About the Validity of Big Data Research*, 19 INFO., COMM. & SOC'Y 1108 (2016).

148. *Loomis*, 881 N.W.2d at 762–63.

149. Richard Berk et al., *Fairness in Criminal Justice Risk Assessments: The State of the Art* 3–8, 18–19 (Univ. of Pa. Dep't of Criminology, Working Paper No. 2017-1.0, 2017), https://crim.sas.upenn.edu/sites/default/files/2017-1.0-Berk_FairnessCrimJustRisk.pdf [http://perma.cc/U6B9-4JLL].

150. Chouldechova, *supra* note 125, at 135–37.

151. DIETERICH ET AL., *supra* note 18, at 9 ("A risk scale exhibits accuracy equity if it can discriminate recidivists and non-recidivists equally well for two different groups such as blacks and whites. The risk scale exhibits predictive parity if the classifier obtains similar predictive values for two different groups such as blacks and whites, for example, the probability of recidivating, given a high risk score, is similar for blacks and whites.").

152. Flores et al., *supra* note 127, at 40.

IV. A PROPOSAL: ALTERNATIVE CLAIMS FOR *LOOMIS*

This Section speculates about claims that could have been made in *Loomis* if the organizational context for risk assessment were part of the debate. A more robust argument about predictive assessment could be made if all sides shared information about the algorithm's origin and norms of practice. Understanding these two points makes it possible to articulate design intentions for the algorithm. Ideally, some form of data could be used to demonstrate and substantiate any of the above points without violating privacy. This Section envisions claims about provenance, implementation practices, and training sets.

The claims I propose below are offered as tools for anyone concerned with risk assessments. The claims might be used by the State to support a risk assessment or by the defendant to challenge it. The organizations that create predictive assessments may consider following these proposals to improve their credibility. Third-party applications can maintain their trade secrets and still meet one or all of these proposals. Northpointe's active participation in the ProPublica-COMPAS debate illustrates the ability to maintain proprietary claims and still address public concerns. Most importantly, criminal justice organizations that are purchasing these systems could use these proposals as criteria during acquisition and procurement. Advanced data analytics tools are expensive, and the public has a right to understand how government money is spent. The proposals presented in this article are intended to decrease the possibility of waste, fraud, and abuse and to increase accountability of government technology in the criminal justice system.

A. Provenance

Provenance establishes the value of an item by documenting its history and ownership. Buyers of fine art use provenance to trace paintings and sculptures across centuries of owners. Although usually considered for data that moves through multiple systems, provenance can also apply to the origins, ownership, and history of any digital asset.[153] Provenance could provide a linkage between who created the algorithm and who currently owns it. How were the COMPAS products introduced to the organization? How long has the Wisconsin circuit court used COMPAS predictive risk assessments? When was the product last updated? Has the algorithm been specially calibrated for Wisconsin populations or is it the standard version of the product? The provenance of the

153. For more on digital provenance, see Lucian Carata et al., *A Primer on Provenance*, 12 ACMQUEUE, Apr. 10, 2014, at 1, and Luc Moreau et al., *The Provenance of Electronic Data*, COMMS. ACM, Apr. 2008, at 52, 54–58.

COMPAS algorithm in Wisconsin could support or challenge the assertion that the sentencing court appropriately employed predictive risk assessments in *Loomis*.

Because predictive algorithms are products designed by organizations for organizations, changes in either the creator or the consumer might change aspects of the algorithm. Each new organization may reflect changes in strategy and business models that impact the algorithm design.[154] The original purpose of the algorithm prediction design would significantly impact how the system is optimized. It might be possible to infer basic design requirements by knowing who created the algorithm and who purchased it and when. The provenance of the algorithm could affirm that the assessments were created by an organization with an appropriate track record.

The provenance of the COMPAS algorithm is complex. COMPAS was originally built by Northpointe Institute for Public Management which was established in 1989 and Northpointe was subsequently acquired by Volaris Systems Group in 2011.[155] The Northpointe subdivision was rebranded as equivant in 2017.[156] Each transfer of ownership is an opportunity to lose organizational knowledge about how the algorithm functions. In the least, these changes might impact the quality of documentation about the design. Provenance is critical when using third-party commercial vendors who may change names and business models.[157] Although not an urgent concern today, in a few decades the history of these algorithms will become increasingly important.[158]

B. Practice

Norms of practice specify how technology is used and implemented. While the *Loomis* decision did recognize existing practice, it merely pointed to language in documents.[159] Stronger support for the practice would have been some indication of training

154. Erna H.J.M. Ruijer & Richard F. Huff, *Breaking through Barriers: The Impact of Organizational Culture on Open Government* Reform, 10 TRANSFORMING GOV'T 335 (2016).

155. *Volaris Group Acquires Northpointe Institute for Public Management,* VOLARIS GROUP (May 03, 2011), https://www.volarisgroup.com/news/article/volaris-group-acquires-northpointe-institute-for-public-management [http://perma.cc/QE4J-B8J5].

156. *CourtView, Constellation, & Northpointe Re-Brand to equivant*, *supra* note 5.

157. For a discussion concerns about using businesses that have different goals and time-scale to long-standing public institutions, see Klievink et al., *supra* note 102, at 72–73.

158. *See generally* Cecelia Klingele, *The Promises and Perils of Evidence-Based Corrections*, 91 NOTRE DAME L. REV. 537, 538–41 (2016) (explaining that the criminal justice system undergoes periodic reforms as values and science change).

159. State v. Loomis, 881 N.W.2d 749, 764 (Wis. 2016) (making references to the State of Wisconsin Department of Corrections Electronic Case Reference Manual, COMPAS Assessment Frequently Asked Questions, and a Practitioner's Guide to COMPAS Core published by Northpointe).

and familiarity with the system. Where was COMPAS first used in Wisconsin? Why did the COMPAS product meet their requirement needs? What other systems did the state consider and why did they not meet their needs? How are employees trained to use COMPAS? How long had the State of Wisconsin been using predictive risk assessments? How long had the Wisconsin circuit court been using predictions in sentencing? When was the decision made to use it throughout the court systems? These are not unusual questions when considering how digital government products go through procurement in the federal executive branch.[160] Although the circuit court in *Loomis* discussed its commitment to evidence-based sentencing, more specific details about implementation and evaluation could have shown that the court had integrated COMPAS risk assessment into normative organizational practices.[161]

The *Loomis* decision was factually correct that both sides had access to the same information because they both had documents that contained the score and questionnaire. Yet, there was a subtle difference in access to information for each party in *Loomis*. The circuit court and the defense had access to the same documents, but not the same context for the information contained in those documents. Anyone with the opportunity to see how COMPAS scores were applied to hundreds of people over time could develop an inductive understanding of what is important on the questionnaire and how it is applied through the Wisconsin population. People external to an organization are unlikely to understand what intuitive internal needs are.

The court probably takes for granted how things work and what aspects of their work they prioritize. Employees of the Wisconsin circuit court, including the judge, would be familiar with the administrative goals of court operations. The COMPAS algorithm is a product marketed to target exactly the efficiency concerns of courts. As a part of an organizational information system, predictive assessments are designed to support employees of criminal justice organizations. The courts have a better grasp on risk thresholds because they see these scores applied over time. Because of this unique retrospective view, the state might have been capable of singularly corroborating its own connection between a COMPAS risk assessment and specific questions while the defendant did not have that inductive experience.

160. Peter Johnson & Pamela Robinson, *Civic Hackathons: Innovation, Procurement, or Civic Engagement?*, 31 REV. POL'Y RES. 349, 352 (2014).

161. A field of public policy, the science of science policy, considers how to evaluate digital investments like this. *See* Sandra Braman, *Technology and Epistemology: Information Policy and Desire*, *in* CULTURAL TECHNOLOGIES: THE SHAPING OF CULTURE IN MEDIA AND SOCIETY 133 (Göran Bolin ed., 2012); Maryann Feldman et al., *The New Data Frontier*, 44 RES. POL'Y 1629 (2015).

C. Training Data

Many public organizations are making digital information available to the public as open data.[162] Training data is one type of open data that could be useful to establish the validity of risk assessment scores. Generated sample data is another type of open data that could be used to establish claims. With shared data, either side in *Loomis* could have engaged statisticians to prove their claims using the same data set.

Training data calibrate and optimize algorithms.[163] In a 2009 article, Northpointe employees pointed out that they trained the COMPAS algorithm on a population sample of 2,328 that was 76% white and only 19% female.[164] Anyone who wanted to argue with COMPAS could point to the training populations in this older article to question the validity of current predictions given their own jurisdiction demographics. Furthermore, knowing the training set could support a claim that the defendant was an outlier and therefore may be easily misclassified.[165] Algorithms use training data as a benchmark for speed, accuracy, or other optimization goals. A training set could also confirm how the algorithm considered characteristics represented by the defendant. A characteristic less probable in the training data could be an argument for a less accurate prediction.

Generated data reflect representative practices and are used in quality assurance to test the breadth of design requirements.[166] Given the sensitivity of criminal justice data, it might be reasonable to generate a data set that would allow for hypothetical testing. Unlike training data, generated data are designed to test the robustness of the system to handle a range of cases. Generated data could confirm how the algorithm handles unusual or under-represented factors in the data. In *Loomis*, Wisconsin could have provided generated data about their populations, or COMPAS could

---

162. Open data are internal organizational files released to the general public usually through the Internet. *See* Anne L. Washington & David Morar, *Open Government Data and File Formats: Constraints on Collaboration*, 18 PROC. INT'L CONF. ON DIGITAL GOV'T RES. 155 (2017).

163. Tom Dietterich, *Overfitting and Undercomputing in Machine Learning*, 27 ACM COMPUTING SURVS. 326, 326–27 (1995); Leslie Scism & Mark Maremont, *Insurers Test Data Profiles to Identify Risky Clients*, WALL ST. J. (Nov. 19, 2010, 12:01 AM), http://www.wsj.com/articles/SB10001424052748704648604575620750998072986 [http://perma.cc/9B6Y-C34S].

164. Brennan et al., *Evaluating the Predictive Validity*, *supra* note 43.

165. An outlier is an observation beyond the general data trend. Algorithms can exclude large segments of society if variations in human populations are not considered. In a series of experiments, computer scientists showed that facial recognition software could have high overall average success but failed at the intersection of gender and skin color. *See* Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACHINE LEARNING RES. 77 (2018).

166. To understand the role of generated data in software engineering, see D. C. Ince, *The Automatic Generation of Test Data,* 30 COMPUTER J. 63 (1987).

have verified their algorithms with Wisconsin population data. A risk assessment reflects a probability that is based on several factors that can vary over time, location, policy environment, or policing behavior. Generated data could clarify how the algorithm handles both unusual outliers and the expected populations. The companies creating risk assessments might consider releasing generated data to help the public and their clients evaluate if they are meeting the standards expected for public sector technology.

The ProPublica-COMPAS debate would not have been possible without open data. ProPublica made their results open for scrutiny by releasing their datasets and writing about their methodology. The open data available from ProPublica was paltry, only 11,000 records,[167] yet it was successful in inspiring hundreds of publications. Future debates on predictive algorithms would be served with access to training data or generated data.

When a court uses proprietary software that is not supported by evidence of validity or open data sources, a defendant does not have sufficient information to deny or explain a prediction.[168] The defendant is in the best position to review the input values, as affirmed in *Loomis,* but the defendant, contrary to *Loomis,* cannot substantially challenge the output of a risk assessment. The criminal justice organization that purchased the decision-making tool solely for its own efficiency needs is in the best position to justify how one individual's assessment matches the algorithm designed for its own administrative needs. Open data, either training data or generated data, could have been used by either side in *Loomis* to make stronger proofs of their claims.

D. Data Science Reasoning

As data science is used in the public sector for vital practices involving human lives, a new form of reasoning is needed to explain and justify algorithmic results. Data science reasoning seeks logical connections between input data, algorithmic procedures, and output interpretations while recognizing that people and organizations make choices at each stage. The court in *Loomis* focused on two pieces of information fed into the predictive algorithm: the score and the questionnaire.[169] Explanations limited to technology and information must be enlarged to embrace the organizational context and daily practice.

Data science reasoning encourages exact descriptions of how algorithms support human decision-making. This is also an

167. *See Data and Analysis*, *supra* note 21.

168. *See* Katherine Freeman, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in* State v. Loomis, 18 N.C. J.L. & TECH. 75, 94 (2016).

169. State v. Loomis, 881 N.W.2d 749, 760–65 (Wis. 2016).

antidote to the psychological tendency of humans to rely on machines instead of making individual decisions.[170] The *Loomis* court recognized three cautions of predictive risk assessment, yet failed to recognize that technology changes the nature of their own work.[171] Research on science and technology studies has shown that technology also impacts expert opinions.[172] How can courts avoid the inherent preference people give numbers and calculated sources? Although there is some recent research on the implementation of risk assessment scores,[173] more research is needed to understand how technology impacts the criminal justice workforce.

While these proposals are intended to assist with legal issues, these matters are equally important to those learning how to become data scientists. Data science reasoning is a term introduced to capture the anticipated growth of critical thinking in data science curriculum.

CONCLUSION

The purpose of this article was to examine the ProPublica-COMPAS debate over risk assessment algorithms. My analysis of the arguments contributes the following three points.

First, the standard set in *Loomis* for challenging a predictive assessment, by reviewing data accuracy, was not supported in the data science literature. Accuracy is just one of many data qualities and does not address how the algorithm produces results or manages the input data.[174] The data quality standard set in *Loomis* is a very low bar for understanding predictive risk assessment. The ProPublica-COMPAS debate revealed claims of fairness, simplicity, and population comparison that can the basis for arguments about predictive algorithms. Accurate data is a necessary but not sufficient standard for assessing the integrity of a risk assessment prediction.

Second, a healthy debate on predictive analytics required shared information such as open data, standard evaluation practices, shared designs, and hypotheses. Sharing information

170. In human factors engineering, the phenomenon is known as Automation Bias. *See* Citron, *supra* note 28, at 1252. For an extensive consideration of the impact of procedural systems before computing, see ANTHONY G. AMSTERDAM & JEROME S. BRUNER, MINDING THE LAW (2000).

171. *Loomis*, 881 N.W.2d at 754.

172. *See generally* Stephen R. Barley, *Technology as an Occasion for Structuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Departments*, 31 ADMIN. SCI. Q. 78 (1986) (suggesting that technology can alter organizational structures by altering roles and patterns of interaction).

173. *See* Angèle Christin, *Algorithms in Practice: Comparing Web Journalism and Criminal Justice*, 4 BIG DATA & SOC'Y 1 (2016) (examining algorithms used in web journalism and criminal justice).

174. *Supra* Section II.C.

about the production and management of risk assessment algorithms could additionally support claims in a courtroom debate.

Third, some algorithms are efficiency tools that are designed to meet the specific needs of an organization's business process. Extra effort is required to make internal predictive algorithms more legible to people outside the organization. Anyone who works inside the criminal justice system, therefore, has a slight advantage to understanding how algorithms work to support their organization's daily operations.

The operational speed that comes with automation also brings sufficient obscurity to raise concerns about equal access to information in an adversarial legal system. The controversy over risk assessment algorithms hints at whether procedural due process is the cost of automating a criminal justice system that is operating at administrative capacity. As predictive tools develop, society and scholars will need a new form of reasoning, data science reasoning, that can serve to facilitate conversations or arguments with algorithms.