6-2019

Beyond State v Loomis: Artificial intelligence, government algorithmization and accountability

Han-wei LIU
Singapore Management University, hanweiliu@smu.edu.sg

Ching-Fu LIN

Yu-Jie CHEN

Beyond *State v. Loomis*: Artificial Intelligence, Government Algorithmization, and Accountability

HAN-WEI LIU,[*] CHING-FU LIN[**] AND YU-JIE CHEN[***]

INTERNATIONAL JOURNAL OF LAW AND INFORMATION TECHNOLOGY
(FORTHCOMING 2019)

DRAFT – DEC. 20, 2018

EMAIL COMMENTS TO HANWEI.LIU@MONASH.EDU;
CHINGFULIN@MX.NTHU.EDU.TW; YU-JIE.CHEN@NYU.EDU.

ABSTRACT

Developments in data analytics, computational power, and machine learning techniques have driven all branches of the government to outsource authority to machines in performing public functions—social welfare, law enforcement, and most importantly, courts. Complex statistical algorithms and artificial intelligence (AI) tools are being used to automate decision-making and are having a significant impact on individuals' rights and obligations. Controversies have emerged regarding the opaque nature of such schemes, the unintentional bias against and harm to underrepresented populations, and the broader legal, social, and ethical ramifications. State v. Loomis, a recent case in the United States, well demonstrates how unrestrained and unchecked outsourcing of public power to machines may undermine human rights and the rule of law. With a close examination of the case, this Article unpacks the issues of the 'legal black box' and the 'technical black box' to identify the risks posed by rampant 'algorithmization' of government functions to due process, equal protection, and transparency. We further assess some important governance proposals and suggest ways for improving the accountability of AI-facilitated decisions. As AI systems are commonly employed in consequential settings across jurisdictions, technologically-informed governance models are needed to locate optimal institutional designs that strike a balance between the benefits and costs of algorithmization.

Keywords: *State v. Loomis*; artificial intelligence (AI); algorithms; black box; human rights; rule of law; accountability

[*] Lecturer, Monash University.
[**] Associate Professor of Law, National Tsing Hua University.
[***] Post-doctoral Scholar, Institutum Iurisprudentiae, Academia Sinica, Taiwan; Affiliated Scholar, U.S.-Asia Law Institute, NYU School of Law.

Introduction

The past decade has witnessed an unprecedented rise of data analytics and algorithms in both private life and public policy. By and large, such technological expansion has been driven by increasing demands for managing the complexity of contemporary society, the availability of vast amounts of information generated by ubiquitous uses of innovative devices enabled by the Internet, and the significant growth in the semiconductor industry and hence computing powers. The United States is at the forefront of applying data-driven techniques: financial institutions capitalize on big data to assess the pattern of transactions and the risks of customers;[1] e-commerce companies leverage data-driven approaches to craft automated systems that recommend products, music, and services to users;[2] fitness trackers monitor personal health data and motivate users to engage in more activities;[3] and manufacturing and agricultural industries can also tap into such new techniques to increase efficiency and productivity.[4]

While many pursue the benefits of big data and algorithms, automated systems can have equal potential for harm. As stated in the White House Report on Big Data of 2016, 'it is a mistake to assume [these big data techniques] are objective simply because they are data-driven'.[5] The way we use data as 'inputs to an algorithm' and 'the inner workings of the algorithm itself' poses critical challenges to policymakers in promoting fairness and overcoming bias and discriminatory effects while we move towards a 'smart world'.[6]

This raises a set of analytical questions: Who invents these systems? In what social and economic contexts are such systems developed? How are algorithms designed and what data are they based on? What are their roles in various decision-making mechanisms adopted by public and private actors to

---

[1] *See eg,* E Bank, 'How Marketplace Lenders Decide If You're a Good Risk' (*Credible*, 12 February 2018), <https://www.credible.com/blog/personal-loan/marketplace-lenders-decide-good-risk/>.

[2] *See eg,* J Markman, 'Amazon Using AI, Big Data to Accelerate Profits' (Forbes, 5 June 2017) <https://www.forbes.com/sites/jonmarkman/2017/06/05/amazon-using-ai-big-data-to-accelerate-profits/#4583b58c6d55>.

[3] D Yeung and AS Cevallos, 'Using Wearable Fitness Devices to Monitor More Than Just Fitness', (*Scientific American*, 11 May 2017) <https://blogs.scientificamerican.com/observations/using-wearable-fitness-devices-to-monitor-more-than-just-fitness/>.

[4] L Columbus, 'Ten Ways Big Data Is Revolutionizing Manufacturing' (*Forbes*, 28 November 2014) <https://www.forbes.com/sites/louiscolumbus/2014/11/28/ten-ways-big-data-is-revolutionizing-manufacturing/#5884b86ce161>; T Sparapani, 'How Big Date and Tech Will Improve Agriculture, From Farm to Table' (*Forbes*, 23 March 2017) <https://www.forbes.com/sites/timsparapani/2017/03/23/how-big-data-and-tech-will-improve-agriculture-from-farm-to-table/#d67a43159891>.

[5] Executive Office of the President, Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights (2016), 6.

[6] For an overview of the development of informationalization and intelligentization in the data-driven age, *see generally* H Ning et al., 'From Internet to Smart World', (2015) 3 IEEE Access 1994.

exercise certain economic and political powers and to shape and reshape individual rights and obligations? Are we able to examine the process as well as the outcome of automated systems? If so, how? If not, should we reject these systems altogether or subject their use to certain legal and ethical parameters? How should we ensure the accountability of AI-enabled tools in public (and in some cases private) decision-making processes?[7] Many of these questions require continuing conversations informed by technological expertise and policy considerations. Such discussion is not just a technical issue. Rather, these automated decision-making systems have normative implications, for not only every aspect of social life, but also the way in which policymakers around the world design social and legal institutions in the age of AI.

This Article primarily focuses on the normative implications of using data-driven techniques (in particular, complex algorithms and artificial intelligence tools) in various government functions; the challenge it poses to due process, equal protection and transparency; and the accountability of the public sector to these important values. To do so, we investigate one of the most controversial areas that illustrate the emerging challenge—the increasing use of data analytics in the criminal justice system, which engages the core of these fundamental human rights and governance issues.

In this regard, the changing landscape of America's criminal justice system offers some critical insight into these debates. For years, the U.S. has been applying data analytics and algorithms in the decision-making process of law enforcement agencies, correction officials, and judges. This trend turns on two interrelated factors. For one, the data-driven approach appears to represent a *cost-effective* solution, aiding criminal justice officials to prioritize government resources in predicting and controlling complex individual behaviors.[8] For another, many have argued that 'evidence-based' tools aided by big data could

[7] For the present purpose, 'accountability' refers broadly to the notion that public (and in some cases, private) powers using automated decision-making should explain and justify the use of AI to those with whom they interact and should uphold important values for the benefits of individuals and society, including but not limited to due process, equal protection and transparency. On the notion of accountability in the context of AI, *see eg* Virginia Dignum, 'Responsible Autonomy', Proceeding of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), 4698 (linking 'accountability' to 'answerability', 'blameworthiness', and 'liability'); Finale Doshi-Velet et al., 'Accountability of AI under the Law: The Role of Explanation', Berkman Klein Center for Internet & Society working paper <http://nrs.harvard.edu/urn-3:HUL.InstRepos:34372584?>

[8] R Brauneis and EP Goodman, 'Algorithmic Transparency for the Smart City' (2018) 20 Yale Journal of Law and Technology 103, 114.

help remove the presence of human beings—and therefore, their inherent *biases*—from the decision-making process.[9]

Both reasons should be understood in the context of the problems that confront America's criminal justice regime today. The U.S. has experienced a record-breaking rise in the penal population with nearly 2.2 million people currently behind bars, which has quadrupled from only 500,000 in the 1980s,[10] and is nearly five times the international average.[11] The current situation reflects the shifted policies and politics at state and federal levels that nowadays rely on lengthy prison sentences for violent and drug-related crimes and repeated offenses.[12] While officials claim that a 'tough on crime' approach protects public safety by harshly punishing criminal conduct and deterring future criminal behavior,[13] the mass imprisonment problem has resulted in unwanted fiscal, social, and political consequences.[14] Escalating correction costs and the high rate of recidivism have driven policymakers to move away from heavy reliance on imprisonment and to map out new strategies of enforcement and rehabilitation.[15]

One convenient way to overhaul the system is to make it 'smarter' by drawing on data analytics and algorithms.[16] By profiling offenders based upon their risks of recidivism, these innovative technologies claim to better inform decisions of criminal justice officials and judges. While these so-called 'evidence-based' risk assessment tools[17] were initially implemented to aid

---

[9] J Salman and E Le Coz, 'Race and Politics Influence Judicial Decisions. But Florida's Bench is a World of Contradictions' Herald Tribute, 12 December 2016) <http://projects.heraldtribune.com/bias/politics/>.
[10] D Hudson and President BH Obama, 'Our Criminal Justice System Isn't as Smart as It Should Be' <https://obamawhitehouse.archives.gov/blog/2015/07/15/president-obama-our-criminal-justice-system-isnt-smart-it-should-be>.
[11] *See* M O'Hear, 'Wisconsin Sentencing in the Walker Era: Mass Incarnation as the New Normal' (2017) 30 (2) Federal Sentencing Report 125 (describing the prison population increase in Wisconsin since the 1970s). R Walmsley, 'World Prison Population List' (Prison Studies,11th edn, 2015) <https://prisonstudies.org/sites/default/files/resources/downloads/world_prison_population_list_11th_edition_0.pdf>; J Conyers, 'The Incarceration Explosion' (2013) 31 Yale Law and Policy Review 377, 377-78.
[12] National Research Council, 'The Growth of Incarceration in the United States: Exploring the Causes and Consequences' (2014), 70-71.
[13] RK Warren, 'Evidence-Based Practices and State Sentencing Policy: Ten Policy Initiatives to Reduce Recidivism' (2007) 82 Indiana Law Journal 1307, 1308.
[14] D Garland, an American socialist and legal scholar, popularized the term 'mass imprisonment.' *See* D Garland, 'Introduction: The Meaning of Mass Imprisonment' in D Garland (ed), *Mass Imprisonment: Social Causes and Consequences* (SAGE 2001) 1-4.
[15] Warren (n 12) 1307, 1308.
[16] AM Barry-Jester et al., 'The New Science of Sentencing' (*Marshall Project*, 4 August 2015) <https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing>.
[17] ibid. The term 'evidence-based assessment tools' or 'risk assessments' generally refer to techniques that 'try to predict recidivism--repeat offending or breaking the rules of probation or parole--using statistical probabilities based on factors such as age, employment history and prior criminal record'. Various federal and state governments have been actively championing the use of evidence-based risk assessment tools in their criminal justice systems. For an overview of evidence-based assessments, *see* 'Algorithms in the

certain post-conviction decisions such as determination of parole and of the types and conditions of supervision, they have now penetrated into many other phases—from policing,[18] pre-trial bail,[19] to post-trial sentencing.[20] As things stand, over 60 automated systems have been adopted in various stages throughout the American criminal justice system.[21] Prominent examples are PredPol,[22] Level of Service Inventory (LSI)—now rebranded as Level of Service Inventory-Revised (LSI-R),[23] Public Safety Assessment (PSA),[24] Post Conviction Risk Assessment (PCRA),[25] and Correctional Offender Management Profiling for Alternative Sanctions (COMPAS),[26] among others.

As promising as these systems are, potential bias and discrimination embedded in their data sources, the algorithmic 'black-box' problem, and the misguided interpretations and inferences resulting from data analytics, have quickly engendered enormous debates among policymakers, practitioners, and academics.[27] The ramifications have been manifested in a recent case, *State v.*

Criminal Justice System' (Electronic Privacy Information Center) <https://epic.org/algorithmic-transparency/crim-justice/>.

[18] *See eg,* AG Ferguson, 'Policing Predictive Policing (2017) 94 Washington University Law Review 1109.

[19] *See eg,* S Baradaran and FL Mcintyre, 'Predicting Violence' (2012) 90 Texas Law Review 497.

[20] *See eg,* DS Sidhu, 'Moneyball Sentencing' (2015) 56 Boston College Law Review 671. For an overview of risk assessments in the criminal justice system, *see generally* Pew Center on the States, Risk/Needs Assessment 101: Science Reveals New Tools to Help Manage Offenders (2011), 2.

[21] ibid.

[22] PredPol grew up out of a collaborative project between professors at UCLA, Santa Clara University, as well as the Los Angeles Police Department by predicting the locations of future crimes, and then directing the deployment of officers. *About PredPol,* PredPol, <http://www.predpol.com/about/> accessed 30 August 2018. For a recount, *see* AG Ferguson, 'Predictive Prosecution' (2016) 51 Wake Forest Law Review 705, 709-11.

[23] LSI-R is an actuarial classification system to evaluate individuals' risks to inform correctional decisions of custody, supervision, and service needs. *See* B Vose et al., 'The Empirical Status of the Level of Service Inventory' (2008) 72-DEC Federal Probation 22; SM Manchak et al., 'Utility of the Revised Level of Service Inventory (LSI-R) in Predicting Recidivism after Long-term Incarceration', (2008) 32 Law and Human Behavior 488.

[24] PSA is a risk assessment tool that helps improve the predictive accuracy in setting the bail at the pretrial stage. *See* 'Developing A National Model for Pretrial Risk Assessment' (Laura and John Arnold Foundation, 2013), 4-5 <http://www.arnoldfoundation.org/wp-content/uploads/2014/02/LJAF-research-summary_PSA-Court_4_1.pdf>.

[25] PCRA is a scientifically based instrument created by the Administrative Office of the U.S. Courts (AO) to improve the effectiveness and efficiency of post-conviction supervision. The PCRA enables officers to focus on the supervision of possible reoffenders who are at the greatest risk of failing. *See* Administrative Office of the United States Courts, Office of Probation and Pretrial Services, An Overview of the Post Conviction Risk Assessment (September 2011), at U.S. Courts, Statistics & Reports, Post Conviction Risk Assessment (2011), <http:// www.uscourts.gov/statistics-reports/publication/post- conviction-risk-assessment >.

[26] JL Skeem and JE Louden, 'Assessment of Evidence on the Quality of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)' (California Dept of Corrections and Rehabilitation, 3 2007) <https://www.cdcr.ca.gov/adult_research_branch/Research_Documents/COMPAS_Skeem_EnoLouden_Dec_2007.pdf>.

[27] *See eg,* F Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (HUP 2015).

Loomis, in which the Wisconsin Supreme Court upheld a lower court's sentencing decision informed by a COMPAS risk assessment report and rejected the defendant's appeal on the grounds of the right to due process.

Using *State v. Loomis* as a vantage point, this Article makes the following contributions. First, we offer a critical appraisal of the *State v. Loomis* decision by discussing the court's failing to understand the workings of the risk assessment tool at its disposal as well as the negative impact on the fundamental rights of the defense (Section 1). Second, and more significantly, we generalize our analysis and discuss broader implications of a data-driven approach in legal processes in the United States and beyond. As illustrated in Section 2, quite a few countries have implemented or are preparing for the AI-enabled future by outsourcing, in one way or another, decision-making processes to automated systems. The sweeping ramifications merit further considerations. Section 3 concludes this Article.

1. *State v. Loomis*: A Revisit

1.1. Facts and Decision

In 2013, Eric Loomis was charged by Wisconsin with five criminal counts in relation to a drive-by shooting in La Crosse, including: (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; and (5) possession of a short-barreled shotgun or rifle. The defendant Loomis admitted that he drove the car in dispute, but denied his involvement in the shooting.[28] Loomis waived his right to trial and entered a guilty plea to two of the lesser charges; the plea agreement also stated that the rest of the counts would be 'dismissed but read in' for sentencing.[29]

Following the plea, the circuit court ordered a pre-sentencing investigation report (PSI) which included a COMPAS risk assessment score to facilitate the sentencing process. The risk assessment—drawn on information built upon the interview and the criminal file of the defendant—generated scores that displayed by way of a bar chart representing different types of risks: pretrial recidivism, general recidivism, and violent recidivism.[30] These three bar charts all indicated that Loomis was as an individual with a 'high risk of recidivism' who was 'at high risk to the community'.[31] Based on the risk assessment and

[28] *State v Loomis* 881 N.W.2d 749 (Wis. 2016) 754 (US).
[29] ibid.
[30] ibid.
[31] ibid 755.

other relevant factors, including the read-in charges, the circuit court ruled out probation and sentenced Loomis within the maximum of the two charges for which he entered into a plea: six-year imprisonment and five-year extended supervision.[32]

In response, Loomis lodged a motion for post-conviction relief, requesting a new sentencing proceeding. The defense argued that the court's reference to the risk assessment in sentencing Loomis violated his constitutional right to due process and incorrectly considered the read-in charges. More specifically, on the due process issue—the focus of this Article—Loomis's arguments proceeded along three lines. First, the decision violated Loomis's due process right to be sentenced based on accurate information since much of the information used by COMPAS was taken as a trade secret that barred the defendant from assessing the accuracy.[33] Second, Loomis challenged the court saying that the decision infringed on his right to an individualized sentence proceeding as COMPAS was designed to evaluate group data.[34] Third, the score involved improper use of a gender-based assessment and therefore violated his right to due process.[35]

Upon appeal, however, the Wisconsin Supreme Court rejected all these claims. On the first argument, while the Court acknowledged that risk scores failed to explain how COMPAS employed data to generate the results, it nevertheless dismissed Loomis' argument on the basis that risk assessments were largely based on available information such as criminal history and the answers to a list of questions provided by the defendant. To this effect, the Court opined that Loomis had the opportunity to verify and challenge his risk scores.[36]

On the second claim, the Court began by reiterating what the case *State v. Gallion* explained in 2004, that the notion of individualized sentencing is 'a cornerstone to Wisconsin's criminal justice jurisprudence'.[37] Yet, according to the Court, the risk assessment formed only part of the basis on which the sentencing decision relied; there were other factors considered by the circuit court. Referring to the *Gallion* decision that underlined the need for 'more complete information upfront at the time of sentencing', the Court saw

[32] ibid 756.
[33] ibid 757, 760-64.
[34] ibid 757, 764-65.
[35] ibid 757, 765-67.
[36] ibid 761.
[37] ibid 764. *See also State v Gallion* 678 N.W.2d. 197 (Wis. 2004) 209 (US).

COMPAS as having the potential to address such an enhanced need and thereby rejected this second claim.[38]

In relation to the gender-based issue, the Court, too, sided with the lower court. It found that Loomis failed to discharge his burden of proof showing that the sentencing decision was based on gender. Citing the expert opinion, which warned about the risk of failing to take gender into account when predicting recidivism, the Court observed instead that 'if the inclusion of gender promotes accuracy, it serves the interests of institutions and defendants, rather than a discriminatory purpose'.[39] The Court therefore concluded that COMPAS's use of gender promotes accuracy that 'ultimately inures to the benefit of the justice system including defendants'.[40]

Although the Wisconsin Supreme Court rejected Loomis's claims, it nevertheless set forth some limitations (which are, in our view, symbolic ones of limited value) on using COMPAS in the sentencing process. According to the Court, while the risk scores can help assess a defendant's risk to public safety and inform the decision-makers of the recidivism and corresponding risk management, they are only part of the relevant factors and 'cannot be determinative'.[41] Citing the report of the Department of Corrections and especially the 'Guiding Principle' of the National Center for State Courts, the *Loomis* Court admonished that:

> Risk and need assessment information should be used in the sentencing decision to inform public safety considerations related to offenders' risk reduction and management. It should not be used as an aggravating or mitigating factor in determining the severity of an offender's sanction.[42]

The Court so determined because it acknowledged that the primary focus of COMPAS was on recidivism, part of the overall goal of a sentence.[43] Other goals such as deterrence, rehabilitation, retribution, and segregation were also crucial rationales behind the criminal punishment; such multifaceted functions therefore made this risk assessment tool 'a poor fit' to determine the length and severity of a sentence.[44]

---

[38] *State v Loomis* (n 27) 765.
[39] ibid 766.
[40] ibid 767.
[41] *State v Loomis* (n 27) 767.
[42] ibid 68.
[43] ibid.
[44] ibid 769.

The Court hence instructed that lower courts 'must explain the factors in addition to a COMPAS risk assessment that independently support the sentence imposed'.[45] Chief Justice Patience Drake Roggensack in his concurring opinion attempted to clarify the role of COMPAS in the sentencing process by qualifying its use as follows:

> [H]owever, I write to clarify that while our holding today permits a sentencing court to *consider* COMPAS, we do not conclude that a sentencing court may *rely on* COMPAS for the sentence it imposes… the majority opinion interchangeably employs *consider* and *rely* when discussing a sentencing court's obligations and the COMPAS risk assessment tool, our decision could be mistakenly be read as permitting reliance on COMPAS.[46] (emphasis added)

According to the Court's opinion, COMPAS can be at best only one of the factors taken into account in a sentencing decision. What remains puzzling, however, is how to distinguish the meanings of 'rely on' and 'consider' in Loomis, or to put it plainly, how significant the COMPAS predictions should be among the many considerations. Writing separately, Justice Shirley S. Abrahamson clarified the term 'consider' by instructing judges to 'set forth on the record a meaningful process of reasoning addressing the relevance, strengths, and weaknesses of the risk assessment tool' when they consider COMPAS or similar tools in a sentencing process.[47]

To ameliorate the concerns about the lack of transparency due to the proprietary nature of COMPAS, the *Loomis* Court introduced a warning label by requiring 'written advertisement of its limitation' as follows to be included in the pre-sentencing investigation report:

> • The proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are determined;
>
> • Because COMPAS risk assessment scores are based on group data, they are able to identify groups of high-risk offenders—not a particular high-risk individual;

[45] ibid.
[46] ibid 772.
[47] ibid 774.

• Some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having a higher risk of recidivism;

• A COMPAS risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed. Risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations;

• COMPAS was not developed for use at sentencing, but was intended for use by the Department of Corrections in making determinations regarding treatment, supervision, and parole.

To challenge the Court's decision, Loomis filed a writ of certiorari to the United States Supreme Court, which dismissed his petition in June 2017.

1.2. Where the *Loomis* Decision Falls Short

Loomis is a landmark case on the legality of using risk assessment tools in a sentencing court.[48] Unsurprisingly, it quickly hit the headlines and stimulated enormous debate among scholars, practitioners, and policymakers.[49] In what follows, we sharpen the focus on the failing of the court to understand potential perils of using algorithms and the harm it does to the fundamental safeguards of due process. In doing so, we take into consideration the state of the field and the trend before and after the *Loomis* decision of increasing judicial use of data analytics. This critique also helps us to look beyond the U.S context by contrasting how other jurisdictions address similar challenges in Section 2.

1.2.1. A Window-Dressing 'Warning Label'

[48] *Malenchik v State* is yet another earlier case which involved an 'evidence-based' risk assessment. The Indiana Supreme Court in this case held that the tools like the LSI-R 'can be significant sources of valuable information for judicial consideration in deciding whether to suspend all or part of a sentence, how to design a probation program for the offender, whether to assign an offender to alternative treatment facilities or programs, and other such corollary sentencing matters' and they did not 'substitute for the judicial function of determining the length of sentence appropriate for each offender'. *Malenchik v State* 928 N.E.2d 564 (Ind. 2010) at 573.

[49] *See eg,* F Pasquale, 'Secret Algorithms Threaten the Rule of Law' (*MIT Technology Review* 1 July 2017) <https://www.technologyreview.com/s/608011/secret-algorithms-threaten-the-rule-of-law/> (It noted that 'a secret risk assessment algorithm that offers a damning score is analogous to evidence offered by an anonymous expert, whom one cannot cross-examine'); M Smith, 'In Wisconsin, a Backlash Against Using Data to Foretell Defendants' Futures' (*The New York Times* 22 June 2016) <https://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html>; A Liptak, 'Sent to Prison by a Software Program's Secret Algorithms' (*New York Times 1 May 2017*) <https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html>. *See generally* N Ram, 'Innovating Criminal Justice' (2018) 112 *Northwestern University Law Review* 659, 683-91.

Admittedly, the Wisconsin Supreme Court did not appear to simply rubber stamp the use of COMPAS or similar systems, and instead, drew the lower court's attention to the limitations in detail. In our view, however, these cautions are questionable and insufficient for several reasons.

First, the *Loomis* Court seemed self-contradictory in explaining the role of COMPAS in the context of sentencing. If, as the Court held, COMPAS was a 'poor fit' for a sentencing decision, should it not direct the lower courts to remove the risk scores from the PSI altogether? Why did the Court accept the COMPAS risk assessment as one of the sentencing factors, but at the same time downplay its significance? The Loomis Court offered little explanation here.

More importantly, in the view of the *Loomis* Court, as long as written cautionary notes about the risks of COMPAS are included in the PSI report and the judges consider other factors at the same time, the use of COMPAS should not be challenged. Yet, the Court's 'warning label' fails to put meaningful constraints in place, leaving the door wide open for judges to be heavily influenced by the risk assessment. In reality, it can be unrealistic to expect a trial judge, after reading the risk scores attached to the PSI, to exercise discretion without any predetermined views of, or even bias against, the defendant. It is hard to think of, as Professor Sonja Starr remarked, a situation where a '"high risk" label will not result in a longer sentence'.[50]

That is, the Court failed to consider the psychological 'anchoring effect' for courts using scientific and technological tools, as numerous studies have demonstrated how judges (and human individuals) are submissive to computer-generated numbers and results that may further frame and condition the view of judges.[51] Mere written warnings do not seem to be able to adequately inform judges as effective gatekeepers, especially when they may not be sufficiently equipped with expertise as to understand the workings of such tools.[52] Yet the Court, insensitive to this reality itself, accepted the claim of the circuit court in the post-conviction proceeding that 'it would have imposed the exact same sentence' even without COMPAS risk scores.[53]

---

[50] D Kehl et al., 'Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing' Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School. <https://dash.harvard.edu/handle/1/33746041 >.
[51] Note, 'State v Loomis: Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing' (2017) 130 Harvard Law Review 1530.
[52] *State v Loomis* (n 27) (Abrahamson J concurring) (warned that the Court's 'lack of understanding' is a 'significant problem' to understanding risk assessment systems).
[53] ibid 771.

1.2.2. Means to Challenge Algorithms Deprived

What is equally problematic is that the Loomis Court pretended to address the issues while leaving aside critical challenges facing courts in the age of big data and AI. The fundamental issue of due process is at stake.

The crux of the present case is, in our view, how to ensure that the defendant has the means to challenge information before the court, namely, the COMPAS risk score. This question is of both factual and normative significance. As a matter of fact, the risk assessment attached to the PSI may, as argued above, shape the judges' predetermined views about the defendants even though it comes with a 'warning label'.[54] Whether the defendant can meaningfully challenge information before the court is a matter of law. The U.S. Supreme Court has recognized, as in *Townsend v. Burke*, that the due process right to a fair sentencing procedure included 'the right to be sentenced on the basis of accurate information'. To this end, the court in *State v Skaff* went on to underscore that a defendant must be given 'means' to investigate and verify the information:

> Skaff does not complain that the trial court relied on inaccurate information; he complains of the denial of means to ascertain whether there was any misinformation. Until Skaff reads his PSI, its correctness is unknown to anyone. If the PSI contains errors, given the wide sentencing discretion possessed by the trial court, a possibility exists that such errors skewed the sentence.[55]

How can the *Loomis* Court satisfy itself with the above tests is far from clear. Although Northpointe, the private company that developed COMPAS, maintained that COMPAS risk scales 'generally fall into the moderate to good range of predictive accuracy' as measured under the receiver operating characteristic curve (AUC)—approximately 0.7, there have been studies showing the opposite.[56] For instance, a non-profit organization, ProPublica, conducted an in-depth survey, observing the conduct of over 7,000 defendants in Broward County, Florida, who also received COMPAS risk scores, to examine how many of them committed a new crime over the two years following their previous arrest. Their studies showed that COMPAS was 'remarkably

[54] Computerized results may serve as a psychological anchor for judges. Note (n 50) 1536.
[55] *State v Skaff* 152 Wis.2d 48, 53, 447 N.W.2d 84 (Ct.App.1989) 58 (US).
[56] 'Practitioner's Guide to COMPAS Core' (*Northpointe, 2015, 13-14*) <http://www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-_031915.pdf> .

unreliable' and 'only 20 percent of the people predicted to commit violent crimes actually went on to do so'.[57] If a full range of crimes—including misdemeanor—were considered, ProPublica found that, 'the algorithm was somewhat more accurate than a coin flip'. In the aftermath of the Loomis decision, another study by computer scientists from Dartmouth also challenged COMPAS, asserting that 'the widely used commercial risk assessment software COMPAS is no more accurate or fair than predictions made by people with little or no criminal justice expertise', which cast 'significant doubt on the entire effort of algorithmic recidivism prediction'.[58] These critiques underscore the problem that COMPAS may mistakenly classify a lower-risk defendant as a higher-risk one—or the other way around. Although the tolerance level of error is debatable, the mixed results have already led us to reflect upon the cornerstone of the American criminal judicial process, as warned by Justice Harlan in the 1970s: 'it is far worse to convict an innocent man than to let a guilty man go free'.[59]

The defendant's right to interrogate the algorithm, therefore, is crucial, especially against the backdrop of potential discrimination hidden in the data. COMPAS' model, for example, has been criticized for working against African Americans. The ProPublica report finds that its algorithms tends to 'falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants, while at the same time mislabeling the whites as low risk more often than the blacks.[60] While COMPAS may not explicitly incorporate race as a factor, it can arguably be programmed in a way that highly correlates to a defendant's ethnic background, thus raising constitutional concerns about due process and equal protection. Although using race as a variable would be, as the U.S. Supreme Court held in *Johnson v. California*, subject to strict scrutiny and run afoul of the spirit of a matrix of equal opportunity laws in other contexts,[61] proxies such as education, vocation, family stability, and socioeconomic status can perfectly aid programmers to work around the unconstitutionality problem. The hidden racism, together with the gender issue raised by Loomis may well set off an alarm for the criminal justice

[57] J Angwin and J Larson, 'Machine Bias' (*Propublica,* 23 May 2016)<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing >.
[58] J Dressel and H Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism' (2018) 4(1) Science Advances.
[59] In re Winship 397 U.S. 358, 371-72. (1970) (Harlan, J concurring).
[60] Angwin and Larson (n 56).
[61] *Johnson v California* 543 U.S. 499, 509 (2005) (US). Federal legislation like the Equal Credit Opportunity Act and Title VII of the Civil Rights Act of 1964 in the U.S also ban discrimination based on protected characteristics, including ethnic background.

system, as reiterated by a recent U.S. Supreme Court case, *Buck v. Davis*: 'our criminal law punishes people for what they do, not who they are'.[62]

Non-transparency is yet another flaw. While COMPAS algorithms drew on public data and information provided by Loomis, it did not explain the breakdown of each variable, relevant weighting, and their correlation. For instance, Northpointe in its Practitioner Guide explains the way it determines the Violent Recidivism Risk as follows: The Violent Recidivism Risk Scale is constructed based on characteristics including: 'History of Noncompliance Scale', 'Vocational Education Scale', 'Current age', 'Age-at-first-arrest', and 'History of Violence Scale' and each item is multiplied by a given weight expressed in 'w', without disclosing what the 'w' actually is.[63] While part of the COMPAS input, as the *Loomis* Court pointed out, drew upon public information and the 137-question survey provided by the defendant, there was no way Loomis could know how Northpointe determined the weight and why, which is taken as a proprietary instrument and trade secret. Simply put, Loomis may have seen the input and output, but had no idea of their relationship. The defendant did not know how the algorithm was designed or how the decision was made; nor could he challenge it. Unlike expert witnesses who must be subject to cross-examination for their opinions and methodology, the software COMPAS cannot be called into court for questioning, and yet it is as powerful—if not more so—than an expert witness in influencing the court's decision.

The *Loomis* Court acknowledged the lack of transparency because of the proprietary nature, but what it did was just to flag this issue by virtue of a warning label and then took its risk scores as a given without examining how the outputs were derived anyway. As a corollary, Loomis had no means to meaningfully challenge the risk scores against him.

2. Beyond Loomis: Taking Stock and Looking Forward

The *Loomis* Court, while acknowledging the lack of transparency and potential biases accompanying the use of COMPAS in the sentencing procedure, appeared reluctant to tackle the core problems of applying AI or algorithmic tools in the court. Rather than solving the challenges posed by technology to due process, equal protection, and transparency, the Loomis decision exacerbates them by brushing away serious concerns by slapping on a warning label. The

[62] *Buck v Davis*, 137 S. Ct. 759, 766 (2017) (US).
[63] The Violent Recidivism Risk Score is calculated using the following equation: (age∗−w)+(age-at-first-arrest∗−w)+(history of violence∗w) + (vocation education∗w) + (history of noncompliance∗w). Practitioner's Guide to COMPAS Core (n 55) 28-29.

Court in effect outsourced its decision-making, at least to some extent, to algorithms that are questionable in quality and insensitive to the fundamental norms in the U.S. legal system, consequently undermining its public accountability.

The challenge is not unique to the *Loomis* Court, and it is worth noting that other approaches exist. Take Australian courts, which seem more cautious about applying risk assessment tools in criminal proceedings. In *Director of Public Prosecutions for Western Australia v. Mangolamara*, for instance, the Supreme Court of Western Australia pointed out 'the research data and methods underlying the assessment tools are assumed to be correct but this has not been established by the evidence'.[64] 'It has not been made clear', as the Court went on to note, 'whether the context for which the categories of assessment reflected in the relevant texts or manuals were devised is that of treatment and intervention or that of sentencing'.[65] In contrast to the Loomis decision, the Supreme Court of Western Australia underscored the fact that the tools in dispute 'were not devised for and do not necessarily take account of the social circumstances of indigenous Australians in remote communities,' thus concluding to have 'grave reservations as to whether a person of the respondent's background can be easily fitted within the categories of appraisal presently allowed for by the assessment tools'.[66]

This careful treatment by the Supreme Court of Western Australia poses a stark contrast to the otherwise prevailing inclination to place unrestrained and unchecked trust with technology, which is manifested beyond the U.S. context as well. Without a thorough understanding of the tools and technical expertise at

[64] *Director of Public Prosecutions for Western Australia v Mangolamara* (2007) 169 A. Crim. R. 379[2007] WASC 71, [165]. N Stobbs et al., 'Can Sentencing Be Enhanced by the Use of Artificial Intelligence?' (2017) 41 Criminal Law Journal 261, 274 (It argues that since such tools are still far 'more accurate than unstructured judicial observations', if they are adapted for Australia, 'they should be incorporated into the algorithmic equation').

[65] ibid, 165

[66] ibid 166. Studies, for example, cast doubt on the appropriateness of the LSI-R for indigenous female Australians. Generally speaking, Australian courts appear concerned about using algorithmic risk assessment tools in sentencing decisions while correction departments in several Australian jurisdictions such as New South Wales and Victoria have used the LSI-R. *See* I Watkins, 'The Utility of Level of Service Inventory- Revised (LSI-R) Assessments with NSW Correctional Environment' <https://www.correctiveservices.justice.nsw.gov.au/Documents/utility-of-level-of-service-inventory-.pdf> (It reports that New South Wales has used the LSI-R since 2002 and for the year 2008 alone, there were 37,211 LSI-R assessments completed for various purposes including pre-sentencing and pre-release court advice reports.); 'Fact Sheet: Assessing the Risk of Reoffending, Corrections Victoria' <http://assets.justice.vic.gov.au/corrections/resources/36957a65-3829-4a79-974e-4c1e6e66e7eb/fs_assessing_risk_reoffending.pdf>.

their disposal, judges will most likely maintain institutional inertia and shy away from vigorous debates about their problems.

2.1. Unfinished Business: Challenges of Legal and Technical Black Boxes

To have an informed discussion, it is crucial to understand the operation of COMPAS and similar risk assessment systems. It essentially consists of a series of steps in the order of (1) data input (2) processing and computation and (3) prediction output. The *Loomis* Court failed to examine the 'processing and computation' phase of the issue, which involves the most critical questions of how to interpret data and how to base prediction output on the interpretation. That is, in interpreting and applying 'the right to be sentenced on the basis of accurate information' according to the U.S. Supreme Court's ruling,[67] the *Loomis* Court, unfortunately, had an uninformed focus on the accuracy of 'data input' and 'prediction output' only. The algorithmic process—the 'processing and computation' phase, namely the core of the COMPAS system—seemed to be irrelevant to the *Loomis* Court and hence excluded from its understanding of 'accurate information'. The Court accordingly made a mistake in discussing only whether the defendant can challenge the accuracy of his criminal history and questionnaire answers ('data input'), without attending to whether he can challenge the critical 'processing and computation' stage.

The *Loomis* Court's ignorance of the processing and computation phase demonstrates exactly the problem of its lack of transparency, which is vividly described by Frank Pasquale as a 'black box.'[68] The problem of the 'black box' refers to the complexity and secrecy of the algorithmic process, which frustrates meaningful scrutiny of automated decision-making that has an immense impact on society. Without understanding and addressing the 'black box' challenge and its legal ramifications, the *Loomis* decision impedes the role of the judiciary as the final gatekeeper to protect individuals' rights and to provide effective remedies.

Legal Black Box

The black box problem should be further disentangled. First, as argued by one of the authors of this Article,[69] this decision highlights the issue of a

[67] The U.S. Supreme Court ruled in *Townsend v Burke*, that the due process right to fair sentencing included 'the right to be sentenced on the basis of accurate information' (see also *Gardner v Florida*).
[68] Pasquale (n 48).
[69] *See* CF Lin, 'Artificial Intelligence in the Court: Black Box, Due Process, and the Need for Greater Algorithmic Accountability', working paper presented at the 6th Annual Conference on Governance of

'*legal black box*'—that is, that the opacity in fact comes from the propriety characteristics of statistical models or source codes, which are legally protected by relevant trade secret statutes. A legal black box is present in cases of traditional computerization of sentencing which automates and simplifies legal reasoning or decision-making into determinable and concrete elements and patterns so as to promote accuracy, cost efficiency, and simplicity (e.g., COMPAS). One way to address this is to unpack the legal black box to the public upon specific conditions to secure a certain level of transparency and accountability. Public disclosure would likely be objected to by companies keen to protect the secrecy. An alternative would be disclosure only to interested parties, or to an expert committee in a confidential manner. For instance, the relevant laws could be amended to compel private firms earning profits while *performing essential public services to* disclose their algorithmic processes to court-approved parties or expert committees or at least to limit trade secret protection in such circumstances.[70] Given the importance of the public interest involved in these public services, secrecy for profit should be reasonably confined.

Technical Black Box

More problematic is what we called a '*technical black box,*' which occurs when AI techniques including machine learning and deep learning are involved.[71] The technical nature of AI techniques is characterized by an inherent lack of transparency, as decisional rules emerge automatically in ways that no one—even the programmers—can adequately explain why and how certain decisions and determinations are made. The technical black box problems, therefore, cannot be readily addressed by way of transparency requirements or the like.

Such problems can be most obvious in an 'artificial neural network' (ANN), one type of algorithms applied extensively across disciplines.[72] As its name suggests, an ANN mimics the human brain: it is comprised of various neurons (i.e. interconnected processors). Unlike 'expert systems' that are based

Emerging Technologies and Science: Law, Policy, and Ethics, Arizona State University Sandra Day O'Connor College of Law (16-18 May 2018), United States [on file with the author].

[70] For instance, the Pennsylvania Commonwealth Court once ruled that a 'trade secret contention ceases to be of any moment when the function is recognized as governmental, rather than that of a private business'. *Hoffman v Pennsylvania*, 455 A.2d 731, 733 (Pa. Commw. Ct. 1983). For a more thorough discussion, *see generally* DS Levine, 'The People's Trade Secrets?' (2011)18 Michigan Telecommunication Technology Law Review 61.

[71] *See* Lin, (n 69).

[72] *See eg,* JV Tu, 'Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regressions for Predicting Medical Outcomes' (1996) 49 (11) Journal of Clinical Epidemiol, 1225.

on a collection of 'hierarchical rules, variables and constants that they apply to a given problem to try and determine a solution,' [73] an ANN is a type of representational learning that does not require too much human intervention. The learning algorithms of the ANN are not programmed *a prior*;[74] rather, they learn the relationship between different information and patterns through a layered structure and develop their own decisional rules which are usually not intelligible to humans.[75] Although an ANN reduces the work of computer code development in expert systems and also increases the accuracy by extracting patterns in a massive amount of data, it comes at the expense of a 'human's ability to substantively explain the inferential reasoning that occurs in each layer'. [76] Therefore, given the highly non-linear nature, 'there is no well-defined method to easily interpret the relative strength of each input and to each output in the network' even for engineers.[77]

This '*technical black box*' problem may significantly frustrate governance efforts to foster transparency and accountability in the government's use of AI-based systems, especially those proposed and designed to undertake the '*legal black box*' problem. As elaborated in Section 2.3 below, a potential solution is to consider certain opt-in or opt-out mechanisms, allowing the defendant or those subject to computational decision-making, a certain degree of autonomy regarding the use of such technology in his or her case. The subject should be fully informed of the potential risks as well as the benefits of the use of the software. He or she should also be informed of the limits on understanding how the software makes projections the way it does. Unfortunately, the *Loomis* Court turned its back on both the legal black box issue and the more challenging, complex problem of the technical black box.

---

[73] M Aikenhead, 'The Uses and Misuses of Neural Networks in Law' (1996) Santa Clara Computer & High Technology Law Journal 31, 33.

[74] The operation of an ANN does require learnings. There are two common techniques: 'supervised learning' and 'unsupervised learning'. The former is more labor-intensive in that algorithms are designed based on a labelled dataset and are trained to look for the 'correct' values assigned to them. In contrast, the latter involves an unlabelled dataset and it is left for the algorithms to 'find regularities in input data without any instructions as to what to look for'. GF Hepner, 'Artificial Neural Network Classification Using a Minimal Training Set: Comparison to Conventional Supervised Classification' (1990) Photogrammetric Engineering and Remote Sensing, 469; Big Data, Artificial Intelligence, Machine Learning, and Data Protection, (2017) Information Commissioner's Office, 7-8 <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>.

[75] L Zhou et al., 'A Comparison of Classification Methods for Predicting Deception in Computer-Mediated Communication' (2004) 20 (4) Journal of Management Information Systems, 139, 149.

[76] P Margulies, 'Surveillance by Algorithms: the NSA, Computerized Intelligence Collection, and Human Rights' (2016) 68 Florida Law Review 1045, 1069.

[77] Zhou et al., (n 75), 150-151.

2.2. Other Perils of Outsourcing Public Powers to Machines

Policy and legal solutions are much needed in response to both *legal* and *technical* black box problems for similar issues can go beyond the context of criminal justice. While COMPAS and many similar tools have not yet reached the level of sophistication of an 'expert system' (or even that of neural networks noted above), the current practice indicates a global trend of increasing use of AI technology in the court. Governments, too, have increasingly moved towards a 'smart' world too by reinventing regulatory infrastructure through big data and algorithms—it has been commonplace for various public sectors to tap into big data analytics to automate their decision-making processes.[78] Without addressing the *legal* and *technical* black box problems, the use of these systems undermines the very core ideals of due process and accountability.

Many troubling questions remain. It is far from clear whether automated systems support, or replace human discretion and judgment in practice.[79] So who *are* the decision-makers—the government officials involved or software programs they are relying on, which are designed by companies for profit? In many cases, private entities wield public power by virtue of the algorithms they design, as demonstrated in the *Loomis* Court's use of COMPAS. On what legal basis is such a power exercised?

While the delegation of power is inevitable in governing today's world, governments must delegate only within their prescribed mandates and within the legal confines, which should be interpreted stringently. One should not assume, as Melissa Perry, a judge of Federal Court of Australia aptly remarks, that 'a statutory authority vested in a senior public servant which extends by implications to a properly authorized officer, will also extend to an automated system; nor that authority to delegate to a human decision-maker will permit 'delegation' to an automated system'.[80] Any delegation of this kind must be authorized by law expressly to put in place at least some *ex ante* scrutiny and democratic accountability.[81]

---

[78] Chicago's SmartData Platform, for instance, was introduced to help the government analyze trend data and engage in predictive problem-solving. Ash Center Mayors Challenge Research Team, Chicago's SmartData Platform: Pioneering Open Source Municipal Analytics <https://datasmart.ash.harvard.edu/news/article/chicago-mayors-challenge-367>.

[79] A Smith, 'iDecide: the Legal Implications of Automated Decision-making' Speech at Cambridge Centre for Public Law Conference 2014: Process and Substance in Public Law, 15-17 September 2014 <http://www.fedcourt.gov.au/digital-law-library/judges-speeches/justice-perry/perry-j-20140915>.

[80] ibid.

[81] In Australia, for instance, section 495A(1) of the Migration Act 1958 (Cth) states that 'the Minister may arrange for the use, under the Minister's control, of computer programs for any purposes for which the Minister may, or must, under the designated migration law: (a) make a decision; or (b) exercise any power,

However, even where there is express authority, several questions arise. First, what type of decisions can governments delegate to machines? Should the authority be shaped broadly enough to allow algorithms to make value-based judgments or only non-discretionary decisions? [82] Moreover, codes and algorithms are value-laden. Development of algorithms is a complex process that can be influenced by humans—such as criteria selection, data mining, training, semantics, and interpretation. [83] In designing the operational parameters, according to Brent Daniel Mitelstadt et al., it is not uncommon for developers to have 'desired outcomes in mind that privilege some values and interests over others'.[84] While sensitive attributes like gender, race, or ethnicity are generally disallowed in the decision-making process of public sectors, they may be encoded, inadvertently or not,[85] when private companies design the systems. The fact that algorithms are implemented throughout government agencies can only magnify and perpetuate the risks of hidden biases and errors.[86]

Yet, the perpetuation of such risks is much less visible for both governments and the public due to the lack of transparency and adequate capacity to interrogate the algorithms. Thus, private participation in public administration could produce negative consequences by 'hurt[ing] citizens and weaken[ing] public authority'.[87] All these problems underscore the challenges to human rights, the rule of law and democratic accountability in the age of big data. Thus, although efficiency and convenience seem to justify implementation of these automated systems, meaningful safeguards must exist.

---

or comply with any obligations; or (c) do anything else related to making a decision, exercising a power, or complying with an obligation'.

[82] The issue of the level of delegation underscores a related yet more problematic question: how can we ensure that machines reflect the laws and social norms cherished in a society? For a machine to be able to make a decision—whatever type it is—it involves a translation process by converting laws and policies into binary codes and algorithms. Such a 'law to code' translation process can be rather vexing, even for lawyers and judges who from time to time disagree on the interpretation of laws and cases, let alone engineers without a solid legal background. Hence, it can be a daunting task to ensure these translators faithfully discharge their duties. Smith, (n 79), 5.

[83] *See eg,* N Diakopoulos, 'Algorithmic Accountability' (2015) Digital Journalism, 3:3, 398, 402.

[84] BD Mittelstadt et al., 'The Ethics of Algorithms: Mapping the Debate' (2016) Big Data & Society 1.

[85] It is true, as Cathy O'Neil points out, 'racists don't spend a lot of time hunting down reliable data to train their twisted models'. C O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown 2016).

[86] ibid 29-31.

[87] Brauneis and Goodman (n 7), 117. According to Brauneis and Goodman, the fact that governments have increasingly moved towards an algorithmic decision-making process also raises concerns about a 'lock-in' effect. Governments' reliance on such a process would then lock them into proprietary technologies, whose 'costs and pace of innovation they cannot control'.

2.3. Stop Big Data from Running Wild: Suggestions for Moving Forward

Improper design and deployment of big data and algorithms in criminal justice or other contexts of public authority undermines due process, equal protection, and transparency. Policymakers and academics are beginning to grapple with the potential risks of the unfettered power of data. While this Article cannot offer an exhaustive list of possible solutions, many of which are still in the making, it identifies and recommends a number of useful approaches that can help harness such powers and increase accountability.

A useful starting point is to rethink the question: who is the decision-maker? In a world that often blindly portrays numbers to be scientific, neutral and objective,[88] human decision-makers are likely to surrender their powers to data. As seen in the *Loomis* decision, ill-informed deference to the privately-made machines marginalizes the role of public authority and public scrutiny in governance.

Conceivably, one way to redress this may be to draw a line between non-discretionary and discretionary decision-making process in relation to public officials' use of automated machines. Discretionary processes must require human intervention, although the fashion and level of such intervention is contingent upon factors like the nature of the decision, issue area, interests involved, available remedy, and resource distribution, etc. Further, a government may secure an appropriate level of protection by resorting to the best practices adopted by its oversea counterparts. For instance, the Australian government has issued a 'Best Practice Guide: Automated Assistance in Administrative Decision-Making', making clear that algorithms 'should not automate the exercise of discretion'.[89] An automated system, moreover, 'must be designed in a way that accurately reflects the government policy it models and agencies should be careful that the system does not fetter the decision-maker in exercising

[88] For a critique of the use in governance of numbers or indicators that create an appearance of objective science, see SE Merry, 'Measuring the World: Indicators, Human Rights, and Global Governance: with CA comment by John M. Conley' (2011) 52 (S3) Current Anthropology), S83-S95 ('Numbers have become the bedrock of systematic knowledge because they seem to be free of interpretation and to be neutral and descriptive. They are presented as objective, with an interpretive narrative attached to them by which they are given meaning. Numbers can be assigned to observed particulars in a way that makes them amenable to such manipulations and makes them amenable to a knowledge system that privileges quantity over quality and equivalence over difference'.); KE Davis et al., 'Indicators as a Technology of Global Governance' (2012) 46(1) Law and Society Review, 71-104; KE Davis et al., Governance by Indicators: Global Power Through Classification and Rankings (2012).

[89] 'Best Practice Guide: Automated Assistance in Administrative Decision-Making' (Australian Government, 2007, 14-15) <https://www.oaic.gov.au/images/documents/migrated/migrated/betterpracticeguide.pdf>14-15.

any discretion'.[90] Admittedly, however, such an approach has its limits. First, at present, data analytics seems to be widely deployed in discretionary decision-making in the government and the judiciary, and as discussed above, human decision-makers are likely to defer to these automated systems in practice. Moreover, even we mandate governments to have humans to make the final judgment call, it remains problematic to address the anchoring effect as long as we permit a role of data-driven approach in the decision-making process of the public sector. At the end of the day, the question comes down to the black box problems. To address the black box problems noted above, transparency, while not the ultimate answer, should be part of the discussion. In theory, a handy tool may be freedom of information laws—statutes that have been widely adopted in many jurisdictions to mandate government agencies to disclose certain data to the general public upon request. In practice, however, requesters still struggle to obtain information on algorithms. Consider the Freedom of Information Act (FOIA) of the U.S. The FOIA is the major instrument at the federal level for citizens to access previously unreleased government information. Generally, it requires 'agency records' be available to 'any person'—without any exemption—insofar as the records are reasonably described and under the 'control' of the agencies. The 'control' element is central to the public's access to government data. The U.S. courts in several cases such as *Gilmore v. Department of Energy*[91] and *Tax Analysts v. US Department of Justice*[92] seem to suggest that software offered by private vendors under government contracts fails to meet the 'control' test, thus would not be available under FOIA.[93] Consequently, so long as the requested algorithm remains privately owned and, as a result, is out of the reach of public agencies, FOIA disclosure is unlikely to succeed.

Additionally, vendors typically require non-disclosure and trade secret clauses in their contracts with government agencies. An empirical study indicates that such contractual arrangements are a major hurdle that frustrates FOIA requests for algorithms.[94] Some would argue that governments should have leverage to negotiate better terms, limiting the scope of non-disclosure and trade secrets.[95] Even so, however, this is subject to the discretion of government agencies on a case-by-case basis. To directly tackle the legal black box, free

[90] ibid.
[91] *Gilmore v Department of Energy* 4 F. Supp. 2d 912 (N.D. Cal. 1998)
[92] *Tax Analysts v United States* 913 F. Supp. 599, 603-04 (D.D.C. 1996).
[93] *See eg,* MJ Madison, 'Legal-Ware: Contract and Copyright in the Digital Age' (1998) 67 Fordham Law Review 1025.
[94] Brauneis & Goodman, (n 7), 137-52.
[95] ibid. 164-66.

information laws must change to require disclosure by private vendors when their software is used for governance.

While disclosure may mitigate the concerns about the legal black box, it seems less useful in resolving the technical black box problem that involves highly sophisticated algorithms. One way to (partially) address this is an opt-in or opt-out mechanism as briefly noted in Section 2.1 above. In this regard, several provisions introduced by the General Data Protection Regulation 2016/679 (GDPR) of the European Union (EU) may be of some value.[96] First of all, the subject should be informed of 'the existence of automated decision-making', and, at least in those cases, 'meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject'. Article 22 (1) of the GDPR provides the data subject with the right to opt out under certain circumstances, eg, 'the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her'.[97] This would enable the data subject to have a final say in the use of automated decision-making. Even where certain exceptions apply (such as the data subject's explicit consent, the necessity for the subject to enter into or perform contracts, or authorization by domestic laws under certain conditions), the data subject retains several crucial rights including those to 'obtain human intervention', 'to express his or her point of view and to contest the decision'.[98] Moreover, to ensure the automated system properly operates, the EU's Article 29 Data Protection Working Party (replaced by the European Data Protection Board) invoked Article 5 of the GDPR, requiring data accuracy at all stages in order to reduce flawed automated decision-making or profiling.[99] Although Article 22 (1) of the GDPR is insufficient in that it does not apply when the process is not 'solely' automated,[100] the safeguards it offers—such as human intervention, the right to explanation, the right to contest the decision and to make oneself heard—shed light on how to alleviate the concerns created by the technical black box. The approaches that have been introduced in Australia or the EU as noted above, however, are some of the many

[96] *See generally* S Wachter et al., 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31 Harvard Journal of Law and Technology 841.
[97] Regulation 2016/679 of the European Parliament and of the Council on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Advancement of Such Data, and repealing Directive 95/46/EC, 2016 O.J. L 119/1 [hereinafter the GDPR].
[98] ibid Article 22 (3).
[99] Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the Purposes of Regulation 2016/679 (17/EN WP 251rev.01, 2018), 11-12.
[100] For a critique, S Wachter et al., 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 (2) International Data Privacy Law, 76, 91-94.

initiatives that aim to remedy negative ramifications. A more thorough solution to ensure the ethics of algorithm designers and to hold them accountable involve lengthy multistakeholder deliberation.[101]

3. Concluding Remarks

Writing in the 1980s, a commentator warned us about the risks of using AI-enabled juries by predicting that 'robot juries would be marketed by profit-making firms...that would have an incentive to create programs biased in favor of conviction, since state officials are the intended purchasers'.[102] In today's world, automated systems aided by big data and algorithms increasingly function like the imagined 'robot juries' in terms of determining the fate of individuals. By revisiting the debate on *State v. Loomis*, we critique the judicial use of automated risk assessment tools in ways that undermine the fundamental values of due process, equal protection, and transparency. More broadly, the global trend of using data analytics and AI tools not just in courts but also throughout all levels of governments warrants a more engaging and informed approach to tackle the lack of accountability in the use of such technology by public powers, a problem well illustrated by *State v. Loomis*. We propose that, as a starting point, the 'black box' problem identified by Frank Pasquale must be treated seriously in both public and private spheres. This Article further lays out such a problem as two more nuanced ones—the legal black box and the technical black box—which pose different challenges in practice. Premised upon this nuanced and informed analysis, we evaluate various possible approaches across the United States, the European Union, and Australia—such as free information, mandatory disclosure, the right to explanation, mechanisms to opt in or opt out of automated decision-making, and other procedural safeguards—that might be helpful in addressing problems confronting the expansive algorithmization of government functions. Searching for a technologically-informed and socially-apt governance model in the age of big data can only be fruitful by engaging multistakeholders through constructive debate and dialogue.

---

[101] Policy debates on data-driven approaches to government functions also appear in many other jurisdictions. *See eg,* 'AI in the UK: Ready, Willing and Able?' (House of Lords Select Committee on Artificial Intelligence Report of Session 2017–19, 16 April 2018) <https://www.politico.eu/wp-content/uploads/2018/04/AI-in-the-UK-ReadyWillingAndAble-April-2018.pdf>; Senate Co-Sponsorship Memoranda, Senate of Pennsylvania, Session of 2017 - 2018 Regular Session (13 August 2018)<http://www.legis.state.pa.us/cfdocs/Legis/CSM/showMemoPublic.cfm?chamber=S&SPick=20170&cosponId=26230> (It notes that '[m]y legislation will require all risk algorithms or artificial intelligence programs to meet certain requirements. They must be shown to be free of bias toward any race, gender, or protected class. They must be periodically re-validated and revised in accordance with national best practices. The report of these revisions must be publicly available, along with information about the programs or algorithms and the risk factors they analyze').

[102] BM Tindall, 'Robot Jury Neither Impartial Nor Randomly Selected' (*Infoworld*, 25 April 1983,) 65.