PANDORA'S ALGORITHMIC BLACK BOX: THE CHALLENGES OF USING ALGORITHMIC RISK ASSESSMENTS IN SENTENCING

Leah Wisser*

Introduction: What is Black Box Algorithmic Sentencing?

It is the year 2050. Sarah has been convicted for driving the "getaway" vehicle in a robbery. Sarah is a good kid, she just got in with the wrong crowd. As she walks into the courtroom for sentencing, she vows to herself to never commit another crime. She wants to turn her life around. Sarah looks around the courtroom. She is alone. A screen sits in front of her. "Begin," it says. She starts to fill out a list of her personal information. Age: 18. Level of education: low. Socioeconomic status: poor. Gender: female. Prior convictions: none. Prior arrests: none. Prior police-initiated contact: 20. Sarah scoffs at how many times she has been unfairly profiled and stopped by the police, but she is honest. The list goes on and on. Eventually, she clicks "Done." The screen turns red. Alarms start blaring. "High-risk!" "High-risk!" An automated voice screams. "Maximum sentence!" is plastered across the screen as handcuffs click around Sarah's wrists. "But, why?" she asks to an empty, automated room.

This is a picture that many may find ludicrous, like an excerpt from a science fiction novel. However, it is less exaggerated than it appears. Algorithmic risk assessments, a new automated way to calculate recidivism risk, take a chilling step toward making this fiction a reality.[1]

Judges have factored recidivism predictions into sentencing for decades.[2] Using recidivism risk as a factor in sentencing is itself controversial.[3] The American Bar Association ("ABA") encourages using recidivism predictions, explaining that placement of criminals with a low-risk of recidivism ("low-risk") with criminals with a high-risk of recidivism ("high-risk") decreases the probability of rehabilitation for low-risk offenders.[4] Another scholar argues that the practice is not only non-progressive and non-scientific, but also verges on unconstitutional.[5]

---

* Georgetown University Law Center, J.D. Candidate 2019. I would like to thank Professor Julie Cohen for her academic support and feedback on this Note.

1. *Algorithms in the Criminal Justice System*, Electronic Priv. Info. Ctr., https://epic.org/algorithmic-transparency/crim-justice/ (last visited Feb. 20, 2019).

2. *See* John Monahan, *Risk Assessment in Sentencing*, 4 Reforming Criminal Justice 77, 79 (Erik Luna ed., 2017), http://academyforjustice.org/wp-content/uploads/2017/10/5_Criminal_Justice_Reform_Vol_4_Risk-Assessment-in-Sentencing.pdf.

3. *See generally id.*

4. American Bar Association, State Policy Implementation Project (2011), https://www.americanbar.org/content/dam/aba/administrative/criminal_justice/spip_handouts.authcheckdam.pdf.

5. Sonja Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 Stan. L. Rev. 803, 805 (2014).

This Note will not delve deeply into the advantages and disadvantages of using risk assessments in general. Instead, it will explore what happens when society allows these risk assessments to slip behind an algorithm's black curtain. Risk assessment algorithms often input more than one hundred personal characteristics—such as socioeconomic status, age, sex, geography, background, employment status, and neighborhood crime—in order to predict whether a convicted person will reoffend.[6] Based on these inputs, the convicted person is assigned a score between one and ten—one being low-risk and ten being high-risk.[7] There is little consistency in the algorithms used across the country.[8] States and even counties within the United States differ on which tools they use.[9] Three main systems are most prevalent, with states often adapting a unique version of one of the three for their own specific use.[10] This Note will focus on COMPAS, one of the most commonly used algorithms nationwide, when discussing the inner workings of the algorithm in more detail.[11]

If society continues to use algorithmic risk assessments as they are currently deployed, without the proper limitations and oversight, vast opacity will inevitably cloud our otherwise transparent criminal justice system and risk the introduction of various forms of bias. There is vast misunderstanding about how these algorithms work, both by society at large and the very judges who factor the risk scores into their sentencing decisions. Members of the Senate have urged the United States Sentencing Commission to conduct an independent study of the innerworkings of these algorithms and to "issue a policy statement to guide jurisdictions implementing these tools."[12] The Senate members raised concerns about fairness, racial discrimination, and lack of transparency.[13] We must protect defendants and our justice system's integrity from these algorithms' flaws. Unfortunately, our constitutional framework does not provide us with the appropriate tools to address the problematic nature of algorithmic risk assessments. Therefore, I offer an administrative solution as a better way of addressing these concerns.

---

6. *See* NORTHPOINTE, RISK ASSESSMENT (2011), https://assets.documentcloud.org/documents/2702103/Sample-Risk-Assessment-COMPAS-CORE.pdf (providing a sample of inputs).

7. 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: MONKEY CAGE (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/?utm_term=.12944835bae4.

8. For a state-by-state table on the varying algorithms use, please see *Algorithms in the Criminal Justice System*, *supra* note 1.

9. *Id.*

10. *Id.* The main systems are as follows: Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), Public Safety Assessment (PSA), and Level of Service Inventory Revised (LSI-R).

11. *Id.* Ten percent of states use COMPAS.

12. Letter from Senators Cory Booker & Brian Schatz to Jefferson Sessions, Attorney General, and Judge William Pryor, Chair of the United States Sentencing Commission (Feb. 5, 2018), https://www.booker.senate.gov/?p=press_release&id=745.

13. *Id.*

This Note is organized into five sections. Section I will discuss how algorithmic risk assessments work and present a case of their use in court. Sections II, III, and IV will explore the inability of the Equal Protection Clause of the Fourteenth Amendment, the Cruel and Unusual Punishments Clause of the Eighth Amendment, and the Due Process Clauses of the Fifth and Fourteenth Amendments to adequately address these algorithms' problematic nature. Section V will propose a solution of regulatory oversight to better address the algorithmic risk assessment problem.

I. ALGORITHMIC RISK ASSESSMENTS AT WORK IN *STATE V. LOOMIS*

This section will explain the functionality of algorithmic risk assessments through the leading court case in which they were deployed: *State v. Loomis*.[14] Explaining how algorithmic risk assessments work is not an easy task. Part of the difficulty arises from the fact that the source codes of many of these algorithms are proprietary.[15] As Rebecca Wexler notes, the "introduction of [a trade secret privilege] into the criminal justice system raises . . . tensions between life, liberty, and property interests."[16] The algorithms' opacity makes it impossible for society to test their accuracy and validity to ensure they are not inappropriately weighing specific variables that have a disparate impact on minorities. If the algorithms are inaccurate or discriminatory, society cannot hold them accountable because they are computer programs. Currently, an effective legal avenue does not exist to hold their creators accountable for the algorithms' flaws.

The leading case on the validity of algorithmic risk assessments, *State v. Loomis*,[17] was decided by the Wisconsin Supreme Court. Eric Loomis was allegedly the driver of a car involved in a drive-by shooting.[18] He was charged with five counts, all as a repeat offender:

> (1) First-degree recklessly endangering safety . . .; (2) Attempting to flee or elude a traffic officer . . .; (3) Operating a motor vehicle without the owner's consent; (4) Possession of a firearm by a felon . . .; (5) Possession of a short-barreled shotgun or rifle . . . .[19]

14. 881 N.W.2d 749 (Wis. 2016).

15. Ed Yong, *A Popular Algorithm Is No Better at Predicting Crimes Than Random People*, THE ATLANTIC (Jan. 17, 2018), https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/.

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

17. 881 N.W.2d 749 (Wis. 2016). For governmental awareness of the problematic nature of algorithmic risk assessments, please see Letter from Senators Cory Booker & Brian Schatz, *supra* note 12; N.Y. Committee on Technology Meeting, N.Y. CITY COUNCIL (Oct. 16, 2017), http://legistar.council.nyc.gov/MeetingDetail.aspx?ID=564867&GUID=9567478C-C9F4-4EDE-89F2-947E95A94ACD&Search=.

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

19. *Id.*

Loomis denied any involvement in the shooting, claiming that he only drove the car after the shooting occurred.[20] He entered into a plea deal where he pled guilty to two of the lesser charges: "attempting to flee a traffic officer and operating a motor vehicle without the owner's consent."[21] The court accepted Loomis' plea and ordered a presentence investigation ("PSI").[22] The report included a COMPAS risk assessment.[23]

COMPAS was created and is owned by a for-profit company named Equivant (previously Northpointe).[24] COMPAS produces a risk assessment in a bar chart consisting of three bars: pre-trial release risk, general recidivism, and violent recidivism.[25] The COMPAS assessment categorized Loomis as "a high risk of recidivism" in all three categories.[26]

The State argued this assessment should be used in "determining an appropriate sentence" for Loomis.[27] The risk assessment was subsequently used as a reason to deny Loomis probation. The circuit court judge explained: "In terms of weighing the various factors, I'm ruling out probation because of the seriousness of the crime and because [the defendant's] history, [the defendant's] history on supervision, and the risk assessment tools that have been utilized, suggest that [the defendant is] extremely high risk to re-offend."[28]

Loomis challenged the use of COMPAS on three grounds. First, Loomis claimed the algorithm violated his right to be sentenced using accurate information (partially due to the proprietary aspect of the algorithm that denied the defendant an opportunity to assess its accuracy).[29] Second, Loomis challenged the algorithm as a violation of his right to have an individualized sentence.[30] Third, Loomis argued that the algorithm impermissibly considered gender when generating its scores.[31]

Although the court ultimately held for the State, its decision was narrow.[32] The court acknowledged the problematic nature of the algorithm, noting: "[a]lthough 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

20. *Id*.
21. *Id*.
22. *Id*.
23. *Id*.
24. *Practitioner's Guide to COMPAS Core*, EQUIVANT (Dec. 19, 2017), http://www.equivant.com/assets/img/content/Practitioners_Guide_COMPASCore_121917.pdf [hereinafter *Practitioner's Guide 2017*].
25. *Id*. at 29.
26. *Loomis*, 881 N.W.2d at 755.
27. *Id*.
28. *Id*.
29. *Id*. at 757.
30. *Id*.
31. *Id*.
32. *Id*. at 772.

potential due process violations."[33] These limitations include the following: courts may not use risk assessments as the determinative factor in sentencing, and courts are required to give the reasons why, in addition to the risk assessment, a specific sentence is chosen.[34] Courts may not use risk assessments to determine whether or not someone will be incarcerated or to determine the severity of a sentence.[35] The court also requires that a "written advisement" including limitations of and cautions about the COMPAS program be given to courts who use COMPAS.[36]

The *Loomis* decision is flawed because the court demonstrates a clear misunderstanding of how COMPAS functions and provides potentially impracticable guidelines for future judges with no mechanism through which we can ensure said guidelines are being followed. First, the court struggles to comprehend how COMPAS actually works.[37] Self-described as an "actuarial risk assessment instrument[]," COMPAS purports to be an "objective method of estimating the likelihood of reoffending."[38] Equivant asserts that the purpose of COMPAS risk scores is "to discriminate between offenders who will and will not recidivate."[39] This is troubling because COMPAS does not predict risk on an individual level.[40] Instead, the algorithm is designed to align the defendant with a group "of offenders who have similar characteristics" and predict risk based on "known outcomes" of said group.[41] In a 2012 version of its Practitioner's Guide, Equivant provided a section on how to understand a COMPAS score. "[A] COMPAS score tells you, relative to other offenders across the United States, the predicted risk of [the defendant]. . . If he scores a 4 on a [risk] scale, then 60% of the population looks more risky in that area than he does, and 30% looks less risky."[42] This explanation, along with the entire section on understanding a COMPAS score, has been removed from subsequent Practitioner's Guides.[43] Without a sufficient and clear explanation as to how the algorithm works, the court is understandably confused.[44]

---

33. *Id*. at 757.

34. *See id*. at 760; *see also id*. at 769 ("[A] circuit court must explain the factors in addition to a COMPAS risk assessment that independently support the sentence imposed.").

35. *Id*.

36. *Id*. at 769.

37. *Id*. at 774 (Abrahamson, J., concurring).

38. *Practitioner's Guide 2017*, *supra* note 24, at 32.

39. *Id*. at 7.

40. *See id*. at 34 ("Risk assessment is about predicting group behavior . . . it is not about prediction at the individual level.").

41. *Id*.

42. *Practitioner's Guide To COMPAS*, NORTHPOINTE 5 (Aug. 17, 2012), http://www.northpointeinc.com/files/technical_documents/FieldGuide2_081412.pdf [hereinafter *Practitioner's Guide 2012*].

43. *Compare Practitioner's Guide 2017*, *supra* note 24, *and Practitioner's Guide to COMPAS*, NORTHPOINTE (Mar. 19, 2015), http://www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-_031915.pdf, *with Practitioner's Guide 2012*, *supra* note 42.

44. State v. Loomis, 881 N.W.2d 749, 774 (Wis. 2016) (Abrahamson, J., concurring) (noting the "court's lack of understanding of COMPAS was a significant problem"). The opacity problem that creates a public lack of understanding of these algorithms equally affects judges since judges are not granted special access to the internal source code. Furthermore, sometimes even complex algorithm creators themselves struggle to

Second, many of the court's suggested limitations are impractical to employ and will have an illusory effect. As stated by one commentator: "the opinion mandates warnings and instructions that might, in reality, be hard for judges to actually follow."[45] For example, courts should not use COMPAS as the determinative factor in whether a defendant will be incarcerated nor should they use algorithms to determine the severity of a sentence.[46] However, despite the *Loomis* court's warnings, "all things being equal, a high-risk score will make it much less likely a person will get the minimum sentence or avoid incarceration."[47] Society has no way of knowing if a risk assessment was in fact the determinative factor in a judge's mind, and the *Loomis* court provides no test which would allow us to ensure the scores are used appropriately beyond prescribing that they must be.[48]

Finally, the *Loomis* court does not address trade secret concerns or the potential of seeing inside the black box. The algorithm's impenetrability is mainly what gives rise to concerns over its use. While we know what information is inputted into the algorithm as well as the score that is outputted, we do not know *how* the algorithm weights or processes the different inputs.[49] The weights matter, for they can make the algorithm inaccurate or discriminatory. As an exaggerated example to illustrate this point: if COMPAS inputs one hundred characteristics, but weights neighborhood crime and socioeconomic status as ten times more important than any other characteristic, the algorithm is likely to inappropriately and inaccurately deem minorities as higher risk.

Though the *Loomis* decision is largely unsatisfactory, the court does get a few things right. The *Loomis* court admits, explicitly and unequivocally, that algorithmic risk assessments raise serious concerns if they are used unfettered and unchecked.[50] The court struggles, however, to match any of the problems it finds with the protections Loomis claims to have under the Constitution.[51] This is because constitutional protections are not the appropriate tools to address this problem.

---

understand why their program made a single, discrete decision. For a discussion of creator confusion surrounding complex algorithms, please see Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. REV. (Apr. 11, 2017), https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.

45. Lauren Kirchner, *Wisconsin Court: Warning Labels Are Needed for Scores Rating Defendants' Risk of Future Crime*, PROPUBLICA (July 14, 2016, 10:35 AM), https://www.propublica.org/article/wisconsin-court-warning-labels-needed-scores-rating-risk-future-crime.

46. *Loomis*, 881 N.W.2d at 769.

47. Kirchner, *supra* note 45 (internal quotation marks omitted).

48. DANIELLE KEHL ET AL., ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM: ASSESSING THE USE OF RISK ASSESSMENTS IN SENTENCING 23 (2017), https://dash.harvard.edu/handle/1/33746041.

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

50. *Id*. at 763–64.

51. The Supreme Court denied Loomis' petition for certiorari in June of 2017, so there will be no additional guidance from our highest Court in Loomis' case. Supreme Court of the United States, Loomis v. Wisconsin, SCOTUSBLOG (June 26, 2017), http://www.scotusblog.com/case-files/cases/loomis-v-wisconsin/.

II. IS THERE NO PROTECTION PROVIDED BY THE EQUAL PROTECTION CLAUSE?

The Equal Protection Clause of the Fourteenth Amendment does not adequately address the problematic nature of algorithmic risk assessments. The Equal Protection Clause requires that the law treat all people equally.[52] In practice, however, problems of proof often defeat equal protection claims.[53] If there is unequal treatment under the law, the legislature is required to prove it has a "rational basis" for the law.[54] This is not a difficult standard to meet.[55] If there is unequal treatment of a quasi-suspect class—like gender—courts apply intermediate scrutiny where the legislature is required to show that the treatment furthers "important governmental objectives."[56] If there is unequal treatment of a suspect class—like race—the standard is higher: the legislature must show "a compelling governmental interest" that is "necessary" to achieving the government's objective.[57] To reach this strict scrutiny standard, the Supreme Court requires more than a "racially disproportionate impact;" there must also be a racially discriminatory purpose.[58] Demonstrating a racially suspect intention is a problem of proof that has haunted Fourteenth Amendment challenges over the years.[59] Furthermore, the burden of proof is on the complainant.[60]

Defendants challenging COMPAS may be able to prove a racially disproportionate impact but will be unable to show a racially discriminatory purpose. A racially disproportionate impact likely exists: studies have shown that the algorithm generates racially biased results, and many of COMPAS's other variables serve as proxies for race.[61] A racially discriminatory purpose, however, will be much more difficult to prove because COMPAS does not use race as a

---

52. U.S. CONST. amend. XIV, § 1.

53. *See* Arlington Heights v. Metro. Hous. Dev. Corp., 429 U.S. 252, 270 (1977).

54. Armour v. Indianapolis, 566 U.S. 673, 680 (2012) (citing Heller v. Doe, 509 U.S. 312, 319–20 (1993)); *see also* LINDA R. MONK, THE WORDS WE LIVE BY: YOUR ANNOTATED GUIDE TO THE CONSTITUTION 219 (2003).

55. *See Heller*, 509 U.S. at 333.

56. United States v. Virginia, 518 U.S. 515, 533 (1996) (internal quotation marks omitted) (citing Miss. Univ. for Women v. Hogan, 458 U.S. 718, 724 (1982)).

57. Palmore v. Sidoti, 466 U.S. 429, 432–33 (1984) (citing McLaughlin v. Florida, 379 U.S. 184, 196 (1964)).

58. Washington v. Davis, 426 U.S. 229, 239 (1976).

59. *See, e.g.*, Arlington Heights v. Metro. Hous. Dev. Corp., 429 U.S. 252, 270 (1977); United States v. Armstrong, 517 U.S. 456, 465, 469–70 (1996); *see also* Michael Selmi, *Was the Disparate Impact Theory a Mistake?*, 53 UCLA L. REV. 701, 732–34, 738–40 (2006) (explaining that *Washington v. Davis*' intent requirement limited the reach of disparate impact theory and providing statistics to show disparate impact cases are often unsuccessful).

60. *Id.*

61. Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, SCIENCE ADVANCES (Jan. 17, 2018), http://advances.sciencemag.org/content/4/1/eaao5580.full; *see also* Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (finding that "[b]lack defendants were . . . 77 percent more likely to be pegged as at higher risk of committing a future violent crime and 45 percent more likely to be predicted to commit a future crime of any kind"); KEHL ET AL., *supra* note 48, at 22. *But see* William Dieterich et al., COMPAS RISK SCALES: DEMONSTRATING ACCURACY EQUITY AND PREDICTIVE PARITY, NORTHPOINTE

variable.[62] When starting from a standpoint of racial proxies, defendants' arguments are already one step removed from race. Because the code is not available to the public, complainants would have no way of proving that said proxies are being weighted inappropriately so as to cause the disparate impact. Therefore, defendants are barred from accessing the only information that would show conclusively whether or not the algorithm was in fact intentionally racially defective. Even if the code were readily available, it may still be difficult to prove a discriminatory purpose because the problem with COMPAS may be implicit not intentional bias. It is unlikely that a programmer will write "unconstitutional code."[63] Defendants are fighting an unwinnable battle. Without a racially-biased purpose, the question of racial discrimination under the Fourteenth Amendment reverts to the rational basis test,[64] and this test would likely be satisfied by making an argument for increased efficiency or providing judges with additional information to inform their sentencing decisions.[65]

Although COMPAS uses gender as a variable, the algorithm's consideration of gender is not an equal protection violation because it promotes the accuracy of the algorithm's predictions. In *Craig v. Boren*, the Supreme Court held that an Oklahoma law treating males and females differently with respect to the drinking age violated the Fourteenth Amendment.[66] Although the law's differential treatment was grounded in statistical evidence showing the higher likelihood of male drunk driving versus female drunk driving, the Court held this law unconstitutional.[67] The Court acknowledged that the generalizations were empirically correct, but emphasized that the accuracy of the statistical data had no bearing on the applicability of the Fourteenth Amendment.[68]

As COMPAS scores are also based upon generalized, statistical data, the *Boren* decision seems to support a Fourteenth Amendment challenge to the algorithm. However, unlike the unequal treatment scrutinized in *Boren*, there is an important objective served by COMPAS's gender variable. Gender is not a suspect class; rather, gender receives intermediate scrutiny from courts—requiring a state to

---

Research Department 1 (July 8, 2016), http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf (refuting ProPublica study).

62. *See* NORTHPOINTE, *supra* note 6.

63. Ellora Israni, *Algorithmic Due Process: Mistaken Accountability and Attribution in* State v. Loomis, HARV. J.L. & TECH JOLT DIGEST (Aug. 31, 2017), https://jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1.

64. *See* Washington v. Davis, 426 U.S. 229, 246–47 (1976); Susannah W. Pollvogt, *Unconstitutional Animus*, 81 FORDHAM L. REV. 887, 898 (2012) ("[T]he vast majority of equal protection claims will be subject only to rational basis review.").

65. *See* Heller v. Doe, 509 U.S. 312, 320 (giving strong deference to State's classification if there is some legitimate governmental purpose); *cf.* State v. Loomis, 881 N.W.2d 749, 759 (Wis. 2016) ("The need to have additional sound information is apparent . . . for sentencing courts.").

66. Craig v. Boren, 429 U.S. 190, 192–93 (1976).

67. *Id.* at 208.

68. *Id.* at 208–09.

show "at least that the [challenged] classification serves important governmental objectives and that the discriminatory means employed are substantially related to the achievement of those objectives."[69] The court in *Loomis* provides, in dicta, what that objective could be: the promotion of accuracy.[70] The court noted that because the gender variable "promotes accuracy," it serves an institutional purpose "rather than a discriminatory purpose."[71] Furthermore, the discriminatory means applied are necessary for the achievement of said objective because "any risk assessment tool which fails to differentiate between men and women will misclassify both genders."[72]

Socioeconomic status, which historically has been insufficient grounds for an equal protection claim, is the third and weakest Fourteenth Amendment challenge. Socioeconomic status is not a suspect class.[73] The furthest the Supreme Court has gone in protecting socioeconomic groups in a sentencing context is to hold that punishment cannot be based solely upon a defendant's economic status.[74]

On the question of recidivism specifically, the Court has decided that a judge may, in fact, consider a defendant's finances when determining a sentence.[75] As socioeconomic status is only one factor considered by COMPAS, the program is constitutional on this front.

The difficulty of a Fourteenth Amendment challenge to COMPAS does not mean that the algorithm's disparate treatment is unproblematic. Rule of Law principles require that laws be "evenly enforced."[76] Laws are not evenly enforced when punishment depends on race, gender, or socioeconomic status, rather than the crime itself. Understanding how the program is trained may reveal that the algorithm is inappropriately weighting a race proxy variable so as to create racist results.[77] An algorithm's creator selects which data the algorithm will be trained on.[78] If the training data selected "reflect[s] existing human biases," the algorithm will reflect these very same biases itself.[79] Nonetheless, the Fourteenth Amendment is not the appropriate tool to address these issues.

---

69. United States v. Virginia, 518 U.S. 515, 533 (1996) (internal quotation marks omitted) (citing Miss. Univ. for Women v. Hogan, 458 U.S. 718, 724 (1982)).

70. State v. Loomis, 881 N.W.2d 749, 766 (Wis. 2016).

71. *Id*.

72. *Id*. at 766–67.

73. *See* Maher v. Roe, 432 U.S. 464, 471 (1977).

74. *See* Bearden v. Georgia, 461 U.S. 660, 671 (1983).

75. *Id*.

76. American Bar Ass'n, Division for Public Education, *What is the Rule of Law*, DIALOGUE ON THE RULE OF LAW, https://www.americanbar.org/content/dam/aba/migrated/publiced/features/Part1DialogueROL.authcheckdam.pdf (last visited Feb. 20, 2019).

77. Israni, *supra* note 63.

78. Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 680–88 (2016).

79. David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 665 (2017); *see also* Barocas & Selbst, *supra* note 78, at 683–84.

III. IS A COMPUTER-GENERATED RISK SCORE CRUEL AND UNUSUAL?

Much like the Fourteenth Amendment protections, the Eighth Amendment protection from cruel and unusual punishments does not appropriately address the issues caused by algorithmic risk assessments. Firstly, although the Eighth Amendment requires a soft proportionality requirement, the proportionality argument as it relates to COMPAS is likely foreclosed.[80] In *Harmelin v. Michigan*, several justices concluded that the Eighth Amendment's "Cruel and Unusual Punishments Clause encompasses a narrow proportionality principle" that "applies to noncapital sentences."[81] Using COMPAS is problematic, then, because the convicted person is openly being punished for actions beyond her instant crime, thus the punishment cannot be said to be proportionate in any meaningful way to the crime actually committed. However, this criticism speaks to risk assessments generally and, even if evidently fairly inaccurate, risk assessments have been allowed in courts since the 1980s.[82] Furthermore, the Supreme Court has infrequently held a punishment for noncapital crimes to be disproportionate.[83] The Court instead chooses to defer judgement to the legislature.[84]

The Eighth Amendment provides another avenue to revisit discriminatory effects, primarily in capital offense cases, but it too is unable to solve the COMPAS problem. Eighth Amendment discrimination jurisprudence has been largely confined to cases involving the death penalty. In *Furman v. Georgia*, the Supreme Court emphasized that the death penalty was "disproportionately imposed and carried out on the poor, the Negro, and the members of unpopular groups."[85] Therefore, the racial bias of COMPAS may create a constitutional deficiency for its discriminatory results under the Eighth Amendment. That being said, *Furman* dealt with capital punishment, which has received specialized treatment and added judicial protection in Eighth Amendment jurisprudence.[86] Therefore, *Furman*'s rationale is only weakly transferable to noncapital offenses.

---

80. *See* MONK, *supra* note 54, at 187.

81. Harmelin v. Michigan, 501 U.S. 957, 997 (1991) (Kennedy, J., concurring in part and concurring in the judgment).

82. *See* Barefoot v. Estelle, 463 U.S. 880, 884–85 (1983) (holding that psychiatric evidence as to defendant's potential future dangerousness was admissible even if potentially inaccurate and the evaluation of accuracy was left to the jury). *See generally* Kirk Heilbrun et al., *Risk Assessment for Future Offending: The Value and Limits of Expert Evidence at Sentencing*, 53 CT. REV. 116 (2017).

83. MONK, *supra* note 54, at 187.

84. *Id*.

85. Furman v. Georgia, 408 U.S. 238, 249–50 (1972) (Douglas, J., concurring). *Furman* also has a complicated legacy following the Supreme Court's decision in Gregg v. Georgia, 428 U.S. 153 (1976). *See* Sherod Thaxton, *Un-Gregg-ulated: Capital Charging and the Missing Mandate of* Gregg v. Georgia, 11 DUKE J. CONST. L. & PUB. POL'Y 145, 145–46 (2016).

86. *See Furman*, 408 U.S. 238 at 286 (Brennan, J., concurring) ("Death is a unique punishment . . . . [D]eath is the ultimate sanction. . . . No other punishment has been so continuously restricted. . . . Juries, of course, have always treated death cases differently, as have governors exercising their commutation powers."); *cf.* Gardner v. Florida, 430 U.S. 349, 357 (1977) ("[D]eath is a different kind of punishment from any other which may be imposed in this country.").

The closest the Eighth Amendment may come to addressing the algorithmic risk assessment problem is through the concept of arbitrary punishment. This constitutional tool's efficacy depends on how courts will choose to define arbitrary. In the dicta of *Furman*, Justice Douglas states: "[a] penalty . . . should be considered 'unusually' imposed if it is administered arbitrarily or discriminatorily."[87] Whether using patterned-based learning is arbitrary is a definitional choice the courts must make. On one hand, there does exist a rationale to the algorithm. The very fact that it is patterned and statistical means that it is reasoned. However, Rule of Law principles require a "fair, robust . . . legal process."[88]

It is difficult to defend an algorithm as "fair"—or at least as more than arbitrary—when a study found that COMPAS was no more accurate at predicting recidivism than a group of random volunteers with "no criminal justice experience [who were] provided with only the defendant's age, sex and criminal history."[89] The credibility and fairness of COMPAS is severely undermined when a group of random, non-experts with limited information can achieve the same level of accuracy as is achieved by the algorithm. Furthermore, judges are relying on an assessment they do not and cannot fully understand. The "reasoned" nature of the algorithm is not something the judges can personally verify. Therefore, the information is not necessarily accurate or applied in a meaningful way.

Whether or not these Rule of Law defects are sufficient to constitute arbitrariness for Eighth Amendment purposes, however, is indeterminate. Unfortunately, arbitrariness has remained a largely opaque concept in Eighth Amendment jurisprudence. For additional guidance, we can analogize to how "arbitrary" has been extensively defined in the administrative law context. Although not applied in criminal law, the Administrative Procedure Act allows courts to strike down any agency action deemed arbitrary and capricious.[90] In order for a court not to find an agency's decision to be arbitrary and capricious, the "agency must examine the relevant data and articulate a satisfactory explanation for its action including a 'rational connection between the facts found and the choice made.'"[91] Importantly, a decision is found to be arbitrary when the agency has "offered an explanation for its decision that runs counter to the evidence before the agency."[92] Therefore, the Court requires not only a reasoned consideration, but also an explanation. It can be argued that because of its proprietary nature, COMPAS does not come with a full explanation.

---

87. *Furman*, 408 U.S. 238 at 249 (Douglas, J., concurring) (quoting Arthur J. Goldberg & Alan M. Dershowitz, *Declaring the Death Penalty Unconstitutional*, 83 HARV. L. REV. 1773, 1790 (1970)).

88. American Bar Ass'n, *supra* note 76.

89. Dressel & Farid, *supra* note 61; Hannah Devlin, *Software 'No More Accurate Than Untrained Humans' at Judging Reoffending Risk*, THE GUARDIAN (Jan. 17, 2018), https://www.theguardian.com/us-news/2018/jan/17/software-no-more-accurate-than-untrained-humans-at-judging-reoffending-risk.

90. 5 U.S.C. § 706(2)(a) (2012).

91. Motor Vehicle Mfrs. Ass'n of U.S., Inc. v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 43 (1983) (quoting Burlington Truck Lines v. United States, 371 U.S. 156, 168 (1962)).

92. *Id.*

Unfortunately, we may never be able to prove that COMPAS actually produces results that are counter to the evidence it receives because the algorithm is predictive. We can never fully test the counterfactual. If the defendant is found to be high-risk, goes to jail, serves her time, is released and reoffends—the algorithm is proved right. However, if the defendant is found to be high-risk, goes to jail, serves her time, is released and does not reoffend—the algorithm is not necessarily proved wrong. Perhaps the jail time worked, and the defendant was rehabilitated or deterred from reoffending. Therefore, it may be impossible to know whether or not the score "runs counter to the evidence before" COMPAS.[93] Perhaps there is room for maneuver, but, as it is currently interpreted, the Eighth Amendment is unequipped to fully address the problematic nature of algorithmic risk assessments.

IV. IS ALGORITHMIC PROCESSING ENOUGH TO GUARANTEE DUE PROCESS OF LAW?

The Due Process Clause creates the highest hurdle for COMPAS to overcome, but still falls short of offering defendants adequate protection. The Due Process Clauses of the Fifth and Fourteenth Amendments guarantee a right to a fair trial: this includes requirements like understanding and having the ability to confront the evidence against you and receiving an individualized sentence.[94] Though due process rights are required in sentencing as well as during the trial itself, the rights are significantly weakened at the sentencing stage because the right to confront the evidence presented is not a guarantee.[95]

One due process challenge to COMPAS is that it erodes a defendant's right to be sentenced according to accurate information. This includes "the right to review and verify information contained in the PSI upon which the circuit court bases its sentencing decision."[96] This challenge becomes more complex, however, when we ask what kind of information and to what degree the defendant has the right to assess. Does the defendant only have the right to assess the inputs and outputs of the algorithm or also how said inputs are weighted so as to arrive at the final score? There is a "plausible distinction" between these two kinds of information, and it is unclear which type of information the defendant has a right to verify.[97]

---

93. *Id.*

94. U.S. CONST. amend. V; U.S. CONST. amend. XIV, §1; State v. Loomis, 881 N.W.2d 749, 757 (Wis. 2016); *see also* Frank Pasquale, *Secret Algorithms Threaten the Rule of Law*, MIT TECH. REVIEW (June 1, 2017), https://www.technologyreview.com/s/608011/secret-algorithms-threaten-the-rule-of-law/ ("Any court aware of foundational rule of law principles, as well as Fifth and Fourteenth Amendment principles of notice and explanation for decisions, would be very wary of permitting a state to base sentences (even if only in part) on a secret algorithm.").

95. *See* Alan Howard Aronson, *Constitutional Rights During the Sentencing Proceeding*, 69 J. CRIM. L. & CRIMINOLOGY 362 (1978). *Compare* Gardner v. Florida, 430 U.S. 349, 358 (1977) (walking back the severity of the *Williams v. New York* decision but not guaranteeing confrontation at sentencing), *with* Williams v. New York, 337 U.S. 241, 245, 251 (1949) (denying a right to confrontation at sentencing).

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

97. KEHL ET AL., *supra* note 48, at 23.

A defendant's ability to ensure the inputs into COMPAS are factually correct is insufficient to verify the algorithm's accuracy. The *Loomis* court disagrees, emphasizing that access to the inputs allowed Loomis to have some opportunity to "refute, supplement, or explain" the information provided.[98] But mere verification of the inputs will not reveal improper reasoning within the algorithm. For example, it is logically improper to draw a causal inference about a specified individual derived solely from a group's characteristics.[99] A cursory look over whether or not the inputs are factually correct would not allow an observer to test whether or not COMPAS is drawing (or even can draw) appropriate individual conclusions from the group data it uses. Thus, the inputs and outputs of the algorithm are only a small piece of the puzzle.

Due process jurisprudence is therefore not robust enough to allow for an inaccurate information claim against COMPAS. This is because of: (1) the uncertainty as to whether or not COMPAS's conclusions are inaccurate, and (2) the impossibility of fully testing said accuracy due to the algorithm's proprietary nature. This problem is apparent when the *Loomis* court struggles to reconcile a multitude of studies that come to differing conclusions regarding COMPAS's accuracy.[100] As a further indication of the Due Process Clause's inadequacy, the *Loomis* court sidesteps the accuracy question and ultimately dictates that "jurisdictions that utilize risk assessment tools must ensure they have the capacity for maintaining those tools and monitoring their continued accuracy."[101]

Defendants' right to an individualized sentence may provide another due process challenge to using COMPAS because risk scores are based on group statistical data, not data tied discretely to the individual. If courts take this requirement seriously, then arguably they cannot use COMPAS. One way to sidestep the issue, however, is to require—as the *Loomis* court does—that courts do not use COMPAS as the determinative factor in any sentence.[102] I could find no court in my research that has invalidated using COMPAS on the grounds that such use violates a defendant's right to have an individualized sentence. If a COMPAS score is only one of many factors that courts consider during sentencing, each defendant retains his or her right to receive an individualized sentence.

Finally, it is arguable that COMPAS defeats a defendant's right to understand and confront the evidence against her. Under basic Rule of Law principles, laws are to be "publicized[] and broadly understood."[103] Algorithmic risk assessments

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

99. Critics of this point argue that group data is used in many contemporary fields of prediction including insurance. The problem with this critique, however, is that the goal of insurance companies is to predict what proportion of a group will suffer a specified outcome, not to determine whether that outcome will befall a specific individual. For a discussion on this debate, please see NATHAN JAMES, CONG. RESEARCH SERV., R44087, RISK NEEDS ASSESSMENT IN THE CRIMINAL JUSTICE SYSTEM 8–11 (2018).

100. *Loomis*, 881 N.W.2d at 762–63.

101. *Id*. at 763.

102. *Id*. at 769.

103. American Bar Ass'n, *supra* note 76.

are neither of those things. Their proprietary nature keeps their contents from defendants. The assessments are not understood by the judges who apply them, let alone the public at large. In the words of a New York Times article: "no one knows exactly how COMPAS works."[104] Algorithmic risk assessments are akin to "an anonymous expert [that the defendant] cannot cross-examine."[105] These concerns are exacerbated due to the weight people give to technology in our society today.[106] Judges may be "likely to assume that quantitative methods are superior to ordinary verbal reasoning."[107] This is a form of automation bias.[108] Automation bias can easily change a technological suggestion into a "final, authoritative decision."[109] The technology, then, can work to anchor decisions in technological certainty that is both improper and inappropriate.[110]

The problem with this constitutional challenge, however, is that some constitutional rights are weakened during sentencing.[111] In *Williams v. New York*, the Supreme Court found it constitutionally appropriate for "the sentencing judge [to] consider . . . information . . . obtained outside the courtroom from persons whom a defendant ha[d] not been permitted to confront or cross-examine."[112]

Cases following *Williams* have placed emphasis on the reliability of the information used in sentencing when it was not brought out during the case in chief.[113] *Gardner v. Florida* walked back some of *Williams*' harsh language and expanded upon the reliability concept by requiring that a defendant at least have an opportunity to refute the evidence used against him or her.[114] The Court held that it was a due process violation to sentence someone to death "on the basis of information

104. Ellora Israni, *When an Algorithm Helps Send You to Prison*, N.Y. TIMES (Oct. 26, 2017), https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html.

105. Pasquale, *supra* note 94.

106. Katherine Freeman, Note, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in* State v. Loomis, 18 N.C. J.L. & TECH 75, 96 (2016).

107. Pasquale, *supra* note 94.

108. *Id.*

109. Danielle Citron, *(Un)Fairness of Risk Scores in Criminal Sentencing*, FORBES (July 13, 2016), https://www.forbes.com/sites/daniellecitron/2016/07/13/unfairness-of-risk-scores-in-criminal-sentencing/#7ca283b4ad21.

110. *See* Israni, *supra* note 63.

111. For information on the derogation of Fifth Amendment rights during sentencing, please see Aronson, *supra* note 95. For a comparative look at Sixth Amendment rights, please see John G. Douglass, *Confronting Death: Sixth Amendment Rights at Capital Sentencing*, 105 COLUM. L. REV. 1967, 1968 (2005), and Betterman v. Montana, 136 S. Ct. 1609, 1613 (2016). For a general discussion regarding the recognition of constitutional rights during sentencing, please see Carissa Byrne Hessick & F. Andrew Hessick, *Recognizing Constitutional Rights at Sentencing*, 99 CAL. L. REV. 47 (2011). *But see* Gardner v. Florida, 430 U.S. 349, 358 & n.9 (1977) (explaining in dicta that Due Process Clause does apply in sentencing, but hedging assertion in a footnote, saying, "[t]he fact that due process applies does not, of course, implicate the entire panoply of criminal trial procedural rights")

112. Williams v. New York, 337 U.S. 241, 245 (1949).

113. *See* United States v. Bass, 535 F.2d 110, 121 (D.C. Cir. 1976) (stressing the significance of the fact that sentencing allegations had an "indicia of reliability"); United States v. Metz, 470 F.2d 1140, 1142 (3d Cir. 1972) (focus on reliability of contested evidence).

114. *Gardner*, 430 U.S. at 362.

which he had no opportunity to deny or explain."[115] However, the fact that the *Gardner* case was a capital case strongly permeated the Court's opinion.[116] The Court distinguished *Gardner* from *Williams* by outlining the changes in public opinion regarding the death penalty that had occurred in the thirty years since the *Williams* decision.[117] Furthermore, the strengthened reliability requirement outlined in *Gardner* simply brings us back to the question of accuracy that began our due process discussion. Even with the refutation right outlined in *Gardner*, defendants would still lose a due process challenge to COMPAS because defendants do have some opportunity to "refute, supplement, or explain" the information contained in a COMPAS report.[118] Therefore, the Due Process Clause, while raising important concerns, is ultimately an inadequate tool to address the issues that arise from algorithmic risk assessments.

V. AN ADMINISTRATIVE SOLUTION: REGULATORY OVERSIGHT

In January 2018, New York City enacted the first algorithmic accountability law in the nation.[119] The law creates a task force charged with investigating "agency automated decision system[s]."[120] This means the law is directed at systems used by government agencies to assist in decision-making. The law's main goals are "fairness, accountability and transparency."[121] In order to achieve these goals, the law seeks to create a group of experts who will identify automated decision systems' disproportionate impacts.[122] It also requires that agency decisions be archived and that a system is created to make information available so that the public can meaningfully assess the systems.[123] Finally, the law will allow anyone affected by an automated decision to request an explanation for the decision and will require a path for redress for anyone harmed by a decision.[124]

Some feel that the New York law does not go far enough.[125] The law contains a large caveat: no compliance with the aforementioned procedures is required if

115. *Id.*

116. *Id.* at 357.

117. *Id.*

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

119. Introduction of Law Number 2018/048, N.Y. CITY COUNCIL, https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0 (last visited Feb. 20, 2019).

120. Local Law No. 49, N.Y. CITY COUNCIL https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0 (last visited Feb. 20, 2019).

121. *Id.* § 2.

122. *Id.*

123. *Id.* § 3(e).

124. *Id.* § 3(b).

125. *See* Zoë Bernard, *The First Bill to Examine 'Algorithmic Bias' in Government Agencies Has Just Passed in New York City*, BUSINESS INSIDER (Dec. 19, 2017), http://www.businessinsider.com/algorithmic-bias-accountability-bill-passes-in-new-york-city-2017-12?IR=T; Cynthia Rudin, *Algorithms and Justice: Scrapping the 'Black Box'*, THE CRIME REPORT (Jan. 26, 2018), https://thecrimereport.org/2018/01/26/algorithms-and-justice-scrapping-the-black-box/.

such compliance "would result in the disclosure of proprietary information."[126] Companies like Equivant can still hide behind their black curtain. Critics of the law argue that for true justice to be had, proprietary models need to be banned entirely.[127] Bernard Harcourt, a law professor at Columbia University who has studied risk assessments, thinks that the New York law is an important first step, but that the true solution to an ever-growing problem is increased transparency.[128] Most seem to agree that it is, at the very least, a step in the right direction.

Although the New York law is a good start, an effective solution should comprehensively address the problems we have identified, while remaining practical and within the framework of broader Rule of Law principles. As there exists no generalized structure for a scheme to address the algorithms' problematic nature, we can begin with a macro Rule of Law framework championed by Ronald Dworkin.[129] Following Dworkin's "law as integrity," we begin by asking what might fit best with the philosophy of our laws and attempt to maintain the "integrity of the legal system as a whole."[130] Dworkin stressed the importance of justice, fairness, and procedural due process.[131] These ideals are not novel in the legal world. Dating as far back as Aristotle, philosophers have emphasized the importance of the Rule of Law.[132] More concretely, the ABA World Justice Project proposed a working definition of the Rule of Law that includes four principles:

> (1) a system of self-government in which all persons, including the government, are accountable under the law; (2) a system based on fair, publicized, broadly understood and stable laws; (3) a fair, robust, and accessible legal process in which rights and responsibilities based in law are evenly enforced; (4) diverse, competent, and independent lawyers and judges.[133]

This administrative solution will attempt to achieve these aforementioned goals.

The simplest solution to the problem of algorithmic risk assessments is not to use them—to disallow technology from usurping the role of judges. However, it seems from the widespread adoption of risk assessment tools by states and their preliminary treatment in the courts that this is not a palatable solution. As noted by the *Loomis* court, the benefit of "additional sound information . . . [in] sentencing courts" outweighs the cost of using algorithmic risk assessments.[134] Therefore, I will outline three practicable suggestions that would improve the New York law and carry forward its strengths into future initiatives.

---

126. Local Law No. 49, *supra* note 120, § 2.

127. *See* Rudin, *supra* note 125.

128. Bernard, *supra* note 125 (quoting Professor Harcourt).

129. RONALD DWORKIN, LAW'S EMPIRE 167 (1986).

130. MIREILLE HILDEBRANDT, SMART TECHNOLOGIES AND THE END(S) OF LAW 136 (2015).

131. DWORKIN, *supra* note 129, at 167.

132. ARISTOTLE, RHETORIC bk. I (W. Rhys Roberts trans., 2009) (c. 350 B.C.E.), http://classics.mit.edu/Aristotle/rhetoric.html.

133. American Bar Ass'n, *supra* note 76.

134. State v. Loomis, 881 N.W.2d 749, 759 (Wis. 2016).

A. Follow the Loomis Requirements

A first step in the right direction would be to make the *Loomis* requirements—forbidding sole reliance on COMPAS and including a warning about COMPAS's flaws—a nation-wide directive. States and the federal government should be encouraged to adopt these restrictions and to make them mandatory as part of their own sentencing practices. First, "consideration of COMPAS is permissible; reliance on COMPAS for the sentence imposed is not permissible."[135] Courts should not use COMPAS to determine whether a person is incarcerated or the severity of a sentence.[136] Many significant due process concerns are reduced if courts only use COMPAS as a single factor in a sentencing decision and do not use COMPAS as the reason for increasing the severity of a sentence. Additionally, a list of warnings about COMPAS's flaws should accompany the risk scores when handed to a judge. These warnings should highlight that: COMPAS scores are generated from group data, studies have shown a disparate impact on minorities, and COMPAS was created to be used by the Department of Corrections, not during sentencing.[137]

We must also increase judges' understanding of algorithmic risk assessments. It is evident that some judges—the actors using these tools—do not understand how they work. As Justice Abrahamson notes in her concurring opinion in *Loomis*, the "court's lack of understanding of COMPAS was a significant problem."[138] An actor cannot appropriately apply a tool she does not fully understand. Judges need accurate and detailed information in order to appropriately tailor sentences to specific defendants in specific cases.[139] As mentioned, whether these tools actually produce accurate results is controversial.[140] Further, COMPAS scores are far from detailed: the judge receives a score between one and ten, without any additional contextual information, that is meant to correlate to recidivism risk.[141] At the very least, then, judges need to understand how these tools work so that they may make a conscious, informed decision about how much to rely on them in any given sentencing decision. Training is also a way to combat automation bias because judges who go through such training will be more critical of the suggestions the programs

---

135. *Id*. at 774.

136. *Id*. at 769.

137. *Id*. at 769–70. The warnings should also discuss the proprietary nature of COMPAS, but, because I argue the code should not be kept hidden, this warning would only need to be included if the rest of my administrative scheme is not adopted.

138. *Id*. at 774 (Abrahamson, J., concurring); *see also* Israni, *supra* note 63 (arguing that the decision in *Loomis* as well as the briefs filed for appeal "reflect fundamental misunderstandings about how an algorithm like COMPAS might work, and what safeguards would allow it to be useful in sentencing"); Cecelia Klingele, *The Promises and Perils of Evidence-Based Corrections*, 91 NOTRE DAME L. REV. 537, 576 (2015) ("Unless criminal justice system actors are made fully aware of the limits of the tools they are being asked to implement, they are likely to misuse them.").

139. CASSIA SPOHN, HOW DO JUDGES DECIDE 123 (2d ed. 2009).

140. *Loomis*, 881 N.W.2d at 762–63.

141. *Id*. at 754, 756.

produce.[142] Some of the remaining *Loomis* requirements have enlightened the discussion of other recommendations expanded upon below.

B. Black Box Testing and Full Accountability

The algorithms and their creators must be held accountable for mistakes the algorithms make. Algorithms are created by people. The suggestions algorithms make are necessarily a product of decisions about design, optimization choices, and what data is chosen to train the program.[143] An algorithm cannot be held to codes of conduct or standards of professionalism, but its creators can be. Furthermore, "[t]he justice system must keep up with the research and continuously assess the use of these tools."[144] This point is imperative if we are to accept the reality of technology. Technology is constantly evolving, especially when it comes to machine learning algorithms. The New York law proposes developing a procedure for testing the impact of algorithms on certain minority groups and allowing redress for any person specifically harmed by an algorithm found to be discriminatory.[145] I suggest the development of a more comprehensive, far-reaching scheme. We need both internal checks by the companies who create these algorithms as well as external review to create a system of regulatory oversight and testing that will hold the algorithm and its creators accountable.

Nicholas Diakopoulos and Sorelle Friedler make some salient suggestions about how to increase algorithmic accountability internally.[146] Diakopoulos and Friedler first stress that someone needs to take responsibility for listening to public comment and criticism concerning the algorithms.[147] This person's position must come with both the authority and the resources to actually implement change.[148] Additionally, the person must be known and available to the public so that people know who they can contact with their concerns.[149] This person need not assume legal liability, but merely practical responsibility.[150] Furthermore, this authority figure or another set of responsible persons needs to continuously "indentif[y], log[], and benchmark[]" sources of error.[151] This sort of internal check is a necessary step in both acknowledging and understanding the flaws of the algorithms and then using this information to make appropriate adjustments.

Externally, states and the federal government should take on the responsibility to test and re-test the validity and accuracy of risk assessment algorithms on a

142. *See* Rudin, *supra* note 125.

143. Nicholas Diakopoulos & Sorelle Friedler, *How to Hold Algorithms Accountable*, MIT TECH. REVIEW (Nov. 17, 2016), https://www.technologyreview.com/s/602933/how-to-hold-algorithms-accountable/.

144. *Loomis*, 881 N.W.2d at 753.

145. *See* Local Law No. 49, *supra* note 120.

146. Diakopoulos & Friedler, *supra* note 143.

147. *Id.*

148. *Id.*

149. *Id.*

150. *Id.*

151. *Id.*

regular basis. As Danielle Citron argues, "procedural regularity" is the best way to address the dynamism of algorithms.[152] Like the New York law, Citron suggests the creation of an expert agency.[153] This can be done at both the state and federal level. Said agency should routinely audit the algorithms. It should be the job of the agency to ferret out discrimination violations by continuously running "expected and unexpected hypothetical scenarios."[154] Citron argues that even the presence of these audits would be beneficial to society.[155] The existence of the audits would likely induce companies to take more precautions and be more careful when crafting and training their algorithms so as to avoid liability.[156] In order to run proper audits, it would prove easier for these experts to have full access to the trade secret protected portions of the algorithms.[157] I will address the possibility of partial disclosure to experts further in my discussion of transparency below.

Finally, Diakopoulos and Friedler encourage that the results produced by the algorithm be accessible to the defendant whom they affect.[158] Similarly, the New York law allows an affected person to "request and receive an explanation of [the algorithm's decision] and the basis therefor."[159] In order to implement this suggestion, however, society would likely need complete transparency of the source code. Complete transparency is an option that I will expound below and which the New York law makes impossible with its proprietary information loophole.[160] Even further, however, we would need experts who could "translate" this code into an explanation that a lay person—the defendant—could understand and digest.[161] The combination of these proposals would enable society to hold the algorithms' creators accountable.

C. Transparency

Transparency is the key to understanding and addressing the flaws of algorithmic risk assessments. As an initial step towards understanding the algorithm, some states have begun to conduct validation studies of COMPAS.[162] Unfortunately, it is unknown how many states have undertaken these sorts of studies.[163] Wisconsin, as noted in *Loomis*, was one of the states that had failed to do so at the time of the

152. Citron, *supra* note 109.

153. *Compare* Local Law No. 49, *supra* note 120 *with* Citron, *supra* note 109.

154. Citron, *supra* note 109.

155. *Id.*

156. *Id.*

157. *Id.*

158. Diakopoulos & Friedler, *supra* note 143.

159. Local Law No. 49, *supra* note 120.

160. *See* Local Law No. 49, *supra* note 120.

161. Diakopoulos & Friedler, *supra* note 143. For a possible method to explain algorithmic decisions to lay people, see MARCO TULIO RIBEIRO ET AL., "WHY SHOULD I TRUST YOU?" EXPLAINING THE PREDICTIONS OF ANY CLASSIFIER (2016), http://www.kdd.org/kdd2016/papers/files/rfp0573-ribeiroA.pdf.

162. State v. Loomis, 881 N.W.2d 749, 762 (Wis. 2016).

163. For a state-by-state table on validation studies, *see Algorithms in the Criminal Justice System*, EPIC (last visited Feb. 20, 2019), https://epic.org/algorithmic-transparency/crim-justice/.

decision.[164] Additionally, conclusions vary. For example, a 2007 study out of California concluded there was insufficient evidence to determine what COMPAS actually assesses.[165] A subsequent 2010 California study, however, concluded that COMPAS was imperfect, but reliable.[166] The vacillation suggests an internal struggle concerning the endorsement of the tool and how to effectively study it. What states are doing is therefore insufficient. A comprehensive test of risk assessment algorithms cannot be completed without full access to the proprietary source code.

Therefore, there needs to be increased transparency surrounding the algorithms. The algorithms' opacity prevents us from truly grasping the underlying flaws in the programs. A lack of information precludes discourse. By housing the source code in a single entity with a profit driven motive, we are precluding ourselves from having a multitude of experts in the field test and improve the algorithm so that it may better serve its purpose: to appropriately and fairly aid in sentencing decisions. Therefore, we must work to break down the trade secret shield Equivant and other companies hide behind. The following suggestions would work to increase transparency and therefore address the most glaring weakness in the New York law: the exception for proprietary information.[167]

One suggestion for addressing the trade secret problem is "qualified transparency."[168] This method would employ a group of experts to assess the "quality, validity, and reliability" of the trade secret protected source code.[169] A step beyond state validation studies, this method would allow a group of carefully chosen, qualified experts to test the algorithms from a standpoint of full information. As this would be a closed group, however, the companies would still be able to protect their code from the general public and maintain their competitive advantage.

The second option is taking this role out of the hands of the private sector entirely and moving it to the public sector. Though intellectual property law protects companies from having to disclose their source code, states and the federal government are not required to use their code in the first place. Cynthia Rudin suggests a viable alternative: use transparent algorithms derived from public data and public source code.[170] Rudin argues these fully transparent algorithms are beneficial not only because they allow a defendant to see exactly why she received a certain risk score, but also because they are free and would save taxpayers money.[171] Rudin designed and presented an example of such a transparent code in a recent article.[172]

---

164. *Loomis*, 881 N.W.2d at 762.
165. *Id*. at 762–63.
166. *Id*. at 763.
167. *See* Local Law No. 49, *supra* note 120.
168. Pasquale, *supra* note 94.
169. *Id*.
170. Rudin, *supra* note 125.
171. *Id*.
172. ELAINE ANGELINO, NICHOLAS LARUS-STONE, DANIEL ALABI, MARGO SELTZER & CYNTHIA RUDIN, LEARNING CERTIFIABLY OPTIMAL RULE LISTS FOR CATEGORICAL DATA (2017), https://arxiv.org/pdf/1704.01701.pdf.

Moreover, if algorithms can usurp part of a judge's decision-making power, then their creators must be held to the same "high ethical and constitutional standards" that judges are and have the same responsibility that judges have.[173] The algorithm creators' responsibility has been compared to that of judges to explain their decisions in written, protracted, published opinions.[174] Therefore, those who design and train the algorithms should also be required to provide similar written, public explanations about the decisions their algorithms make. If algorithms want to act like judges, then they—and their creators—should be treated like judges. Being part of the criminal justice system comes with a lengthy list of limitations and responsibilities. Algorithm creators should not be allowed to hide behind the actions of the tools they have designed, sold, and deployed.

A final suggestion is to have the algorithms themselves "explain" their reasoning for decisions.[175] This suggestion would apply only to machine learning algorithms. A group of researchers at MIT designed a training process for algorithms that allows the algorithm to match a rationale with its conclusion.[176] This research was done on medical diagnoses, but it can be abstracted to risk assessment algorithms. There is value in an algorithm "explaining itself," especially if its creators will not or cannot do so.

CONCLUSION

We cannot allow the use of algorithmic risk assessments to go unchecked. Concerns over "sentencing by computers" date back to 1970s debates over indeterminate sentencing.[177] We have now allowed technology to permeate our judicial system. This is a positive thing: it increases efficiency and cost-effectiveness. A tool like COMPAS, however, goes too far: it infiltrates our system's decision-making function. This usurpation should concern us all, especially when these algorithms are created and controlled by for-profit, private companies who refuse to reveal their source code to the public.

Use of algorithmic risk assessments in sentencing runs afoul of many general Rule of Law principles because their outcomes are tainted by discrimination and the algorithms are not and cannot be held accountable for their mistakes. The United States Constitution is not the appropriate tool to address the algorithmic risk assessment problem because the opaque, technological aspects of the algorithms preclude defendants from proving the necessary elements of their constitutional challenges. Instead, we should work together as a country to create a

173. Israni, *supra* note 63.

174. *Id*.

175. Matt Burgess, *Holding AI to Account: Will Algorithms Ever Be Free from Bias if They're Created by Humans*, WIRED (Jan. 11, 2016), http://www.wired.co.uk/article/creating-transparent-ai-algorithms-machine-learning.

176. Larry Hardesty, *Making Computers Explain Themselves*, MIT NEWS (Oct. 27, 2016), https://news.mit.edu/2016/making-computers-explain-themselves-machine-learning-1028.

177. RICHARD G. SINGER, JUST DESERTS: SENTENCING BASED ON EQUITY AND DESERT 49 (1979).

comprehensive administrative scheme to solve the problem. Such a solution could ensure a more unified, limited use of these tools. This solution should importantly require increased accountability and transparency.

As a final note, we must be prepared for what we find once we do look inside the black box. It is likely that the algorithms are not facially discriminatory and that the programmers have not created unconstitutional code. It is also possible, however, that these algorithms do in fact have a discriminatory purpose or are so error-ridden as to render the scores arbitrary. If we find such egregious flaws inside the black box, perhaps our only option is to stop using the algorithms entirely or to create different, improved algorithms to take their place. Judgment day is coming for these algorithms and unfortunately for them, they have been classified as high-risk.