The Rise and Fall of AI and Algorithms in American Criminal Justice

Lessons for Canada

LCO
CDO

LAW COMMISSION OF ONTARIO
COMMISSION DU DROIT DE L'ONTARIO

ABOUT THE LAW COMMISSION OF ONTARIO

The Law Commission of Ontario (LCO) is Ontario's leading law reform agency.

The LCO provides independent, balanced and authoritative advice on complex and important legal policy issues. Through this work, the LCO promotes access to justice, evidence-based law reform and public debate.

The LCO evaluates laws impartially, transparently and broadly. The LCO's analysis is informed by legal analysis; multi-disciplinary research; contemporary social, demographic and economic conditions; and the impact of technology.

The LCO is located at Osgoode Hall Law School, York University, Toronto.

More information about the LCO is available at www.lco-cdo.org.

Law Commission of Ontario Reports

Legal Issues in the Last Stages of Life (Forthcoming Early 2021)
Indigenous Legal Issues in the Last Stages of Life (Forthcoming Early 2021)
AI, Algorithms and Government Decision-Making (Forthcoming Fall 2020)
Regulating AI: An International Survey (Forthcoming Fall 2020)
Defamation Law in the Internet Age (March 2020)
Class Actions: Objectives, Experiences and Reforms (July 2019)
Legal Capacity, Decision-making and Guardianship (March 2017)
Simplified Procedures for Small Estates (August 2015)
Capacity and Legal Representation for the Federal RDSP (June 2014)
Review of the Forestry Workers Lien for Wages Act (September 2013)
Increasing Access to Family Justice (February 2013)
Vulnerable Workers and Precarious Work (December 2012)
A Framework for the Law as It Affects Persons with Disabilities (September 2012)
A Framework for Teaching about Violence Against Women (August 2012)
A Framework for the Law as It Affects Older Adults (April 2012)
Modernization of the Provincial Offences Act (August 2011)
Joint and Several Liability Under the Ontario Business Corporations Act (February 2011)
Division of Pensions Upon Marriage Breakdown (December 2008)
Fees for Cashing Government Cheques (November 2008)

CONTRIBUTORS

The following individuals contributed to research or drafting this Final Report:

Law Commission of Ontario staff:

Nye Thomas, Executive Director, Law Commission of Ontario
Ryan Fritsch, Counsel, Law Commission of Ontario
Susie Lindsay, Counsel, Law Commission of Ontario
Natasha Daley, LCO Student Scholar, University of Windsor, Faculty of Law
Michael Piaseczny, LCO Student Scholar, University of Ottawa, Faculty of Law
Erin Chochla, LCO Student Scholar, Lakehead University, Bora Laskin Faculty of Law

Citation
Law Commission of Ontario, *The Rise and Fall of AI and Algorithms In American Criminal Justice: Lessons for Canada,* (Toronto: October 2020).

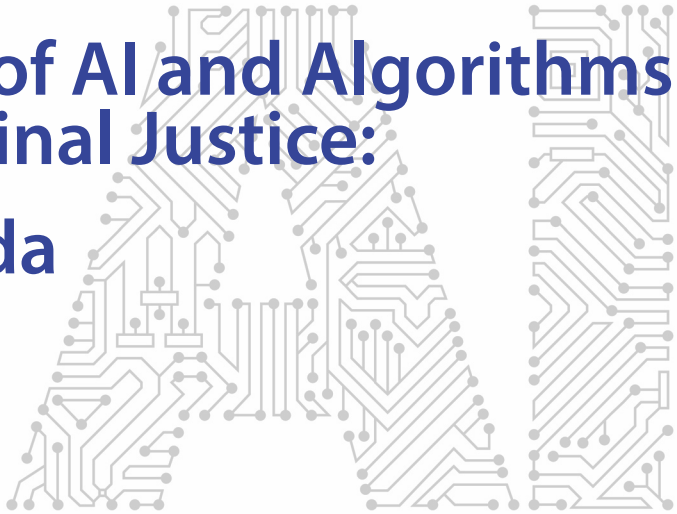
Law Commission of Ontario

2032 Ignat Kaneff Building
Osgoode Hall Law School, York University
4700 Keele Street
Toronto, Ontario, Canada
M3J 1P3

Email: LawCommission@lco-cdo.org
Web: www.lco-cdo.org
Twitter: @LCO_CDO
Tel: (416) 650-8406
Toll-free: 1 (866) 950-8406

The Rise and Fall of AI and Algorithms in American Criminal Justice: Lessons for Canada

October, 2020



I. INTRODUCTION

This is the first of three Law Commission of Ontario (LCO) Issue Papers considering the use of artificial intelligence (AI) and algorithms in the Canadian justice system. This paper considers the use of this technology in the criminal justice system.

The LCO's second Issue Paper, *Regulating AI: An International Survey,* considers current efforts to regulate AI and algorithms in government decision-making. The LCO's third Issue Paper, *AI, Algorithms and Government Decision-Making*, considers the use of AI and algorithms in civil and administrative law decision-making, such as determining welfare entitlements, administrative proceedings and government investigations.

The LCO's three Issue Papers provide an important first look at the potential use and regulation of these technologies in Canadian justice systems. Each paper identifies a series of important legal, policy and practical issues and choices that Canadian policymakers and justice system stakeholders should consider before these technologies are widely adopted in this country. The LCO's analysis is cumulative, with each paper building upon and adding to our work.

The LCO's research and consultations in this area draw heavily from the international experience with these systems, particularly in the United States. This experience provides both insights and important lessons.

The specific subject of this paper is algorithmic pretrial risk assessments. These are AI or algorithmic tools that aid criminal courts in pretrial custody or bail decision-making. According to the Partnership on AI (PAI), an American research organization focussing on best practices for AI, criminal risk assessment tools "present a paradigmatic example of the potential social and ethical consequences of automated AI decision-making."[1]

Algorithmic pretrial risk assessments are an important case study in the use of AI and algorithms in criminal justice. Bail proceedings adjudicate and balance fundamental liberty and public safety issues while needing to ensure high standards of due process, accountability and transparency.

The use of these tools has expanded rapidly across the United States, to the point where these systems are probably the most widely implemented AI or algorithmic tools to aid decision-making in criminal proceedings in the world. This expansion has been the catalyst for an unprecedented and rapid evaluation of how algorithmic tools in criminal proceedings are designed, developed and deployed. Suffice to say, this reform has not gone smoothly. In the space of a few short years, there has been an extraordinary backlash against the use of these systems, including by many of the same organizations and stakeholders who enthusiastically supported their development in the first place.

The LCO believes an analysis of the American debate can provide Canadians with important insights and lessons about the use of AI and algorithms in criminal proceedings. The LCO further believes that many, if not most, of these lessons are applicable to the use of these tools in civil and administrative decision-making as well.

Consistent with our mandate, the LCO's objective is to provide an independent, balanced and forward-looking analysis of the legal and policy issues that will likely arise in the Canadian context if, or more likely when, these tools are considered more extensively by Canadian governments, courts and tribunals.

Some of the issues discussed in this paper will likely be familiar, such as widely discussed "black box" criticisms of AI and algorithmic decision-making. Other issues are likely to be new or unfamiliar, such as questions regarding the "metrics of fairness" and emerging best practices.

The paper pays particular attention to issues regarding racism and data discrimination. Anti-Black and anti-Indigenous racism has been a long-standing concern in Ontario's justice system. Accordingly, any analysis of AI and algorithms in criminal proceedings must address these issues clearly and comprehensively. Perhaps not surprisingly, the LCO has identified many unexplored, unregulated and poorly understood issues respecting data, discrimination, algorithms and the law. The paper also considers this technology from the related and overlapping perspective of access to justice for low-income and vulnerable communities.

The paper concludes with an analysis of regulatory issues and options to assist Canadian policymakers and stakeholders identify, discuss and decide appropriate options and instruments for the Canadian criminal justice system.

II. THEMES AND LESSONS LEARNED

The LCO begins by noting several important themes that run through this Issue Paper and the LCO's technology-related work:

- **AI, algorithms and automated decision-making are a significant new frontier in human rights, due process and access to justice**. The use of AI, algorithms and automated-decision making are expanding rapidly in justice systems across the world. This expansion raises new and crucial questions about equality, bias, access to justice and due process in areas of law affecting fundamental rights.
- **Simple solutions and complex problems.** AI and algorithms offer many benefits, including the potential to provide consistent, "evidence-based" and efficient predictions. Unfortunately, experience demonstrates the risk of adopting unproven and under-evaluated technologies too quickly to address long-standing, complex and structural problems, such as systemic racism in the justice system.
- **AI and algorithms often embed, and obscure, important legal and policy choices.** AI and algorithms are not "objective" or neutral because they are based on data. Seemingly technical decisions often embed far-reaching policy or legal choices without public discussion or accountability. The distinction between "code choices" and "policy choices" is sometimes difficult to appreciate.
- **There are data issues and choices at every stage of AI and algorithmic decision-making.** Data issues and choices are endemic to every aspect of an AI or algorithmic system. Data issues and choices can be both consequential and controversial. For example, many AI and algorithms are criticized on the basis of historically racist, discriminatory or biased data. Other data issues include questions regarding statistical "metrics of fairness" and the accuracy, reliability and validity of datasets. Simply stated, data issues and choices are foundational to the success and legitimacy of any AI or algorithmic tool used by government, courts or tribunals.
- **AI and algorithmic systems make predictions; they do not set policy, make legal rules or decisions.** AI, algorithms and risk assessments are statistical tools that help make predictions. Courts, legislatures and policymakers decide how to turn those predictions into "action directives" or legal decisions. Whether algorithms worsen or lessen bias, or are coercive or supportive, depends on how human decision-makers decide these tools are used.
- **Legal protections regarding disclosure, accountability, equality and due process for these systems are often inadequate.** In the criminal justice system, AI and algorithmic tools must be held to high legal standards. Unfortunately, many of the legal issues raised by this technology are unexplored, unregulated and poorly understood. Current models of legal regulation and accountability have not kept pace with technology. Emerging models of "technological due process" suggest a constructive way forward.
- **AI and algorithms in the criminal justice system raise important access to justice issues.** Using AI and algorithms in criminal proceedings means that criminal accused potentially face even higher hurdles in presenting a full answer and defence to the charges against them. These additional hurdles may compound existing barriers to access to justice and lead to greater over-representation of low-income and racialized communities in the criminal justice system.

- **The criticisms of AI and algorithms are legitimate, but there are also opportunities and emerging best practices.** There are many significant and legitimate criticisms of algorithmic tools. At the same time, much has been learned about how to design, develop, implement and evaluate these systems. Many legal organizations, technologists and academics have begun to develop best practices and/or legal regimes necessary to improve these tools.
- **There must be broad participation in the design, development and deployment of these systems.** Unequal access to information and participation in decision-making can significantly worsen existing biases and inequality. This participation must include technologists, policymakers, law makers, and, crucially, the communities who are likely to be most affected by this technology.
- **Comprehensive law reform is needed.** The systemic legal issues raised by this technology cannot be addressed through individual litigation, best practices or piecemeal legislation. Comprehensive law reform is required. There are many potential legislative or regulatory responses, but the choices and options between these responses are complex and consequential. The Canadian legal system must proactively address important issues and options prior to widespread implementation of these systems.
- **Incremental reforms, deliberately.** Canadians need to be thoughtful, deliberate and incremental when adopting technologies in the justice system that potentially have such an extraordinary impact on individual rights and justice system fairness and transparency.

III. ABOUT THE LCO

The Law Commission of Ontario (LCO) is Ontario's leading law reform agency. The LCO provides independent, balanced and authoritative advice on complex and important legal policy issues. Through this work, the LCO promotes access to justice, law reform and public debate.

LCO reports are a practical and principled long-term resource for policymakers, stakeholders, academics and the general public. LCO reports have led to legislative amendments and changes in policy and practice. They are also frequently cited in judicial decisions, academic articles, government reports and media stories.

The LCO's current portfolio includes projects respecting Digital Rights, family protection orders, Last Stages of Life, Indigenous Last Stages of Life, and environmental rights. More information about the LCO is available at www.lco-cdo.org.

This paper is part of the LCO's ongoing Digital Rights project. The first phase of this project brings together policymakers, legal professionals, technologists, NGOs and community members to discuss the development, deployment, regulation and impact of AI and algorithms on access to justice, human rights, and due process. The project considers this technology in both the criminal[2] and civil/administrative law[3] justice systems. Other LCO Digital Rights projects include:

- Consumer Protection in the Digital Marketplace[4];
- Defamation Law in the Internet Age[5];
- AI, Automated Decision-Making: Impact on Access to Justice and Legal Aid;
- AI for Lawyers: A Primer on Artificial Intelligence in Ontario's Justice System with Element AI and Osgoode Hall Law School; and,
- Roundtable on Digital Rights and Digital Society with the Mozilla Foundation.

IV. ORGANIZATION OF THE PAPER

This paper is organized as follows:

Section 5 is an overview of the benefits and drawbacks of AI and algorithms in government decision-making.

Section 6 summarizes the use of AI and algorithmic tools in government and judicial decision-making, including the criminal justice system.

Section 7 describes the growth, use and reconsideration of algorithmic risk assessments in the United States.

Section 8 describes how pretrial risk assessments work, including a detailed look at the Public Safety Assessment (PSA), the most common algorithmic risk assessment tool in the US.

Section 9 summarizes a number of important issues, choices and lessons that can be learned from the American experience with algorithmic risk assessments, including:

Data Discrimination, Transparency and Scoring

Issue #1 Bias In, Bias Out
Issue #2 The "Metrics of Fairness"
Issue #3 Data Transparency
Issue #4 Data Accuracy, Reliability and Validity
Issue #5 Data Literacy: Risk Scores and Automation Bias

Law, Policy and Litigation

Issue #6 The Difference Between Predictions, Law and Policy
Issue #7 Best Practices in Risk Assessments
Issue #8 Public Participation
Issue #9 Algorithmic Accountability
Issue #10 The Limits of Litigation

Section 10 discusses the elements of a comprehensive law reform approach to AI and algorithms in criminal proceedings.

Section 11 is the LCO's Conclusion in which we describe current initiatives to rethink algorithmic risk assessments and offer suggestions for moving forward.

V. THE POTENTIAL AND PERILS OF AI AND ALGORITHMS

AI and algorithms are often referred to as "weapons of math destruction."[6] Many systems are also credibly described as "a sophisticated form of racial profiling."[7] These views are widespread in many current discussions of AI and algorithms.

The debate about pretrial algorithmic risk assessments in the United States is a vivid and important illustration of how and why these views have become commonplace. This paper will discuss (and to some extent confirm) these interpretations. The US debate also clearly demonstrates why AI and algorithms were so readily adopted and what has been learned in the intervening years.

Importantly, the catalyst for AI and algorithms to address American bail issues was neither cost-cutting nor "law-and-order" politics. Rather, the growth of these systems was driven in large part by a progressive, bipartisan American bail reform movement that many Canadians would, in principle, recognize and agree with. In this view, algorithmic risk assessments were important because of their potential to be neutral, consistent and evidence-based tools that would help transform the arbitrary, opaque, and often-racist pretrial decision-making of individual judges, prosecutors and justice systems.

These tools were seen, at least to some, as being in some respects superior to subjective, human assessments of risk in that they "eliminate the variability, indeterminacy, and apparent randomness—indeed, the subjectivity—of human prediction that has long pervaded criminal justice….[Algorithms can] bring uniformity, transparency, and accountability to the task."[8] Key to their growth and popularity was "harnessing the power of data to aid decision-making,"[9] an objective well-known to Canadian policymakers.[10]

As will be seen below, many of these early assumptions proved to be problematic. Indeed, one of the key lessons of this story is the risk of adopting unproven and under-evaluated technologies too quickly to address long-standing, complex and structural problems in the justice system.

The challenge of reconciling the potential risks and benefits of AI and algorithms is summed up concisely in the Preamble to the *Toronto Declaration on AI*, a 2018 statement prepared by AccessNow, an international organization that promotes understanding of human rights in the use of technology.[11] The Preamble reads, in part,

> *As machine learning systems advance in capability and increase in use, we must examine the positive and negative implications of these technologies. We acknowledge the potential for these technologies to be used for good and to promote human rights but also the potential to intentionally or inadvertently discriminate against individuals or groups of people.*
>
> *From policing, to welfare systems, online discourse, and healthcare – to name a few examples – systems employing machine learning technologies can vastly and rapidly change or reinforce power structures or inequalities on an unprecedented scale and with significant harm to human rights. There is a substantive and growing body of evidence to show that machine learning systems, which can be opaque and include unexplainable processes, can easily contribute to discriminatory or otherwise repressive practices if adopted without necessary safeguards.*[12]

The LCO will explore these contrasting views throughout this paper. The paper will also highlight recent developments that potentially offer a constructive way forward.

VI. AI AND ALGORITHMS IN GOVERNMENT AND JUDICIAL DECISION-MAKING

A. What Is AI and Algorithmic Decision-Making?

What are algorithms, automated decision-making, and AI?

The AI Now Institute defines algorithms and AI as follows:

> *An Algorithm is generally regarded as the mathematical logic behind any type of system that performs tasks or makes decisions…*
>
> *Artificial Intelligence (AI) has many definitions, and can include a wide range of methods and tools, including machine learning, facial recognition, and natural language processing. But more importantly, AI should be understood as more than just technical approaches. It is also developed out of the dominant social practices of engineers and computer scientists who design the systems, and the industrial infrastructure and companies that run those systems. Thus, a more complete definition of AI includes technical approaches, social practices and industrial power.*[13]

The difficulty of defining AI and algorithms has been noted by many. For the purpose of this report, the LCO is focussing on pretrial risk assessments, which are forms of AI and algorithms that operate as predictive models or tools.[14] In this paper, the LCO will use the phrases "algorithmic decision-making" and "AI" to describe this range of technologies.

B. How Is AI and Automated Decision-Making and AI Being Used in Justice Systems?

It is well-known that automated decision-making and AI is increasingly being used in a wide range of public and private contexts. It is also generally known that automated decision-making and AI is likely to have an impact on justice systems. What is less well-known, however, is the extent to which automated decision-making systems and AI is *already* being used in justice systems internationally, including US, the UK and Europe, including:[15]

- **Access to Government Benefits:** Automated decision-making is being used to determine eligibility for access to health and other government benefits.
- **Access to Housing:** Automated decision-making is being used to prioritize and determine eligibility for permanent or temporary housing.
- **Child Welfare:** Automated decision-making is being used to assess risk of current or future harm to a child.
- **Domestic Violence.** Automated decision-making is being used to assess victims' level of risk for future abuse.
- **Education:** Automated decision-making is being used to predict whether students are a high risk for school-related violence.
- **Surveillance Technologies:** Automated decision-making is being used by law enforcement agencies to support police surveillance.
- **Immigration:** Automated decision-making is being used to recommend immigration eligibility or status.

- **Predictive Policing:** Automated decision-making is being used to analyze data to help predict either where crime will occur or who will be involved in crime.
- **Bail:** Automated decision-making is being used to assess the suitability of releasing criminal accused on bail.
- **Sentencing:** Automated decision-making is being used to recommend sentencing for criminal accused, including whether an accused is at high or low risk of reoffending.
- **Inmate Housing Classification:** Automated decision-making is being used to recommend prison classification and conditions for inmates.
- **Parole:** Automated decision-making is being used to recommend parole eligibility or conditions.

The applications currently in use internationally affect fundamental rights, significant government entitlements, crucial human rights and important access to justice issues, including "poverty law", child welfare, criminal law, and refugee/immigration issues. This is not a complete list.[16] The range and complexity of these applications will be discussed more extensively in the LCO's second and third Issue Papers.

Transposed to the Canadian context, the applications currently in use internationally would affect some of Canadian's most important government services and the jurisdiction and workload of many Superior Courts, provincial courts, administrative tribunals, government ministries, agencies and municipalities.

The potential reach of AI and algorithms to aid justice system decision-making is staggering. For example, the US Social Security Administration (SSA), which is described as "the largest adjudication agency in the western world,"[17] is a leader on using AI and algorithmic tools to assist "mass adjudication."[18] SSA applications currently include "clustering" (grouping cases together to improve case processing), triaging (accelerating appeals based on their likelihood of success) and quality assurance (analyzing draft decisions against more than 30 "quality flags" that are suggestive of policy non-compliance or internal inconsistencies).[19]

A recent Stanford/New York University study of these developments notes that

> *AI could transform what it means to adjudicate a case. To be sure, current use cases are a far cry from full automation of adjudication, but the trajectory raises profound implications…*[20]
>
> *[T]he trajectory of AI tools in adjudication raises the normative question about the desired extent of discretion in adjudication.*
>
> *Formal adjudication requires that a decision be based on the exclusive record, but AI tools involve a transfer of decision-making authority away from line-level adjudicators toward AI developers.*
>
> *…the adoption of AI tools could potentially erode the decisional independence of and de novo review by [adjudicators].*
>
> *…the adoption of such tools can heighten concerns of bias.*[21]

The long-term implications of, and response to, AI and algorithms in the justice system is well beyond the scope of this paper. For present purposes, it is important to note that the use of AI and algorithms in government decision-making is growing rapidly. The Stanford/New York University study noted above canvassed the use of AI in 142 of the most significant US federal departments and agencies. The study found that "nearly half of [American] federal agencies studied (45%) have experimented with AI and related machine learning (ML) tools."[22] Growing use of AI and algorithms has also been found in government-use studies in the UK, Australia and New Zealand.[23]

Unfortunately, at present there are no equivalent Canadian studies. Nor is there a central list or repository of automated decision-making systems in use in the Canadian justice system or in other Canadian government

applications. As a result, it is very difficult to assess how widespread this technology is being used in Canada. In Canada (and elsewhere), these systems are often disclosed as a result of litigation, freedom of information requests, press reports or review of government procurement websites.

Disclosure of automated decision-making in Canada is further complicated by the fact that these systems can be used by a wide range of government actors, including federal and provincial government ministries, municipal governments and government agencies (including but not limited to tribunals, school boards, police services, and others).

C. Algorithms in the Criminal Justice System

Several recent reports provide a useful overview of the many ways in which AI and algorithms are being used in the criminal justice system in the United States, the UK and elsewhere, including: [24]

- Photographic and video analysis, including facial recognition;
- DNA profiling and evidence, including predictive genomics;
- Predictive crime mapping (predictive policing);
- Mobile phone and extraction tools;
- Data mining and social media intelligence; and,
- Individual risk assessment and prediction.

These reports note several potential uses of AI and algorithms within each of these broad categories. For example, within the risk assessment category, AI and algorithms are being used in the following contexts:[25]

- Bail algorithms that predict recidivism;
- Sentencing algorithms that predict recidivism;
- "Scoring at arrest" algorithms that advise how to charge an individual;
- "Scoring suspects" algorithms that analyze an individual's behaviour in the future;
- "Scoring victims" algorithms that predict likelihood of being a victim of crime; and,
- Correctional algorithms that predict propensity to reoffend within an institution.

The most comprehensive Canadian analysis of AI and algorithms in policing in Canada is a recent report by The Citizen Lab, *To Surveil and Predict, A Human Rights Analysis of Algorithmic Policing in Canada*.[26] This report analyzes the use and implications of predictive policing in Canada, including location- and person-focussed algorithmic policing and algorithmic surveillance technology, such as facial recognition and social network analysis. The report discusses the use of these technologies by police departments in Vancouver, Toronto, Calgary and Saskatchewan and other departments, including the RCMP. The report notes that, at this point, "algorithmic policing systems does not appear to be widespread in Canada."[27]

VII. CASE STUDY: RISK ASSESSMENT AND BAIL REFORM IN THE UNITED STATES

A. The "Breathtaking" Growth of Pretrial Risk Assessments

As noted in the Introduction, the use of pretrial algorithmic risk assessments has expanded rapidly across the United States to the point where these systems are probably the most widely implemented algorithmic tools in use in criminal proceedings in the world. Algorithmic pretrial risk assessment tools are used to predict how likely it is that an accused will miss an upcoming court date or commit a crime before trial.

The growth of algorithmic pretrial risk assessments appears to be driven by several factors, including increases in computing power, growing academic interest in criminal justice research, the "evidence-based criminal justice movement" [28] and, most importantly, the American bail reform movement. According to Logan Koepke and David Robinson, the central goal of American bail reform is to "end wealth-based [bail] system and move pretrial justice systems to a risk-based model."[29]

American bail reform advocates proposed several strategies to address these issues, including mandatory release for minor criminal offenses, greatly expanded diversion and community programs, improved electronic monitoring, expanded pretrial services and algorithmic pretrial risk assessments. Of all these initiatives, however, algorithmic risk assessments quickly emerged as the "favored reform."[30]

The attraction to these tools was understandable: Algorithmic risk assessments were considered neutral, consistent, evidence-based tools that would help transform the arbitrary, opaque, and often-racist pretrial decision-making of individual judges, prosecutors and justice systems. This view was confirmed by several early research studies which appeared to validate the positive impact of these tools.[31] The anticipated benefits of algorithmic risk assessments was articulated by the Center on Court Innovation, a New York-based non-profit research organization, as follows:

> *The appeal of pretrial risk assessment—especially in large, overburdened court systems—is of a fast and objective evaluation, harnessing the power of data to aid decision-making.*[32]

Similarly, a widely distributed commentary in the *New York Times* stated:

> *Bail decisions have traditionally been made by judges relying on intuition and personal preference, in a hasty process that often lasts just a few minutes…*
>
> *To combat such arbitrariness, judges in some cities now receive algorithmically generated scores that rate a defendant's risk of skipping trial or committing a violent crime if released. Judges are free to exercise discretion, but algorithms bring a measure of consistency and evenhandedness to the process.*
>
> *Studies [show that] data and statistics can help overcome the limits of intuitive human judgments, which can suffer from inconsistency, implicit bias and even outright prejudice.* [33]

Notably, pretrial risk assessments had broad bipartisan political support in the United States, at least initially. The *Pretrial Integrity and Safety Act of 2017,* jointly sponsored by Democratic Senator Kamala Harris and Republican Senator Rand Paul, proposed broad reforms to the American bail system, including

> *(A) replacing money bail systems with individualized, pretrial assessments that—*
>
> *(i) measure the risk of flight and risk of anticipated criminal conduct posed by a defendant while on pretrial release; and*

> *(ii)* ***shall use risk-based decision making that includes objective, research-based, and locally validated assessment tools*** *that do not result in unwarranted disparities on the basis of any classification protected under Federal nondiscrimination laws or the nondiscrimination laws of the applicable State;*[34] [Emphasis added.]

Significantly, risk assessments were also "strongly" endorsed as a "necessary component of a fair pretrial release system" by a broad coalition of public defender and civil rights organizations, including the American Council of Chief Defenders (ACCD), Gideon's Promise, the National Association of Criminal Defense Lawyers (NACDL), the National Association for Public Defense (NAPD), and the National Legal Aid and Defender Association (NLADA).[35]

With this kind of support, it is not surprising the use of pretrial risk assessments grew rapidly across the United States, which has been variously described as "breathtaking"[36] and "hard to overestimate".[37] In 2017 alone, as many as 14 states made provisions to adopt or investigate the use of pretrial risk assessment tools.[38] In California, 49 of 58 counties use algorithmic risk assessment tools.[39]

B. Pretrial Risk Assessments Reconsidered

Notwithstanding this rapid expansion, many researchers, organizations and civil rights advocates began to quickly re-evaluate pretrial risk assessments. The result has been a remarkable reversal in the legal and political support for these tools.

This reassessment has been driven by several factors, including the experience of jurisdictions that implemented pretrial risk assessments, new research asserting that pretrial risk assessments may perpetuate racial bias, and a reconsideration of the utility of risk assessments relative to other bail reform strategies.

Perhaps the most significant public milestone in the pretrial risk assessment debate was the 2016 publication of an article by ProPublica, an American investigative journalism organization. The article, titled "Machine Bias," summarized ProPublica's research on the use of the COMPAS risk assessment tool to inform criminal sentencing in Broward County, Florida.[40]

Broward County used COMPAS to help predict whether individual defendants were likely to be recidivists. ProPublica obtained the COMPAS risk scores assigned to more than 7,000 people arrested in 2013 and 2014 to determine how many were charged with new crimes over the next two years. The ProPublica article was damning:

> ***The [COMPAS] score proved remarkably unreliable in forecasting violent crime: Only 20 percent of the people predicted to commit violent crimes actually went on to do so.***
>
> ***When a full range of crimes were taken into account — including misdemeanors such as driving with an expired license — the algorithm was somewhat more accurate than a coin flip.*** *Of those deemed likely to re-offend, 61 percent were arrested for any subsequent crimes within two years.*
>
> *We also turned up significant racial disparities…In forecasting who would re-offend, the algorithm made mistakes with black and white defendants at roughly the same rate but in very different ways.*
>
> ***The formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants.***
>
> *White defendants were mislabeled as low risk more often than black defendants. [Emphasis added.]*[41]

ProPublica concluded by stating that COMPAS was "biased against blacks" because Black defendants were over-classified as risky.[42]

The ProPublica analysis was vigorously debated on both methodological and policy grounds.[43] Nevertheless, the ProPublica article "supercharged" the emerging debate on risk assessments and racial bias.[44]

C. Algorithmic Risk Assessments in the US Today

It is hard to generalize about the current status of pretrial risk assessments in the United States.

On the one hand, many American jurisdictions have adopted these tools, with more appearing ready to follow. On the other hand, many of the original supporters of these systems now argue that algorithmic risk assessments should play no role in pretrial administration. For example, organizations such as Human Rights Watch now argue that pretrial risk assessment tools should be opposed "entirely."[45] Opposition to pretrial risk assessments is widespread and growing. Just recently, more than 100 civil rights, social justice, and digital rights groups issued "A Shared Statement of Civil Rights Concerns" declaring that risk assessment instruments should not be used in pretrial proceedings, or at least that their use should be severely circumscribed.[46]

The most striking repudiation of algorithmic risk assessments came from one of their most ardent supporters, the Pretrial Justice Institute (PJI), an American research and advocacy organization dedicated to bail reform. The PJI was one of the most prominent and influential supporters of algorithmic risk assessments. Yet, in a statement dated February 7, 2020, the PJI stated

> ***The intense studying and listening we have done over the last year has provided us with a deeper sense that there is no pretrial justice without racial justice.*** *We now see that pretrial risk assessment tools, designed to predict an individual's appearance in court without a new arrest, can no longer be a part of our solution for building equitable pretrial justice systems.* ***Regardless of their science, brand, or age, these tools are derived from data reflecting structural racism and institutional inequity that impact our court and law enforcement policies and practices. Use of that data then deepens the inequity…***
>
> *We made a mistake—we did not have the right people at the table when we were designing our roadmap to decarceration, particularly individuals directly impacted by the system. As we pushed forward, some places saw significant increases in pretrial liberty, but many did not, and racial disparities persisted in both.*
>
> ***We were wrong for having risk tools as part of our "smart" pretrial justice framework.*** *In the places that have undertaken reform, success hasn't hinged on an assessment tool; it has been driven by a commitment to decarceration, values-based discussions about the purpose of detention, a willingness to acknowledge the humanity of everyone, and each system's openness to change...*[47] *[Emphasis added.]*

Later that month, the developers of the Public Safety Assessment, the most widely used risk assessment tool in the US, released a statement in which they also partially stepped back from these tools. This statement reads, in part,

> ***Assessments, including the Public Safety Assessment, can play a positive role in a jurisdiction's pretrial system. However, implementing an assessment alone cannot and will not result in the pretrial justice goals we seek to achieve.*** *No single strategy, practice, or tool can achieve these ends.*[48] *[Emphasis added.]*

These statements are having an impact. Several jurisdictions have or are considering withdrawing their initial support for algorithmic risk assessments.[49]

D. What Should Canadians Learn?

From a Canadian perspective, the American experience with pretrial risk assessments can seem bewildering. The rapid expansion of algorithmic risk assessments, coupled with their equally rapid demise (at least in some quarters), represents an extraordinary – and extraordinarily quick – reversal and critical reassessment.

Notwithstanding these circumstances, the LCO believes Canadians can learn important lessons about both algorithmic pretrial risk assessments and algorithms in the criminal justice system generally.

Before addressing these issues, however, it is important to get a better understanding of how algorithmic pretrial risk assessments work.

VIII. HOW DO PRETRIAL RISK ASSESSMENT SYSTEMS WORK?

Risk assessments have been used in the American criminal justice system in some form or another since the early twentieth century. [50] *Algorithmic* risk assessments, the subject of this paper, are sometimes described as "fourth generation" risk assessment tools.[51]

Risk assessments are statistical models used to predict the probability of a particular future outcome. In the pretrial context, risk assessment tools are used to predict how likely it is that an accused will miss an upcoming court date or commit a crime before trial. This is done by using an algorithm, or statistical model, to compare and measure an accused's *individual* characteristics (such as age, gender, criminal history, etc.) against historic data for *groups* of people. These tools use a checklist of risk factors and data that are said to statistically correlate with nonappearance in court or the commission of a crime before trial.

The Partnership on AI notes "[t]hough they are usually much simpler than the deep neural networks used in many modern artificial intelligence systems, criminal justice risk assessment tools are basic forms of AI."[52]

Generally speaking, the outcome of a risk assessment is some kind of risk score. Risk scores are then transformed into a generalized risk category. For example, a tool might identify an accused as being low, moderate or high risk of missing an upcoming court date or committing a crime before trial.

Risk assessments typically *do not* directly recommend or mandate whether or not an accused should be detained, released, or what conditions should be ordered. Rather, risk assessments are often accompanied by a separate "decision-making framework" or "matrix" that sets out explicit directives or recommendations specifying how an accused in each risk category should be treated. "Decision-making frameworks" are not algorithmic or mathematical determinations, but rather policy decisions determined by the tool designers, local laws and practices, and/or policymakers.

A. A Closer Look at the PSA

This section takes a closer look at the Public Safety Assessment (PSA), perhaps the most widely used and sophisticated pretrial risk assessment tool, to demonstrate how pretrial risk assessment tools have been developed, administered and used in criminal courts across the United States.[53]

According to its developers, the PSA is a research-based pretrial assessment that provides judicial officers with information to help assess a person's likelihood of returning to court for future hearings and remaining crime free while on pretrial release.

The PSA generates scores that predict three pretrial outcomes:

- Failure to appear in court pretrial;
- New criminal arrest while on pretrial release; and,
- New violent criminal arrest while on pretrial release.

The PSA's developers state that the PSA was created using the largest, most diverse set of pretrial records ever assembled—approximately 750,000 cases from roughly 300 jurisdictions across the United States. They further state that the developers identified and tested hundreds of variables that were ultimately reduced to nine factors that, in their view, most effectively predicted new criminal arrest, failure to appear and new violent criminal arrest.

According to its developers, the PSA does not rely on factors such as race, ethnicity, or geography. Rather, the factors used by the PSA are:

- The person's age at the time of arrest;
- Whether the current offense is violent;
- Whether the person had a pending charge at the time of the current offense;
- Whether the person has a prior misdemeanor conviction;
- Whether the person has a prior felony conviction;
- Whether the person has prior convictions for violent crimes;
- Whether the person has failed to appear at a pretrial hearing in the last two years;
- Whether the person failed to appear at a pretrial hearing more than two years ago; and
- Whether the person has previously been sentenced to incarceration.[54]

The PSA is currently being used statewide in Arizona, Kentucky, New Jersey, and Utah. It is also being used in a number of major cities and surrounding areas, including Allegheny County (Pittsburgh), Pennsylvania; Cook County (Chicago), Illinois; Harris County (Houston), Texas; Lucas County (Toledo), Ohio; Mecklenburg County (Charlotte), North Carolina; Milwaukee County, Wisconsin; New Orleans, Louisiana; San Francisco County, California; Santa Cruz County, California; and Tulare County, California.[55] According to the PSA website, "[h]undreds of localities across the United States use the Public Safety Assessment (PSA). Dozens more are in the process of implementing it."[56]

B. A PSA Report

In most, but not all, jurisdictions, it appears the PSA is prepared by a court-based office, often a local pretrial services office. A PSA report may be ordered by a court or prepared automatically for all criminal accused in a jurisdiction.

Once a PSA assessment is completed, a report is generated that summarizes the results, including a person's scaled risk scores and results for each risk factor.

Each system-generated report is unique to the person who has been assessed; it is defendant-specific, charge-specific, and assessment-specific. The standard PSA Report (reproduced below) provides an example of the layout and types of information presented to a judicial officer who makes pretrial decisions, including the following:

- Basic demographic information, including the defendant's name, arrest data and the date of the PSA;
- A list of current charges;
- PSA prediction scores for three pretrial outcomes (failure to appear before trial, new criminal activity and new violent criminal activity), ranked on a scale from 1 to 6;
- The defendant's answers to each of the nine risk factors identified by the PSA; and,
- The defendant's "Presumptive Release Level", based on the Release Conditions Matrix.

12A Standard PSA Report

Name: John Defendant
PID: 123456
Arrest Date: 06/15/17
PSA Completion Date: 06/16/17

Current Charge(s): 14-113.9 FINANCIAL CARD THEFT F 1

PSA Score

FAILED TO APPEAR

1	2	3	4	5	6

NEW CRIMINAL ACTIVITY

1	2	3	4	5	6

NEW VIOLENT CRIMINAL ACTIVITY FLAG: NO

Risk Factor:

1. Age of Current Arrest:	**23 or Older**
2. Current Violent Offense	**No**
2a. Current Violent Offense 20 Years Old or Younger:	**No**
3. Pending Charge at the Time of Offense:	**No**
4. Prior Misdemeanor Conviction:	**Yes**
5. Prior Felony Conviction:	**Yes**
5a. Prior Conviction:	**Yes**
6. Prior Violent Conviction	**0**
7. Prior Failure to Appear in Past 2 Years	**1**
8. Prior Failure to Appear Older than 2 Years	**Yes**
9. Prior Sentence to Incarceration:	**Yes**

Presumptive Release Level

Based on the Release Conditions Matrix, the defendant's presumptive release level is **Release Level 2.**

C. The PSA and Decision Framework and Release Conditions Matrix

The PSA's developers explicitly state that the PSA Report itself is only used to help *measure* an accused's pretrial risk, whereas a second set of policy frameworks (called the Decision Framework and the Release Conditions Matrix) are used to help *manage* that risk.

The Decision Framework and the Release Conditions Matrix are developed by local stakeholders when implementing the PSA. The Decision Framework and Release Conditions Matrix help match pretrial release conditions with an accused's PSA scores. They are supposed to reflect local statutes, court rules, and policy preferences regarding pretrial release, conditions or other "activities."

This distinction between a PSA score and the Decision Framework/Release Conditions Matrix is important, as the LCO will discuss later.

The charts below reproduce a PSA Release Conditions Matrix.

The first chart is a grid that matches a person's scores on two PSA scales (Fail to Appear and New Criminal Activity) to presumptive levels of pretrial release:

	New Criminal Activity (NCA)Scaled Score					
Failure to Appear (FTA) Scaled Score	**1**	**2**	**3**	**4**	**5**	**6**
1	**Release Level 1**	**Release Level 1**				
2	**Release Level 1**	**Release Level 1**	**Release Level 1**	**Release Level 1**	**Release Level 2**	
3		**Release Level 1**	**Release Level 1**	**Release Level 1**	**Release Level 2**	**Release Level 3**
4		**Release Level 1**	**Release Level 1**	**Release Level 1**	**Release Level 2**	**Release Level 3**
5		**Release Level 2**	**Release Level 2**	**Release Level 2**	**Release Level 2**	**Release Level 3**
6				**Release Level 3**	**Release Level 3**	**Release Level 3**

The second chart is a table detailing the specific conditions associated with each level of pretrial release:

	Pretrial Release Level		
Release Activities and Conditions	**1**	**2**	**3**
Mandatory Statutory Conditions	**Yes**	**Yes**	**Yes**
Court Reminders	**Yes**	**Yes**	**Yes**
Criminal History Checks Once per Month		**Yes**	**Yes**
Check-in Once per Month			**Yes**
Other Case-Specific Conditions			**If court-ordered**

D. Variations Amongst Pretrial Risk Assessments

There is considerable variability among pretrial risk assessments in use in the United States.

Pretrial risk assessments often vary in how they are administered, what data or factors they rely on, and the sophistication of their algorithms. For example, the data used in a pretrial risk assessment may be automatically prepopulated or input by humans. The form and content of risk assessment outcome reports can also vary considerably.

In the US, there may even be several different types of risk assessments used within the same jurisdiction. In California, for example, at least four different risk assessment tools are in use in different counties across the state, each with different attributes, definitions of misconduct and predictors.[57]

Generally speaking, inputs into a risk assessment include data about an accused's criminal history or criminal-related misconduct. Some risk assessment tools may also include data about the accused's socio-economic factors such as education, age, gender, marital status or neighbourhood. Risk assessment tools may include both static and dynamic factors.[58]

Some risk assessment tools may also rely on data from lengthy personal interviews that attempt to discern (and measure) the accused's history of drug and alcohol use, mental health, family and employment history. For example, the COMPAS tool, depending on how it is implemented, can rely on an interview with up to 137 questions or inputs covering both *risk* factors and *needs assessment* factors, such as

- Current charges
- Offense History
- Associates/Peers
- Family
- Financial Status
- Leisure/Recreation
- Residential Stability
- Social Environment
- Vocation
- Education
- Mental Health
- Substance Abuse
- Criminal Attitudes-Thinking[59]

Finally, one of the most controversial and significant differences between tools is whether they are developed by government agencies, courts, non-profit foundations or private companies. For example, the PSA was developed the Arnold Foundation, a non-profit foundation. The Arnold Foundation has publicly disclosed considerable information about how its algorithm operates. COMPAS, by way of contrast, was developed privately. COMPAS relies upon a proprietary algorithm and the company prevents the disclosure of information revealing how factors are weighed or how risk scores are determined. The failure to disclose this information has significant implications for algorithmic transparency and due process, as will be discussed below.

IX. LESSONS FOR THE CANADIAN CRIMINAL JUSTICE SYSTEM

In this section, the LCO summarizes a number of important lessons and observations regarding AI and algorithms in the American criminal justice system. The purpose of this review is to highlight the issues and questions that will likely arise in Canada if, or more likely when, Canadian policymakers begin to consider the use of AI or algorithmic tools in our criminal justice system. This review will also highlight the legal analysis, academic research, operational experience, community evaluations and best practices that have developed in response to algorithmic pretrial risk assessments. Importantly, the LCO believes that many of these issues and lessons have implications for the use of algorithmic tools in the civil or administrative justice systems as well.

This review is by no means definitive. The LCO's objective, rather, is to introduce important questions, choices, controversies and proposals so that Canadian policymakers and stakeholders are better-informed.

The LCO has identified ten important lessons and observations, divided into two broad categories:

Data Discrimination, Transparency and Scoring

Issue #1 Bias In, Bias Out
Issue #2 The "Metrics of Fairness"
Issue #3 Data Transparency
Issue #4 Data Accuracy, Reliability and Validity
Issue #5 Data Literacy: Risk Scores and Automation Bias

Law, Policy and Litigation

Issue #6 The Difference Between Predictions, Law and Policy
Issue #7 Best Practices in Risk Assessments
Issue #8 Public Participation
Issue #9 Algorithmic Accountability
Issue #10 The Limits of Litigation

A. Data Discrimination, Transparency and Scoring

All AI or algorithmic systems rely on data. Data trains the system. Data calibrates the system. And the "outputs" of these systems are typically some kind of data "score" or statistical prediction. As a result, data issues are a fundamental feature of AI and algorithmic design, development, implementation and oversight.

The "bias in, bias out" issue is probably the most widely discussed algorithmic data issue. The US experience illustrates, however, that questions about data discrimination, data transparency and data literacy go far beyond questions of historic bias.

Issue #1: Bias In, Bias Out

The most trenchant and troubling criticism of pretrial risk assessments – and many other forms of AI and algorithms in criminal justice – is that that they are racist. For these reasons, organizations such as Human Rights Watch believe algorithmic risk assessments to be "a sophisticated form of racial profiling."[60]

This argument has many aspects, but the most common concern relates to the use of historic data. The issue is summarized concisely by David Robinson and Logan Keopke:

> *Pretrial risk assessment instruments face an inherent legitimacy problem: The world of mass incarceration and racially inequitable criminal law that exists today also provides the data upon*

> *which pretrial risk assessment instruments are based. There is, justifiably, distrust that tools developed on data reflecting this racially inequitable system will avoid perpetuating these patterns, let alone advance substantial reform.*[61]

In its most reductive form, this argument is straightforward: Because the training data or "inputs" used by risk assessment algorithms – arrests, convictions, incarceration sentences, education, employment – are themselves the result of racially disparate practices, the results or scores of pretrial risk assessments are inevitably biased. In other words: "bias in, bias out."[62]

It is important to note that race is not included as an explicit variable in of any of these systems. However, excluding race itself does not necessary mean that factors that correlate heavily to an individual's race are excluded. Nor are factors that have disparate impact based on the race of the individual, such as arrests or a question that asks a criminal defendant the number of times he or she has been stopped by the police.

For many in the US, the "bias in, bias out" argument is conclusive proof that algorithmic risk assessments and other algorithmic tools should *never* be used in criminal justice decision-making. For others, however, algorithmic risk assessment tools are valuable because they have the potential to *reveal* systemic bias and discrimination. For example, a number of scholars and advocates believe that

> *With the appropriate requirements in place, algorithms create the potential for new forms of transparency and hence opportunities to detect discrimination that are otherwise unavailable. The specificity of algorithms also makes transparent tradeoffs among competing values. This implies algorithms are not only a threat to be regulated; with the right safeguards, they can be a potential positive force for equity.*[63]

The contrast between these perspectives – algorithms as perpetuating bias versus algorithms as revealing bias – runs through the entire algorithmic debate.

From a Canadian perspective, the American "bias in, bias out" debate has clear implications for any Canadian jurisdiction considering algorithmic tools in any number of justice contexts where racially disparate practices have been proven or alleged, including policing, bail, sentencing and child welfare.

In the Canadian criminal justice system, the over-representation of Black people in arrests and convictions has been demonstrated repeatedly, most recently by the Ontario Human Rights Commission (OHRC).[64] The OHRC's August 2020 report, *A Disparate Impact,* demonstrated the racial disparity in arrests and charges by the Toronto Police Service. The OHRC's analysis, based on Toronto Police Service data, demonstrated that:

- In data from 2013 to 2017, Black people are "grossly" over-represented in discretionary, lower-level charges and are more likely than White people to face low-quality charges with a low probability of conviction.
- The charge rate for Black people was 3.9 times greater than for White people and 7.1 times greater than the rate for people from other racialized groups.
- Although they represented only 8.8% of Toronto's population in 2016 Census data, Black people represented 42.5% of people involved in obstruct justice charges and were 4.8 times more likely to be charged with obstruct justice offences than their representation in the general population would predict.[65]

Equally important is the long-standing overrepresentation of Indigenous persons in the Canadian criminal justice system. In a statement released earlier this year, the Federal Correctional Investigator reported that:

- While accounting for 5% of the general Canadian population, the number of federally sentenced Indigenous people has been steadily increasing for decades. More recently, custody rates for Indigenous people have accelerated, despite an overall decline in the inmate population.

- Since April 2010, the Indigenous inmate population has increased by 43.4%, whereas the non-Indigenous incarcerated population has declined over the same period by 13.7%.
- Indigenous women now account for 42% of the women inmate population in Canada.

The Correctional Investigator stated "[o]n this trajectory, the pace is now set for Indigenous people to comprise 33% of the total federal inmate population in the next three years."[66]

The OHRC and Federal Correctional Investigator data is consistent with earlier reports and studies addressing policing and convictions of Black and Indigenous persons in the Canadian justice system. This research is thoughtfully summarized in The Citizen Lab's report, *To Surveil and Predict,* discussed above.[67]

As in the US, any AI or algorithmic tool used in the Canadian criminal justice system is likely to be carefully scrutinized and challenged on the basis of "data discrimination." These challenges could take several forms, including, but not limited to, challenges to a system's training data. Discrimination-related challenges could also be based a system's statistical and predictive reliability, accuracy or bias.

These challenges could be framed on one or more of several grounds:

Most obviously, data discrimination could be challenged on basis of *Charter* sections 7 or 15. For example, section 7 could be used to challenge an automated decision-making process if data discrimination results in a depravation of life, liberty or security of the person. Similarly, section 15 could be used to challenge an AI or algorithmic tool if it is alleged to disproportionately impact, or be biased against, vulnerable groups on the basis of disability, socio-economic disadvantage, race or other factors. Data discrimination could also be challenged on the basis of the "intersectionality" of categories of discrimination (such as race, sex, mental disability, Indigenous identity).[68]

Data discrimination could also potentially be challenged though human rights provisions, privacy rights, tort law, class actions, or even the "precautionary principle" in environmental law.

The LCO believes it is incumbent on Canadian governments, legal professionals, academics, technologists, NGOs and community representatives to develop a comprehensive plan to analyze and address these important legal issues. Racial data discrimination by AI and algorithms has been studied and analyzed extensively in the American justice system. The LCO is not aware of equivalent scholarship in which racial discrimination, AI and algorithms have been analyzed according to Canadian law, history and practices.[69] Nor is the LCO aware of equivalent Canadian scholarship considering data discrimination, AI and algorithms and Indigenous peoples or disabled persons.[70]

This is an obvious need that should be proactively and comprehensively addressed before developing or implementing any AI or algorithmic tool in criminal justice (or any other context, for that matter). However, legal analysis alone is not sufficient to ensure AI and algorithmic systems do not discriminate. Equally important is the need to identify legal rules or frameworks that ensure these systems are legally accountable. The LCO will outline several potential strategies in the discussion titled "Algorithmic Accountability" below.

Issue #2: The "Metrics of Fairness"

In addition to the "bias in, bias out" argument, Canadian policymakers seeking to understand risk assessments in either the criminal or civil context must understand the "metrics of fairness." Canadian policymakers should also understand that the word ""discrimination" carries a very different meaning in engineering conversations than it does in public policy."[71]

According to some, an algorithm may be free from racial bias or race-neutral even though it relies on historic data. How can this be? The answer lies in complex questions concerning statistical measures of racial equality and statistical definitions of fairness.

The COMPAS controversy illustrates this issue. ProPublica compared COMPAS risk scores with criminal accused's actual outcomes over two years. ProPublica's analysis showed that a Black defendant who was *not* rearrested within the study period was approximately twice as likely to be classified as high risk as a White defendant who was not rearrested. ProPublica saw these differences as evidence of racial bias and concluded that COMPAS was "biased against blacks"[72] because COMPAS erroneously labeled Black accused as twice as likely to reoffend as White accused.

Northpointe, the company that owns COMPAS, rejected this analysis and argued that ProPublica's data showed that COMPAS was race neutral. Northpointe claimed that Black and White defendants were *actually* rearrested at equal rates. In their view, the tool was fair and neutral because a high-risk COMPAS classification meant the same chance of re-arrest for Black and White accused.

At a technical level, the COMPAS controversy demonstrated how different measures of statistical fairness are crucial in determining whether an algorithm should be considered discriminatory or race-neutral.[73] More importantly, the controversy demonstrated that the burden for statistical errors may not be shared equally. For example, a statistical measure that over-classifies criminal accused as risky may effectively replicate (or worsen) existing patterns of racial disparity. This was part of ProPublica's argument: COMPAS was racist because the burden of statistical errors fell on the same racialized communities who are over-policed/sentenced in first place.

The American discussion on how to reconcile, or at least understand, the relationship between statistical measures of fairness and legal principles is just beginning. This debate is unlikely to be resolved soon or conclusively, as there appears to be up to six different statistical measures that can be used to evaluate pretrial risk assessments.[74]

Unfortunately, American law has not yet provided a clear framework for how to evaluate measures of racial equality in statistics or how to evaluate if an algorithm is biased or not. According to Professor Sandra Mayson, "the law provides no useful guidance about which to prioritize."[75] Professor Aziz Huq goes further, stating that American constitutional law "provides no credible guidance" to assess racial equity in risk assessments.[76]

The LCO is unaware of any equivalent debate or discussion in the Canadian justice context. The LCO notes, however, that this debate will very likely arise in Canada as well, as the choice of statistical measure can have significant implications for personal liberty, racial equity and the credibility of decision-making in the criminal justice system.

In these circumstances, Canadian policymakers, academics, technologists, justice professionals and community representatives would be well-advised to consider these issues before a large scale implementation of AI or algorithms in the criminal justice system.

Issue #3: Data Transparency

One of the most public and significant issues that has arisen in the United States regarding pretrial risk assessments, and many other types of AI or algorithmic tools, is the lack of transparency about data and how these tools work. These criticisms often are part of a larger "black box" critique of AI and algorithms. In the data context, however, transparency concerns generally relate to:

- The inputs or data used by the system, including training data;
- How those inputs/data are weighted by an algorithm; and,
- Whether specific factors (or combinations of factors) are proxies for problematic or impermissible variables, such as race and poverty.[77]

(The LCO will discuss other aspects of transparency in the due process discussion below.)

The challenges presented by the lack of data transparency are two-fold. First, a lack of data transparency makes it difficult for researchers and outside experts to evaluate and audit algorithms in order to test for accuracy and bias. Second, a lack of data transparency makes it harder to bring legal challenges to the use of these tools.

In response to these criticisms, many American organizations have argued forcefully for much greater transparency for all aspects of pretrial risk assessments, including data transparency.

In the international context, there are many proposals and ideas to improve the transparency and accountability of datasets and data variables in AI and algorithms. A small sample of such proposals would include competing approaches suggested by the New York City Automated Decisions Systems Task Force,[78] the AI Now Institute[79] and Harvard University's Berkman Klein Center.[80] Unfortunately, there is no international consensus on how to best ensure data transparency and accountability. Nevertheless, many best practices and law reform proposals have been developed that offer promising starting points. These initiatives will be discussed later in this paper and in the LCO's second and third Issue Papers.

Any introduction of algorithmic risk assessments or tools in the Canadian justice system will inevitably raise questions about data transparency and accountability. Canadian readers will be familiar with data transparency debates surrounding issues like police carding and Sidewalk Labs.[81] These debates mirror American debates on data transparency and risk assessments closely. As a result, there is an urgent need to consider these issues from a Canadian perspective. This effort must be multidisciplinary and involve multiple stakeholders, especially from communities who are most likely to be affected by these technologies.

The LCO will outline several potential strategies to ensure data transparency in AI and algorithmic systems used in criminal proceedings in the discussion titled "Algorithmic Accountability" below.

Issue #4: Data Accuracy, Reliability and Validity

Issues regarding the accuracy, reliability and validity of data are foundational to the success and/or legitimacy of any AI or algorithmic tool used by government, courts or tribunals.

Experience demonstrates that data issues and choices are both consequential and controversial. Canadians considering or developing algorithmic tools must be mindful of the choices, consequences, best practices and requirements inherent in data practices before implementing risk assessments or any other AI or algorithmic tools in the criminal justice system.

It is self-evident that any algorithmic tool will need to rely on data that is accurate, reliable and valid.[82] The following examples demonstrate how relying on data that is *not* accurate, reliable or valid can have significant negative consequences:

- **Timeliness and "zombie predictions."** Keopke and Robinson warn that algorithms or risk assessment tools based on data that does not reflect current practise will likely make "zombie predictions."[83] These are predictions based on data from times and places that are materially different from where the prediction is being made.[84]
- **Legality.** New York City recently announced it was developing a new risk assessment tool. According to reports, this new tool will be based, in part, on data from 2009 to 2015, when the New York City Police were using stop and frisk arrest practices, which were subsequently held unconstitutional.
- **Geography.** Many risk assessment tools in the US (including the PSA) are trained on data that comes from multiple jurisdictions, meaning that *local* predictions may not be accurate.

Questions about data accuracy, reliability and validity are not technical questions best left to developers or statisticians. Issues such as the reasonableness or appropriateness of a dataset, whether or not a dataset is sufficiently reliable, whether new data should be given more weight, and the characteristics selected by developers as most relevant can have important practical and legal consequences. These issues, like so many others, have to be surfaced, debated and decided publicly.

There are many articles, reports and recommendations regarding best practices for data in pretrial risk assessments and similar tools, including several Partnership on AI proposals, addressing requirements on how measure for intended variables,[85] bias,[86] distinct predictions[87] and data retention and reproducibility.[88]

It is not clear whether Canadian criminal databases, such as Ontario's ICON (Integrated Courts Offenses Network) or CPIC (Canadian Police Information Centre), meet these standards.

In any event, Canadian policymakers need to consider these issues if and when they develop similar tools in the Canadian justice system. Most importantly, the LCO believes that governments should work towards transparent data standards applicable to AI and algorithmic systems in collaboration with appropriate stakeholders.

Issue #5: Data Literacy: Risk Scores and Automation Bias

Scoring is endemic in algorithmic tools, be it in the justice system or private sector applications, such as credit scores. The growing use of scoring by public institutions and private corporations has led some to argue that we now live in a "scored society."[89]

As noted above, AI and algorithms typically give an individual a low, medium or high score for some kind of activity (including but not limited to recidivism, criminality, welfare fraud, eligibility for services, likelihood of child abuse, likelihood of defaulting on a loan, etc.).

At first glance, scoring appears to provide simple, easy to understand and usable summaries of complex statistical predictions. The PSA sample report reproduced above is a good example of this. It is important to understand, however, that the determination of what constitutes a low, medium or high score is an explicit policy choice, not a statistical or technical outcome. An algorithm will produce a numerical prediction. Human decision-makers decide whether or how to label a statistical prediction as a low, medium or high risk. Nor does scoring always correspond to our common sense understanding of risk. As a result, risk scoring can be extremely misleading or prejudicial unless users are literate about what the score really means and how it was determined.

The labelling of risk has important consequences: A high risk score in criminal justice is obviously stigmatizing. A person unaware of the mechanics of risk prediction may intuitively think that a high score means an accused is more likely than not to be rearrested. This may be incorrect. Koepke and Robinson report the majority of people who are labeled as high risk by many current American risk assessment tools *will not* be rearrested if released pretrial.[90] In other words, in contrast to popular perceptions, a person with a high risk score is actually more likely to be *successful* than not.

Similarly, many algorithmic tools make predictions about whether something *negative* (such as rearrest, committing a crime, welfare fraud, etc.) is likely to occur. Many critics note that by emphasizing the prospect of *failure* — rather than the more likely prospect of *success* — AI and algorithmic tools effectively stigmatize individuals, particularly low income, racialized or vulnerable communities. In the criminal justice context, critics further suggest that a tool's emphasis on the prospect of failure can erode the presumption of innocence.[91]

Koepke and Robinson also identify an important but subtle example of how a risk score may mistakenly appear to be consistent with legal rules.[92] Current algorithmic risk assessment tools predict the likelihood of re-arrest and fail to appear over a fixed time period, typically two years. If a defendant's period of pretrial release is half as long as a tool's time horizon, then the defendant will be *less likely* to be rearrested than the tool predicts. Unless this aspect of the risk score is disclosed and understood, the risk score and tool may be misleading.

Finally, experience in the United States demonstrates that risk assessment tools often raise issues concerning "human computer interface" or automation bias. The Partnership on AI addresses these issues at some length. The PAI notes that

> *One of the key challenges of statistical decision-making tools is the phenomenon of automation bias, where information presented by a machine is viewed as inherently trustworthy and above skepticism. This can lead humans to over-rely on the accuracy or correctness of automated systems.*[93]

> *…one often overlooked aspect of fairness is the way risk scores are translated for human users. Developers and jurisdictions deploying risk assessment tools must ensure that tools convey their predictions in a way that is straightforward to human users and illustrate how those predictions are made. This means ensuring that interfaces presented to judges, clerks, lawyers, and defendants are clear, easily understandable, and not misleading.*[94]

The impact of data literacy issues can be significant for both individuals and for systemic fairness and bias. The qualifications and subtleties about risk scores can be easily overlooked in individual cases, especially in busy courts. A high-risk score can become a convenient, critical and quick measure of a defendant's suitability for release. Absent procedural protections and a proper understanding of the limits of data scoring, there may be significant risk or prejudice to a defendant's right to fair hearing and to present arguments on their own behalf.

As in many cases with data issues, the choices and options surrounding risk scores appear to be technical but are not. In these circumstances, it is crucial that the data literacy of stakeholders in the Canadian justice system be improved. Everyone involved (including governments, courts, tribunals, judges, counsel, and the individuals and communities affected by these systems) must be able to understand the data and policy choices behind statistical predictions. As in other areas, failure to makes this literacy equal risks further entrenching existing biases and inequality.

B. Due Process, Public Participation and Law Reform

Many of the issues discussed so far highlight the importance of understanding and discussing the significant data issues and choices inherent in the design, development and deployment of these systems. The following sections discuss important legal and policy issues that are inclusive of, but go further than, data questions. More specifically, these sections discuss:

Issue #6 The Distinction Between Predictions, Law and Policy
Issue #7 Best Practices in Risk Assessments
Issue #8 Public Participation
Issue #9 Algorithmic Accountability
Issue #10 The Limits of Litigation

Issue #6: The Distinction Between Predictions, Law and Policy

Canadian policymakers and stakeholders considering AI or algorithms in criminal justice need to understand the distinction between a statistical prediction and the policy decision that renders that prediction into an "action directive." This is another example of how an apparently technical decision can embed significant law and public policy decisions.

Recall that the PSA tool has two basic parts: 1) the PSA risk score and 2) an accompanying "release matrix" that recommends actions (release/detention/conditions/services/etc.) based on that score. In this manner, the PSA risk score is used to *measure* risk and the "release matrix" is used to *manage* risk. As discussed above, the

recommendations included in release matrix are *not* determined by the PSA algorithm. They are, rather, determined by local policymakers who decide what restrictions/conditions/services are suggested by different risk scores.

In these circumstances, it can be argued that the "release matrix" may actually have greater impact on pretrial release than the actual risk scores. Consider some of the issues that must be considered when developing a pretrial release matrix:

- Does the release matrix conform with constitutional law, relevant statutes, judicial decisions and practice guidelines?
- What conditions or recommendations are suggested for high, medium or low risk scores?
- What risk score justifies pretrial release or pretrial detention?
- How should the release matrix account for local services?

Needless to say, these are complicated, contested and consequential questions of law, criminology, social policy and social services. These questions cannot and should not be answered by an algorithm's developers or a small, closed group of decision-makers. Unfortunately, American experience demonstrates that the choices embedded in decision frameworks or release matrices can sometimes lack transparency or appropriate public participation/scrutiny.

The American experience should also remind Canadian policymakers and stakeholders not to focus exclusively on the accuracy or operation of an algorithmic tool. This insight is easily obscured given the complexity of these systems and the issues they raise.

This discussion raises two further thoughts:

First, it is important to remember that the distinction between predictions, law and policy apply in non-criminal areas of algorithmic risk assessment as well. For example, there are many algorithmic tools in use that provide predictions/recommendations regarding child abuse, fraud in government services, and a range of other activities.[95] The analysis here applies to these applications as well: Algorithmic risk scores *measure* risk. Policymakers decide how to *manage* risk.

Second, many of the criticisms of pretrial algorithmic risk assessment in criminal law reflect concerns that these tools will be used to *over*-incarcerate or *overly*-restrict criminal accused, particularly for racialized accused. Once we appreciate the distinction between predictions and policy, however, we can begin to think about redirecting these tools positively or constructively. This idea will be discussed further in the LCO's Conclusion, "Rethinking Risk Assessments".

Issue #7: Best Practices in Risk Assessments

As noted earlier, the introduction and widespread implementation of pretrial risk assessments in the US has spurred an extraordinary outpouring of research, policy-development, community organizing, reassessment and reflection. From a Canadian perspective, it appears the debate and issues have shifted enormously in a very short period.

One of the major advances in this period has been the development of a wide number of best practices and reform proposals. For example, many organizations, notwithstanding their opposition to pretrial risk assessments in general, have proposed detailed protocols or "minimum requirements" for the development, deployment and governance of these systems.

An important, consistent theme in these proposals is the need to incorporate the principles of equality, due process, the presumption of liberty, community participation, transparency and accountability into all aspects of pretrial risk assessment.

The 2018 "Shared Statement of Civil Rights Concerns", discussed above, offers a good example of the pretrial risk assessment reform approach.[96] The Statement, signed by more than 100 civil rights, digital justice and community-based organizations, identified six principles to guide their use:

> ***Principle 1***
> *If in use, a pretrial risk assessment instrument must be designed and implemented in ways that reduce and ultimately eliminate unwarranted racial disparities across the criminal justice system*
>
> ***Principle 2***
> *If in use, a pretrial risk assessment instrument must never recommend detention; instead, when a tool does not recommend immediate release, it must recommend a pretrial release hearing that observes rigorous procedural safeguards. Such tools must only be used to significantly increase rates of pretrial release and, where possible, to ascertain and meet the needs of accused persons before trial, in combination with individualized assessments of those persons…*
>
> ***Principle 3***
> *Neither pretrial detention nor conditions of supervision should ever be imposed, except through an individualized, adversarial hearing. The hearing must be held promptly to determine whether the accused person presents a substantial and identifiable risk of flight or (in places where such an inquiry is required by law) specific, credible danger to specifically identified individuals in the community…*
>
> ***Principle 4***
> *If in use, a pretrial risk assessment instrument must be transparent, independently validated, and open to challenge by an accused person's counsel. The design and structure of such tools must be transparent and accessible to the public.*
>
> ***Principle 5***
> *If in use, a pretrial risk assessment instrument must communicate the likelihood of success upon release in clear, concrete terms.*
>
> ***Principle 6***
> *If in use, a pretrial risk assessment instrument must be developed with community input, revalidated regularly by independent data scientists with that input in mind, and subjected to regular, meaningful oversight by the community.*

The Shared Statement also includes detailed specifications for how these principles could be achieved.

The Shared Statement is just one of many best practice models. Another important model is the AI Now's Algorithmic Impact Assessment, which is a sophisticated framework for assessing, evaluating and monitoring AI and algorithmic systems in use by government, including systems in the criminal justice system.[97]

Other helpful proposals and practices have been developed by the Berkman Klein Center for Internet and Society (identifying four phases when adopting a risk assessment tool: development, procurement, implementation, and testing),[98] the Electronic Frontier Foundation (policy guide for judges and judicial officers),[99] the National Legal Aid Defenders Association (NLADA),[100] ethical design standards (with leading examples drawn from the Toronto Declaration on Machine Learning[101] and the work of the Institute of Electrical and Electronics Engineers (IEEE)),[102] and initiatives from the Government of Canada (discussed below) and New Zealand.[103]

In the LCO's view, these best practices are an important benchmark against which to evaluate the design, development, deployment and governance of pretrial risk assessments or other AI or algorithmic tools that may be considered in the Canadian justice system.

Importantly, best practices development has not been limited to the legal community. The AI technical community has also developed best practices for the responsible development and deployment of algorithmic risk assessments. For example, the PAI has developed principles agreed upon by a broad cross-section of the AI community, including requirements addressing:

- Accuracy, validity and bias of risk assessment tools;
- Human-computer interface issues; and
- Governance, transparency and accountability issues.

The PAI's ten "minimum requirements for the responsible deployment of criminal risk assessment tools" include:

Requirement 1: Training datasets must measure the intended variables[104]
Requirement 2: Bias in statistical models must be measured and mitigated[105]
Requirement 3: Tools must not conflate multiple distinct predictions[106]
Requirement 4: Predictions and how they are made must be easily interpretable[107]
Requirement 5: Tools should produce confidence estimates for their predictions[108]
Requirement 6: Users of risk assessment tools must attend trainings on the nature and limitations of the tools[109]
Requirement 7: Policymakers must ensure that public policy goals are appropriately reflected in these tools[110]
Requirement 8: Tool designs, architectures, and training data must be open to research, review, and criticism[111]
Requirement 9: Tools must support data retention and reproducibility to enable meaningful contestation and challenges[112]
Requirement 10: Jurisdictions must take responsibility for the post-deployment evaluation, monitoring, and auditing of these tools[113]

The LCO is not offering the best practices or "minimum requirements" identified here as ready-made solutions for the Canadian justice system. These practices and standards were developed in the United States in response to American issues. That said, the LCO certainly agrees with many of these best practices in principle, particularly those that address racial bias, due process and the need for broad participation in algorithmic development and oversight.

More prosaically, the LCO wants to emphasize that there is no reason for Canadian policymakers and stakeholders to believe they are starting from scratch. There are *many* available resources to start these discussions.

Issue #8: Public Participation

The American experience demonstrates the need for broad participation in the design, development, deployment and governance of AI and algorithmic systems in the justice system. This insight is confirmed by the LCO's earlier work in this area.[114]

There are many competing proposals and ideas about how to improve public participation in data collection, systems development and oversight. A particularly high-profile, recent example of the controversies and issues surrounding public participation concerns the New York City AI Task Force.[115] This Task Force was set up in 2018 by New York City to provide recommendations on broad range of topics related to the use of AI and algorithms by New York City agencies. In so doing, New York City became the first US jurisdiction to create a task force to come up with recommendations for government use of AI and algorithms.

The Task Force report includes a comprehensive list of recommendations but did not reach a consensus on many issues. Shortly after Task Force's report was published, a large group of community advocates and NGOs published a "shadow report" which included blistering criticisms of the Task Force's recommendations and public process.[116]

The New York City example illustrates the importance of process and participation when major reforms are being considered. These issues will come to the forefront in Ontario and Canada when AI and algorithmic tools are more widely introduced here. Once again, the debates about public participation, AI and algorithms in the American criminal justice system echo debates in Canada regarding police carding and racial profiling. Canadians would be well advised to take steps to address the need for meaningful public participation thoughtfully and proactively.

The 2018 "Shared Statement of Civil Rights Concerns" again offers a good example of how and why community input and public participation is so important. Principle 6 of the Statement reads:

> *If in use, a pretrial risk assessment instrument must be developed with community input, revalidated regularly by independent data scientists with that input in mind, and subjected to regular, meaningful oversight by the community.* [117]

This principle underscores the significance of risk assessment tools being developed and implemented with input from diverse members of the local community in which the tool is being employed.

Taking a local and jurisdictional approach to community engagement was also emphasized in a recent panel discussion by Cherise Fanno Burdeen, the Executive Partner of the Pretrial Justice Institute (PJI).[118] She suggests that each jurisdiction ought to think broadly and critically about the *impact* of reforms, instead of solely focussing on the *intent* of the tool. Ms. Fanno Burdeen notes that even the most carefully prepared tools can fail to be implemented fairly when there is a lack of communication and engagement with the local community.[119]

In a recent webcast, PJI members candidly discussed how intentional community engagement and having discussions with individuals from diverse racial groups became the catalyst for their strong opposition to risk assessments.[120] The term "colorblind racism" was raised during the webcast by Dr. Angela Hattery, co-author of the book, *Policing Black Bodies*.[121] According to Dr. Hattery, colorblind racism refers to what takes places

> *when one group of people in power create a system without the voices of anyone else; they often don't ask the questions that produce the outcomes that would actually result in racial equity, and by not asking those questions, they produce policies that not only extend racial injustice, but make it very easy for white people to buy-in to them.*[122]

Webcast participants further explained that effective community engagement from the onset, not just at the evaluation stage, is a way to prevent colorblind racism from impacting the implementation and regulation of risk assessment tools.

Participants also emphasized that when stakeholders and policymakers consult with those most impacted by pretrial risk assessments—low-income communities and communities of color—they ought to lay aside the need to *intellectualize* the dialogue. Instead, they should simply listen with compassion and empathy, embracing a sort of "testimonial justice", that is grounded in hearing people's lived experiences and recognizing that experiences differ across racial lines.[123] Proposed ways of having these dialogues involved explicitly naming the issues and thinking critically about how white supremacy has affected systems, communities, and individuals; embracing uncomfortable conversations; examining our understanding of culture and history; as well as each stakeholder being grounded in their own *personal commitment* to think about what it means to contribute to racial justice.

The LCO strongly supports broad participation in the design, development, deployment and governance of AI and algorithmic systems in the Canadian justice system. This participation must include technologists, policymakers, law makers, and, crucially, the communities who are likely to be most affected by this technology. Unequal access to information and participation in decision-making about data and technology can significantly worsen existing biases and inequality.

Issue #9; Algorithmic Accountability

The American and international "digital rights", legal, and technology communities have been focussed on overlapping questions regarding the transparency, accountability and legal protections regarding AI and algorithmic systems for several years.[124] The American debate on these topics is extensive and evolving.

The discussion below outlines some of the specific strategies and proposals that have developed in the United States to improve algorithmic accountability, including proposals respecting:

- Technological due process;
- Algorithmic transparency;
- Bias and equity;
- Public participation and data literacy;
- Due process, evidence, remedies and the right to counsel; and,
- Technical accountability.

Collectively, these proposals represent a robust regime for addressing legal accountability regarding data, transparency, bias and due process concerns in AI and algorithms in criminal justice.

Technological Due Process

Many of the emerging proposals to ensure AI and algorithmic accountability are based on "technological due process" principles and priorities. This concept, based on a seminal 2008 article by Professor Danielle Keats Citron, suggests that AI and algorithms require deeper analysis of due process and regulatory issues than traditional legal models may suggest.[125] This concept is grounded in a belief that

> *the accountability mechanisms and legal standards that govern decision processes have not kept pace with technology. The tools currently available to policymakers, legislators, and courts were developed primarily to oversee human decisionmakers…our current frameworks are not well-adapted for situations in which a potentially incorrect, unjustified, or unfair outcome emerges from a computer.* [126]

The key elements of technological due process are transparency, accuracy, accountability, participation, and fairness. This includes standards over scoring systems; the right to inspect, correct, and dispute inaccurate data; the right to know the source of data; the right to public disclosure and oversight of data; and the creation of an audit trail for all algorithmic decisions or recommendations. [127]

Algorithmic Transparency

Many of the ideas regarding how to ensure AI and algorithms are legally accountable are organized within a broad range of issues sometimes called "algorithmic transparency." Algorithmic transparency is intended to remedy, or at least mitigate, concerns about the opacity of algorithmic systems and decision-making. As noted by Deven Desai and Joshua Kroll,

> *Transparency has been proposed as a solution to mitigating possible undesired outcomes from automated decision-making… A related fear is that the human designer of a program could have bad intent and seek to discriminate, suppress speech, or engage in some other prohibited act. Transparency in this context is the claim that someone "ought to be able to 'look under the hood' of highly advanced technologies like . . . algorithms" as a way to police such behavior.*[128]

Advocates identify various methods/strategies for achieving algorithmic transparency, including 1) disclosure and 2) auditing, impact assessments, evaluation and testing.

Disclosure

In order to understand the use and impact automated decision-making, one must be aware of its existence and use. As a result, *disclosure* of AI and algorithms has become a high-priority justice issue. Disclosure includes both the existence of a system and a broad range of tools and processes used by the system. Data disclosure is a notable high priority. Legally, some of the issues that have arisen regarding disclosure of AI and algorithmic systems include:

- What is the definition of AI, algorithms or automated decisions that would require disclosure?
- When does the disclosure requirement engage? Upon implementation, development, proposal or procurement?
- Does the disclosure requirement apply to new systems, existing systems or both?
- Who has the responsibility to disclose?

There are many different responses to these issues. The first issue (the definition of AI, algorithms or automated decision-making) has proven particularly vexing.[129] Equally complicated are questions about the extent of disclosure. The best practices discussed above generally would require any or all of the following information to be disclosed:

- Training data;
- Source code;
- Complete description of design and testing policies and criteria;
- List of factors that tools uses and how they are weighted;
- Thresholds and data used to determine labels for score;
- Outcome data used to validate tools;
- Definitions of what instrument forecasts and for what time period; and,
- Evaluation and validation criteria and results.

Finally, a particularly important disclosure issue relates to proprietary AI or algorithmic tools, like COMPAS. Advocates strongly urge governments not to deploy proprietary tools that may rely on trade secret claims to prevent disclosure and transparency.[130]

Auditing, Impact Assessments, Evaluation and Testing

In the US, there has been a strong emphasis on ensuring that risk assessments and other AI or algorithmic systems are subject to extensive auditing, evaluation and testing. In the words of the PAI,

> *Jurisdictions must periodically publish an independent review, algorithmic impact assessment, or audit of all risk assessment tools they use to verify that the requirements listed in this report have been met.74 Subsequent audits will need to examine the outcomes and operation of the system on a regular basis.*
>
> *To ensure transparency and accountability, an independent outside body (such as a review board) must be responsible for overseeing the audit. This body should be comprised of legal, technical, and statistical experts, currently and formerly incarcerated individuals, public defenders, public prosecutors, judges, and civil rights organizations, among others. These audits and their methodology must be open to public review and comment. To mitigate privacy risks, published versions of these audits should be redacted and sufficiently blinded to prevent de-anonymization.*[131]

AI Now's Algorithmic Impact Assessment is a particularly helpful critical framework for assessing, evaluating and monitoring AI and algorithmic systems in use by government.[132]

Bias and Equality

In addition to transparency issues, American policymakers and advocates are considering how to ensure AI and algorithmic systems comply with American constitutional and human rights principles, specifically the Equal Protection Clause of the 14th Amendment.[133]

Some scholars are pessimistic that American constitutional law will offer a useful framework for assessing bias and equality issues in criminal algorithms, including one who suggests "[i]f there is a lesson here, indeed, it is about the woeful inadequacy of our constitutional equality norms for the contemporary world."[134]

In the face of these challenges – or perhaps because of them – many stakeholders, scholars and advocates offer a wide range of policy-based or regulatory initiatives to ensure AI and algorithms do not discriminate, including

- Improved transparency, testing, auditing and evaluation of algorithmic systems;
- Improved the collection, accuracy, reliability and disclosure of algorithmic data;
- Statutory or regulatory prohibitions on the use of certain factors in algorithmic decision-making, including race or potential proxy factors, such as education, employment, geography;
- Statutory or regulatory provisions stating that no racial group should bear the undue burden of errors made by an algorithmic instrument;
- Refocussing algorithmic tools towards eliminating racial disparities;
- Ensuring greater community participation in design, development, implementation and oversight of AI and algorithms, particularly from racialized communities; and,
- Mandatory education on the racial effects of algorithmic risk assessments.

It should be noted that for many racial justice advocates, these initiatives, while important, are supplemental to addressing bias and discrimination as part of larger criminal justice strategy.

Public Participation and Data Literacy

As noted earlier in this report, there has been considerable discussion within the US regarding the need for public participation, training and data literacy to ensure AI and algorithmic transparency and accountability, both legally and to the community at large. Some of the notable proposals suggested for achieving accountability through these means include:

- Ensuring tools are designed with input from members of the local community, including, but not limited to:
  - Independent data scientists;
  - Criminal accused;
  - Criminal justice system participants and policymakers, including public defenders, judges, and district attorneys; and
  - Community groups focused on racial justice.
- Establishing community advisory boards; and,
- Regular training on tools.

Due Process, Evidence, Remedies and the Right to Counsel

There have been extensive American debates about how to ensure algorithmic risk assessment systems comply with due process rights regarding fairness, notice, hearing, reasons, appeals and remedies. Legally, due process issues raise a host of complex questions, including but not limited to:

- What does due process require for AI and algorithms in criminal proceedings?
- What is the nature of notice or disclosure?
- How to ensure a fair hearing and respect for constitutional rights?
- What rules of evidence apply?
- How can or should AI or algorithmic decisions/recommendations be explained?[135]
- How to ensure the right to challenge AI or algorithmic evidence?
- How to ensure the right to an individualized hearing and decision?

- What remedies are available?
- Is there a right to appeal and on what grounds?
- Is there a right to counsel?[136]

American constitutional law has also been inconclusive on due process issues. It appears there have been a limited number of constitutional challenges to algorithm risk assessments in criminal proceedings based on the US Constitution's 5th Amendment and Due Process Clause of 14th Amendment.[137]

The leading American due process case is a 2016 Wisconsin decision, *State v. Loomis*.[138] In this case, the Wisconsin Supreme Court held that a trial court's use of the COMPAS risk assessment tool when sentencing Loomis did not violate his due process rights, even though the methodology used to produce the assessment was not disclosed to either the court or the defendant.[139]

The *Loomis* decision was an early but incomplete analysis of algorithmic due process in criminal proceedings. As a result, many American scholars and advocates offer a wide range of policy-based or regulatory initiatives to ensure procedural safeguards if and when algorithmic tools are used in their criminal justice system.

Many of these proposals are directed at ensuring the oversight role of courts while placing explicit restrictions on how and when the tools used. Other proposals are directed at ensuring an effective right to challenge the operation or use of a tool in individual cases.[140] The range of these proposals include

- Explicit prohibitions of algorithms recommending detention;
- Explicit requirements that tools be applied in manner consistent with the presumption of innocence and the right to an individualized hearing;
- Explicit directions as to how tools may be used by decision-makers;
- Explicit recognition of right to inspect and cross-examine risk assessment tools and recommendations in individual cases, including right to introduce evidence that contradict algorithmic recommendations;
- Explicit rules governing how scoring is to be developed and framed;
- Expedited and broad appellate review of decisions based, in part, on algorithmic risk assessments;
- Modification of rules of practice or evidence to support procedural safeguards;
- Mandatory training for all justice system professionals; and,
- Ensuring defence counsel have time, training and resources to challenge risk assessment recommendations.

Technical Accountability

Finally, some computer scientists believe AI or algorithmic accountability can be enhanced through a concept called "technical accountability". "Technical accountability" is distinct from "technological due process" and is grounded in computer science, not law. It originates in a critique of the "algorithmic transparency" model described above. From this perspective, the essential problem with the algorithmic transparency model is that "misunderstands the nature of computer systems"[141] and "will not only fail to deliver critics' desired results but also may create the illusion of clarity in cases where clarity is not possible."[142]

In the technical accountability model, AI and algorithmic systems must be designed from the very outset with oversight and accountability in mind. The goal should not be to achieve (or necessarily strive for) perfect transparency, but rather to have systems generate reliable evidence to verify they function correctly. This could include complex technical approaches to ensure computer systems achieve procedural regularity[143] and to assure fidelity to substantive policy choices[144]

Supporters of this view do not suggest that technical accountability is a substitute for legal accountability. Rather, they believe that technical accountability is a necessary precondition for legal accountability.

Summary

The LCO is not offering the proposals identified above as ready-made solutions for the Canadian criminal justice system. These initiatives are obviously grounded in American experience and law. In principle, however, the LCO supports the American concept of "technological due process." The LCO also agrees in principle with many of the specific proposals, particularly those that address data issues, disclosure, racial bias, due process and the need for broad participation in algorithmic development and oversight.

Issue #10: The Limits of Litigation

It is important to note there has been very little litigation or caselaw considering data discrimination issues in Canada within the context of data-based AI and algorithms.[145] American litigation has been more frequent but is inconclusive.

An important legal milestone in the American debate about risk assessments is the 2016 Wisconsin state court decision discussed above, *State v. Loomis*.[146] *Loomis* remains the leading American judicial decision considering the use and limitations of risk assessments in the US criminal justice system.[147]

In *Loomis*, a defendant from La Crosse County, Wisconsin challenged the use of the COMPAS risk assessment tool. COMPAS had been used by Wisconsin since 2006 and was not subject to any legislative provisions, regulations or dedicated rules of practice. After Loomis pled guilty, the court requested a presentence investigation report, which included a risk score calculated using COMPAS. COMPAS identified Loomis as high risk for three types of recidivism measured by the program: pre-trial recidivism, general recidivism, and violent recidivism.

Loomis ultimately received a six-year prison sentence. The sentencing judge told Loomis "The risk assessment tools that have been utilized suggest that you're extremely high risk to reoffend." Loomis challenged his sentence on due process grounds, arguing that the judge's use of the risk assessment score violated his constitutional right to a fair trial. More specifically, he argued that:

1. His right to be sentenced based on accurate information was violated because the proprietary nature of COMPAS meant could not review it;
2. His right to an individualized sentence was violated because COMPAS data relies on group information; and,
3. COMPAS improperly used "gendered assessments."[148]

Northpointe (now Equivant), the company that owned COMPAS, refused to provide its proprietary software for the court's or Loomis' review. The Wisconsin Supreme Court upheld this refusal. The court noted, however, that even though the Loomis was prevented from seeing how his score was calculated, he was still able to verify most of the inputs into the algorithm because they were based on public records and a questionnaire that he filled out.

The Wisconsin Supreme Court further stated that Loomis' due process rights were not violated because COMPAS did not appear to be the *sole* basis for the trial judge's sentence.

Notwithstanding these rulings, the court also held that COMPAS had to be used cautiously, and put several limits on the use of COMPAS in future state court proceedings:

First, the court ruled that COMPAS could only be used to address treatment needs and the risk of recidivism, not for sentencing purposes.

Second, the court ruled that if and when a COMPAS report was included in a presentence investigation report, the presentence report had to include a five-part written warning, specifying that:

> *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.*
>
> *Some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having higher rates 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 renormed 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.*

Loomis has been both complemented and criticized.[149] The Wisconsin Supreme Court's elucidation of safeguards has been praised as an important limitation on the use of AI and algorithms in criminal justice. Criticisms have centred on the court's refusal to order disclosure of COMPAS's proprietary software to allow Loomis to effectively challenge (or even understand) his COMPAS assessment; the court's lack of clarity about how judges are to use risk scores; criticisms of the court's product warning approach; and the difficulty of assessing whether the risk score was a *determinative* factor in the judge's decision-making process. Other comments on *Loomis* centre on the fact that Loomis did not raise an equal protection claim, meaning that the court did not evaluate COMPAS against a 14th Amendment Equal Protection challenge.

Loomis obviously has no precedential value in Canada. Given the differences between American and Canadian constitutional law, it is arguable if the case even has a modest persuasive value. The LCO is highlighting *Loomis*, however, to illustrate why litigation alone is not likely to be an effective means of providing transparency, procedural safeguards and oversight of algorithmic tools in criminal justice.

Consider just some of the complex statistical, technical and policy issues that could (or should) have been litigated in *Loomis* or equivalent cases:

- Is the historic data used to train the COMPAS tool biased, accurate, reliable and valid?
- Are COMPAS risk factors and scores weighed and calculated appropriately?
- Which communities bear the burden of statistical errors?
- Are the confidence estimates for COMPAS predictions appropriate?
- Are COMPAS predictions validated appropriately?
- Does COMPAS use factors such as education or employment as impermissible statistical proxies for race or gender?

Systemic, technical and normative questions like these (and others) are best addressed through public policy and multidisciplinary participation, not litigation.

Litigation obviously has an important role in regulating AI and algorithms in the criminal justice system. Many issues will always be best addressed in open court with the benefit of an evidential record and high-quality and experienced counsel. To this end, readers should note there is an emerging American community of practice devoted to "litigating AI."[150] A helpful new Canadian study of this issue is expected later in 2020.[151]

Litigation, while obviously necessary to address specific cases, is insufficient to address the *systemic* statistical, technical, policy and legal issues that have been addressed in this report so far. As a result, the LCO believes the most effective response to these issues must ultimately be grounded in some kind of systemic regulation or statutory framework, in addition to litigation, best practices, algorithmic audits, evaluations, etc.

In the LCO's view, comprehensive regulation is justified on access to justice principles as well:

Loomis was a comparatively simple, State-court criminal proceeding *in which Loomis had already plead guilty*. The COMPAS issue arose at sentencing. It is inconceivable that Loomis or any other criminal defendant (particularly one represented by a public defender/legal aid or self-represented) would be in a position to mount an effective challenge to the complex statistical, technical and legal issues raised by COMPAS. This analysis is equally true in cases where COMPAS, the PSA or any other algorithmic risk assessment tool is used in bail proceedings.

The importance of access to justice (or equal justice, in the American vocabulary) in discussions about AI or algorithmic accountability in the criminal justice system is perhaps under-appreciated. In the absence of comprehensive regulation and improved legal aid funding, a criminal accused confronting an algorithmic risk assessment faces even more difficulty in presenting a full answer and defence to the charges against them. Unfortunately, these additional hurdles may actually compound the over-representation of low-income and racialized communities already present in the criminal justice system.

X. COMPREHENSIVE LAW REFORM

The LCO has tried to demonstrate the urgency and importance of addressing the many novel and complex legal and policy issues raised by the prospect of AI and algorithms in Canadian criminal proceedings. The LCO has argued that our existing legal analysis and protections governing the disclosure, accountability, equality and due process requirements for these systems are likely inadequate. There is a need to ensure Canadian legal standards and rules keep pace with technology.

Part of the issue in the US, it would seem, is the fact that many, perhaps most, of these systems were implemented *prior* to a broad public discussion about how they should be regulated. This approach has caused many problems that have been subsequently revealed through litigation, press reports and community and academic scrutiny.

Questions regarding disclosure, accountability, equality and due process will surface quickly, repeatedly and urgently in Canada if and when these systems are used in the Canadian criminal justice system. Fortunately, Canadians have an opportunity to approach regulatory issues thoughtfully and deliberately. Canadians are also fortunate that there are many good examples and precedents to help guide and inform our discussions. The American experience offers an impressive body of legal analysis, academic research, operational experience, community evaluation, best practices and lessons learned.

Given the complexity of the issues, and the fundamental rights at stake, it is clear to the LCO that the systemic legal issues raised by this technology cannot be addressed through individual litigation, best practices or piecemeal legislation. Comprehensive law reform is required. There are many potential legislative or regulatory responses, but the choices and options between these responses are complex and consequential. What's important is that the Canadian legal system proactively address a comprehensive series of issues and options *prior* to widespread implementation of these systems.

Government of Canada AI Directive and Government of Ontario Initiatives

Canada's most important effort to regulate governmental use of AI to make decisions that impact individuals is Government of Canada's *Directive on Automated Decision-Making* (the "Directive").[152]

The Directive states that its objective is to:

> *ensure that Automated Decision Systems are deployed in a manner that reduces risks to Canadians and federal institutions, and leads to more efficient, accurate, consistent, and interpretable decisions made pursuant to Canadian Law.*[153]

The Directive states that its expected results are that

> *Decisions made by federal government departments are data-driven, responsible, and complies with procedural fairness and due process requirements.*
>
> *Impacts of algorithms on administrative decisions are assessed and negative outcomes are reduced, when encountered.*
>
> *Data and information on the use of Automated Decision Systems in federal institutions are made available to the public, where appropriate.*[154]

The Directive came into force on April 1, 2019 and was created following an important White Paper and limited public consultations.[155] The Directive applies "systems, tools, or statistical models used to recommend or make an administrative decision about a client of a federal government department."[156]

The Directive's requirements are linked to "core administrative law principles such as transparency, accountability, legality, and procedural fairness"[157] and are divided into five categories or stages of use of AI:

- Performing an Impact Assessment[158]
- Transparency[159]
- Quality Assurance[160]
- Recourse[161]
- Reporting[162]

The Directive requires an algorithmic impact assessment for every automated decision-making system (ADM), including the impact on rights of individuals or communities. The algorithmic impact assessment must be released publicly.[163]

Based on the impact assessment, an ADM system will be designated as impact level I, II, III or IV. Depending on the impact level, the Directive requires greater or lesser levels of

- Notice before ADM decisions and explanations after ADM decisions;
- Peer review;
- Employee training; and,
- Human intervention.[164]

The Directive establishes some requirements that apply to all ADM systems, regardless of their impact level,[165] including:

- Access, diligence, testing and auditability requirements for licensed software;
- Release of custom source code that is owned by the Government of Canada;
- Quality assurance and monitoring requirements, including:
  - Testing "before launching into production…[to ensure ADM systems] are "tested for unintended data biases and other factors that may unfairly impact outcomes."[166];
  - Monitoring "outcomes of ADM Systems to safeguard against unintentional outcomes and verify compliance with institutional and program legislation."[167];
- Validating the quality of data collected and used;
- Consultations with government legal services to ensure the use of the ADM complies with applicable laws;
- Providing individuals with "recourse options that are available to challenge the administrative decision"[168]; and,
- Reporting information on effectiveness and efficiency.

The Directive is an important initiative. Even in its early stages, the Directive is a major step forward in supporting algorithmic accountability at the federal level. Experience will tell whether or not this effort is successful and/or how the Directive is implemented.

At this point, it is not clear if or how the Directive would apply to the criminal justice system, as the Directive refers to "administrative decisions" at several points.[169] Nor does the Directive explicitly include (or exempt) automated decision-making tools used in federal criminal proceedings.

The Directive is the most advanced and sophisticated effort to regulate AI and algorithmic accountability in government decision-making in Canada. The LCO is aware of similar, but early, initiatives at the provincial level, including the Government of Ontario. The LCO will comment upon those initiatives when more information is publicly available. The LCO will consider the Directive and similar national frameworks (such as the proposed US *Algorithmic Accountability Act*[170]) in more detail in our second and third Issue Papers.

The LCO believes that both federal and provincial government initiatives are important, proactive steps to address some of the accountability issues raised by AI and algorithms in government decision-making. The LCO does not, however, believe these initiatives alone will address and protect the important disclosure, accountability, equality and due process issues raised in the criminal justice system. In the LCO's view, more comprehensive regulation is needed, the elements of which are discussed below.

Mixed Model of AI and Algorithmic Regulation

The LCO believes that governance of these systems can be achieved through what is sometimes called a "smart mix" or "mixed model" of AI and algorithmic regulation.[171] This model is premised on the belief that no one single statute, rule or practice is likely to be sufficient to governing AI and algorithmic systems.

In the criminal context, a comprehensive regulatory regime should include both federal and provincial initiatives, consistent with each government's jurisdiction in criminal justice. These initiatives should likely address both 1) the systemic regulation of AI and algorithmic systems used by government, and 2) specific legislation or rules addressing the use of AI and algorithms in criminal proceedings, including the oversight role of courts and due process protections. Accordingly, the LCO suggests that a comprehensive regime to ensure algorithmic accountability in the Canadian criminal justice system should likely include the following elements:

- National standards or regulations governing the development, disclosure and use of AI and algorithmic systems used by the federal government;
- Provincial standards or regulations governing the development, disclosure and use of AI and algorithmic data and systems used by the provincial government;
- Amendments to federal and provincial evidence legislation;
- Criminal justice-specific statutory or regulatory provisions prescribing the parameters of use for AI and algorithmic tools;
- Criminal justice-specific disclosure and due process regulations, legislation or practice directions;
- Federal and provincial standards or regulations guaranteeing public participation in the design, development, implementation and oversight of these systems;
- Training and education for criminal justice system participants; and,
- Ethical design standards.

Identifying the *types* of legal reforms needed is one thing; identifying the *content* of those reforms is quite another. As a starting point, the LCO suggests a valuable approach might be for Canadian policymakers and stakeholders to consider the proposals and strategies identified in the "Algorithmic Accountability" discussion above, including proposals respecting:

- Technological due process;
- Algorithmic transparency;
- Bias and equity;
- Public participation and data literacy; and,
- Due process, evidence, remedies and the right to counsel.

The first step in this process would be to establish a broadly-based, participatory strategy to address these issues.

XI. CONCLUSION: RETHINKING RISK ASSESSMENTS

Canadian governments considering AI or algorithmic tools in our criminal justice system cannot escape the complex issues, hard choices and public engagement described in this paper. Implementing any tool with potentially such an extraordinary impact on individual rights and the justice system must be approached thoughtfully, incrementally and transparently with early and appropriate engagement with a broad range of stakeholders, particularly the communities who are most likely to be affected by this technology.

The rush to adopt algorithmic pretrial risk assessment tools in the United States was understandable: Algorithmic risk assessments were considered objective, consistent, evidence-based tools that would help transform the arbitrary, opaque, and often-racist pretrial decision-making of individual judges, prosecutors and justice systems. In short, these tools appeared to offer a scientific, objective, modern and transparent way to address long-standing, complex, entrenched issues of structural racism in the American criminal justice system.

As noted, the expansion of these tools did not go smoothly. In the space of a few short years, there has been an extraordinary backlash against the use of these systems, including by many of the same organizations and stakeholders who enthusiastically supported these systems in the first place.

For Canadians, both the outcome and the process of this debate are important. Equally significant are the lessons we can learn from this experience about the use of AI and algorithms in the criminal justice system.

From a Canadian perspective, the American debate suggests policymakers and stakeholders in this country need to begin by addressing at least four threshold questions:

1. Should there be a moratorium on algorithmic risk assessments or similar tools in the Canadian criminal justice system?
2. What is the potential for algorithmic risk assessments in the Canadian criminal justice system?
3. Is there a future where algorithmic risk assessments are used as part of a comprehensive strategy to advance equity, access to justice and systemic efficiency?
4. What is the path forward?

Many advocates in the United States would answer the first question affirmatively. This belief is based on the many significant and legitimate criticisms of these systems as presently deployed. In this view, the emphasis placed on algorithmic risk assessments is either intrinsically wrong (due to insurmountable racial bias or due process concerns) or strategically misguided (because risk assessments displace other, more effective reforms).

This is a compelling argument, but perhaps not the final word.

Based on our analysis so far, the LCO would certainly agree that widely deploying algorithmic risk assessments in the Canadian justice system at this time would be a mistake. The LCO's research demonstrates the risks of adopting unproven and under-evaluated technologies too quickly. The US experience also demonstrates the hazards of relying too heavily on a simple, technological solution (algorithmic risk assessments) to address complex, long-standing and structural problems (bail reform, racism in the American justice system).

At the same time, many American NGOs, legal groups and academics have begun to rethink and refocus algorithmic risk assessments in significant ways.

For example, there has been serious consideration of how to focus algorithmic risk assessments more effectively as a tool to promote pretrial or racial justice. As a result, many recent American proposals for algorithmic risk assessments move away from using risk assessments to assess liberty issues, such whether to detain an accused before trial. According to some, pretrial risk assessments have the potential to be used more strategically or

progressively, including using such tools to facilitate releasing criminal accused from police station without a need for a bail hearing. Algorithmic risk assessments or similar tools might be used to more effectively identify criminogenic needs, to identify biased decision-making, to identify community supports or to support evidence-based recommendations about bail conditions.

These proposals reconsider the objective or goal of algorithmic risk assessments. Other proposals reconsider how to develop, implement, monitor and evaluate algorithmic risk assessments. These proposals (which have been considered at length in this report) reconsider the procedural and legal regimes necessary to improve the operation and accountability of algorithmic risk assessments, consistent with the technological due process model described above. Taken together, comprehensive reforms of this nature, *if implemented properly*, may have the potential to significantly improve the design, development, deployment and outcome of these systems.

There is also recent evidence that algorithmic risk assessments, *when combined with other strategies*, may have a positive effect on criminal justice system reforms: In 2017, New Jersey implemented comprehensive reforms to replace monetary bail with a system based on risk. These risks were assessed using the PSA and a New Jersey-specific decision-making framework. In this program, New Jersey used the PSA at two stages in criminal proceedings:

- At the time of arrest, when a police officer must decide whether to detain an accused pending a hearing or to release the accused with a notice of appearance; and,
- At the accused's first appearance, when judges set release conditions for defendants who were detained at arrest.

The first evaluation of these reforms was very positive, reporting:[172]

- Fewer arrest events after implementation, including fewer arrest events for the least serious charges.
- Police officers appear to be releasing accused at the police station more frequently.
- Pretrial release conditions imposed on defendants changed dramatically, including larger proportions of defendants being released without conditions.
- Significantly reduced the length of time defendants spend in jail in the month following arrest.

Importantly, the implementation of the PSA was just one part of a comprehensive criminal justice reform package that largely eliminated monetary bail; established the possibility of pretrial detention without bail; established a pretrial monitoring program; and instituted speedy-trial laws that impose time limits for case processing.

The report does not discuss the effect of the PSA or decision-making framework on these outcomes at length. Nor does the report speculate if these outcomes could have been achieved without implementing the PSA. One can hope that these issues will be addressed in subsequent evaluations. As a result, at this point the New Jersey example only represents a *potential* demonstration of how the PSA and risk assessments can be used thoughtfully and strategically to advance comprehensive criminal justice reform.

Finally, some commentators believe that algorithmic risk assessments have the potential to be more transformative than examples like New Jersey would suggest. This is because, in the words of Professor Sandra Mayson, the long-term "accountability prospects" of algorithmic prediction "are far better for algorithmic prediction than for subjective prediction."[173] In this view, algorithmic systems have the potential to "eliminate the variability, indeterminacy, and apparent randomness—indeed, the subjectivity—of human prediction that has long pervaded criminal justice. They bring uniformity, transparency, and accountability to the task."[174]

Mayson's support for this potential is qualified, however, as she notes

> *[t]his is not to overstate the case for algorithms. The evidence for the superior accuracy of actuarial over subjective prediction is not watertight; a great deal depends on the algorithm at issue and the details of its use.*[175]

Nevertheless, Mayson states that even the racialized history of algorithmic data could potentially be turned from a weakness into a strength

> *…because predictive algorithms transparently reflect inequality in the data from which they are built, they can also be deployed in reverse: as diagnostic tools to identify sites and causes of racial disparity in criminal justice.*[176]
>
> *If algorithms targeted the disadvantaged for support rather than for further disadvantage, their effects in the world would be very different.*[177]

From a racial equity and access to justice perspective, this is an attractive vision. Unfortunately, achieving this vision will not be easy, nor is it inevitable.

Where, then, should we go from here? Is there potential for algorithmic risk assessments in the Canadian criminal justice system? Can these tools be used as part of a comprehensive strategy to advance equity, access to justice and systemic efficiency? If so, what is the path forward?

In response to these questions, the LCO offers a modest recommendation: This paper has identified a series of issues and options that should be addressed prior to the widespread implementation of any AI or algorithmic system in the Canadian criminal justice system. In these circumstances, perhaps the first step is for policymakers and stakeholders to collectively agree to address these issues and on an appropriate process for doing so.

XII. HOW TO GET INVOLVED

The LCO believes that successful law reform depends on broad and accessible consultations with individuals, communities and organizations across Ontario. As a result, the LCO is seeking comments and advice on this report. There are many ways to get involved:

- Learn about the project on the LCO website (www.lco-cdo.org);
- Contact us to ask about the project; or,
- Provide written submissions or comments on this report.

The LCO can be contacted at:

Law Commission of Ontario
2032 Ignat Kaneff Building
Osgoode Hall Law School, York University
4700 Keele Street
Toronto, Ontario, Canada
M3J 1P3

Email: LawCommission@lco-cdo.org
Web: www.lco-cdo.org
Twitter: @LCO_CDO
Tel: (416) 650-8406
Toll-free: 1 (866) 950-8406

ENDNOTES

1 Partnership on AI (PAI), *Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System*, (April 2019) [PAI] at 7, online: **https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/**. See generally, **www.partnershiponai.org.**

2 In March 2019, the LCO organized Canada's first multidisciplinary forum on AI in Canada's criminal justice system. The event brought together more than 50 policymakers, Crown Attorneys, defence counsel, jurists, technologists, academics and community organizers to discuss predictive policing, algorithmic risk assessments, how to "litigate algorithms" and related human rights and due process issues. Material for this event is available at: **https://www.lco-cdo.org/wp-content/uploads/2019/11/LCO-Crim-AI-Background-Package.pdf.**

3 In December 2019, the LCO organized Canada's first multidisciplinary forum considering the use of AI and algorithms in regulatory investigations, government benefit determinations, and to support decision-making in the civil and administrative justice systems. This event brought together almost 40 policymakers, lawyers, jurists, technologists, academics, and community organizers to share experiences, discuss issues and consider law reform options in civil and administrative law applications.

4 This project considers how to better protect consumers in Canada's "digital marketplace." The LCO's partner in this project is the Centre for Law, Technology and Society at the Faculty of Law, University of Ottawa.

5 The LCO's Defamation in the Internet Age project considered how to balance reputation, freedom of expression, access to justice and legal regulation of new technologies, particularly internet intermediaries. The LCO's *Final Report* was released in March 2020 and is available at: **https://www.lco-cdo.org/wp-content/uploads/2020/03/Defamation-Final-Report-Eng-FINAL-1.pdf.**

6 This term comes from the title of a book by Cathy O'Neil, a former Wall Street data scientist and mathematician. Her 2016 book, *Weapons of Math Destruction*, popularized the idea that AI, algorithms and big data reinforce and worsen bias and discrimination in public and private sector decision-making.

7 Human Rights Watch, *Not In It For Justice,* (April 2017) [Human Rights Watch], online: **https://www.hrw.org/report/2017/04/11/not-it-justice/how-californias-pretrial-detention-and-bail-system-unfairly.**

8 Sandra Gabriel Mayson, *Bias In, Bias Out* (September 28, 2018), 128 Yale Law Journal 2218 (2019), University of Georgia School of Law Legal Studies Research Paper No. 2018-35 [Mayson] at 2280, online: **https://ssrn.com/abstract=3257004.**

9 Sarah Picard-Fritshe et al, *Beyond the Algorithm: Pretrial Reform, Risk Assessment, and Racial Fairness*, Center on Court Innovation, (July 2019) [Pitcard-Fritshe] at 3, online: **https://www.courtinnovation.org/sites/default/files/media/document/2019/Beyond_The_Algorithm.pdf.**

10 For a good analysis of the "evidence-based criminal justice movement", see Megan Stevenson, *Assessing Risk Assessment in Action* (Feb 2018) [Stephenson] at 312-13, online: **http://dx.doi.org/10.2139/ssrn.3016088.**

11 Access Now, *The Toronto Declaration,* (May 2018) [Toronto Declaration], online: **https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/.** This statement was released at the 2018 RightsCon international conference in Toronto.

12 *Ibid.*

13 AI Now Institute, *Algorithmic Accountability Policy Toolkit*, (October 2018) [AlNow Accountability Toolkit] at 2, online: **https://ainowinstitute.org/aap-toolkit.pdf.**

14 For a general primer on the operation of algorithms, see Deven Desai and Joshua Kroll, *Trust But Verify: A Guide to Algorithms and the Law* (April 27, 2017). Harvard Journal of Law & Technology, Georgia Tech Scheller College of Business Research Paper No. 17-19 [Desai and Kroll] at 23-29, online: **https://ssrn.com/abstract=2959472**. For an introduction to the range of AI definitions, history and the operation of AI and algorithmic

predictive systems, see Colin Gavaghan et al, *Government Use of Artificial Intelligence in New Zealand: Final Report on Phase 1 of the NZ Law Foundation's AI and Law in NZ Project*, (Wellington, 2019) [NZ AI and Law] at 5-19, online: **https://www.cs.otago.ac.nz/research/ai/AI-Law/NZLF%20report.pdf.**

15 The following list taken from surveys of the current AI and algorithmic tools in use in governments, including AINow Accountability Toolkit; David Freeman Engstrom, Daniel E. Ho, Catherine M. Sharkey and Mariano-Florentino Cuéllar, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies* (February 1, 2020) [US Federal Administrative Agencies], online: **https://ssrn.com/abstract=3551505;** Michele Gilman, *Poverty Algorithms: A Poverty Lawyer's Guide to Fighting Automated Decision-Making Harms in Low-Income Communities,* Data and Society (September 2020) [Gilman], online: **https://datasociety.net/wp-content/uploads/2020/09/Poverty-Lawgorithms-20200915.pdf;** Dr. Michael Veale et al, *Algorithms In The Criminal Justice System*, Law Society of England and Wales, (June 2019) [UK Law Society], online: https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-in-the-criminal-justice-system-report/; Australian Human Rights Commission, *Human Rights and Technology Discussion Paper,* (December 2019) [Australian Human Rights and Technology], online: **https://tech.humanrights.gov.au/sites/default/files/2019-12/TechRights_2019_DiscussionPaper.pdf** and NZ AI and Law.

16 The AI Now Institute notes "[a]*utomated decision systems can exist in any context where government bodies or agencies evaluate people or cases, allocate scarce resources, focus scrutiny or surveillance on communities, or make nearly any sort of decision.*" AI Now Accountability Toolkit at 9.

17 US Federal Administrative Agencies at 38.

18 *Ibid* at 43.

19 *Ibid* at 37-41.

20 *Ibid* at 44.

21 *Ibid* at 44-45.

22 *Ibid* at 16.

23 See the current use surveys identified in note 15 above.

24 See generally AI Now Accountability Toolkit at 7, UK Law Society and NZ AI and Law at 20-30.

25 UK Law Society at 45-46.

26 Kate Robertson, Cynthia Khoo and Yolanda Song, *To Surveil and Predict, A Human Rights Analysis of Algorithmic Policing in Canada*, Citizen Lab and International Human Rights Program, University of Toronto Faculty of Law, (September 2020) [Citizen Lab], online: **https://citizenlab.ca/2020/09/to-surveil-and-predict-a-human-rights-analysis-of-algorithmic-policing-in-canada/**

27 *Ibid* at 68.

28 This movement emphasizes using social science research techniques to identify practices that improve criminal justice system outcomes. For a good analysis, see Stevenson at 312-13.

29 John Logan Koepke and David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform* (Dec 2018) [Danger Ahead] at 1746, online: **https://ssrn.com/abstract=3041622.**

30 The Champion, *Making Sense of Risk Assessments,* American National Association of Criminal Defense Lawyers, (June 2018) [The Champion], online at **https://www.nacdl.org/Article/June2018-MakingSenseofPretrialRiskAsses**.

31 Stephenson at 321-22.

32 Picard-Fritshe at 3.

33 See Sam Corbett-Davies, Sharad Goel and Sandra Gonzales-Bailon "Even Imperfect Algorithms Can Improve the Criminal Justice System", *New York Times,* December 20, 2017, online: **https://www.nytimes.com/2017/12/20/upshot/algorithms-bail-criminal-justice-system.html.**

34 **https://www.congress.gov/bill/115th-congress/senate-bill/1593/text.** For a good summary of the then-bipartisan consensus on bail reform, see Senator Harris and Paul's October 2017 opinion piece in the New York

Times, titled *Let's Reform Bail,* available at **https://www.nytimes.com/2017/07/20/opinion/kamala-harris-and-rand-paul-lets-reform-bail.html.**

35 National Association for Public Defence et al, *Joint Statement in Support of the Use of PreTrial Risk Assessment Instruments,* (Oct 2017) [NAPDA Joint Statement], online: **https://www.publicdefenders.us/blog_home.asp?Display=563**.

36 The Champion.

37 Stephenson at 314, n 65.

38 Heather Harris, Justin Goss and Alexandria Gumbs, *Pretrial Risk Assessment in California*, Public Policy Institute of California, (December 2019) [Risk Assessment in California] at 4, online: **https://www.ppic.org/wp-content/uploads/pretrial-risk-assessment-in-california.pdf.**

39 *Ibid.*

40 COMPAS is an acronym for Correctional Offender Management Profiling for Alternative Sanctions.

41 Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016) [ProPublica], online: **https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.**

42 *Ibid.*

43 For a good summary of this debate, see Anne Washington, *How to Argue with an Algorithm: Lessons from the COMPAS ProPublica Debate* (February 2019), online: **https://ssrn.com/abstract=3357874.**

44 Mayson at 2230.

45 Human Rights Watch.

46 See "A Shared Statement of Civil Rights Concerns" [Shared Statement] signed by more than 100 community organizations, including the ACLU, the Center for Race, Inequality and the Law at NYU, Civil Rights Corps, the Electronic Frontier Foundation, the NAACP, and numerous community and public defender organizations. Online at **http://civilrightsdocs.info/pdf/criminal-justice/Pretrial-Risk-Assessment-Full.pdf.**

47 Pretrial Justice Institute, *Updated Position on Pretrial Risk Assessments,* (February 7, 2020) [PJI Update], online: **https://www.pretrial.org/wp-content/uploads/Risk-Statement-PJI-2020.pdf**.

48 Advancing Pretrial Policy and Research, *APPR Statement on Pretrial Justice and Pretrial Assessment,* February 24, 2020) [APPR], online: **https://mailchi.mp/7f49d0c94263/our-statement-on-pretrial-justice?e=a01efafabd.**

49 See, for example, Massachusetts Special Commission on Bail Reform, *Final Report of the Special Commission to Evaluate Policies and Procedures Related to the Current Bail System,* (December 31, 2019), online: **https://d279m997dpfwgl.cloudfront.net/wp/2020/01/0102_bail-reform-report.pdf.**

50 For a history of the use of risk assessment in criminal justice system in the United States, see Danielle Kehl, Priscilla Guo, and Samuel Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, (2017) [Kehl et al], Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School, online **https://dash.harvard.edu/handle/1/33746041.**

51 *Ibid* at 8.

52 PAI at 7.

53 Much of the following section is based on information provided by Advancing Pretrial Policy and Research (APPR). APPR is a project of the National Partnership for Pretrial Justice with support from Arnold Ventures, the developers of the Public Safety Assessment. More information about APPR and the PSA is available at **https://advancingpretrial.org/psa/about/. PSA implementation guides are available at https://advancingpretrial.org/implementation/guides/.**

54 **https://advancingpretrial.org/psa/factors/#psa-factors**

55 **https://advancingpretrial.org/psa/psa-sites/**

56 *Ibid.*

57 Risk Assessment in California at 7-10.

58 According to Eugenie Jackson and Christina Mendoza," [s]tatic factors, such as past Criminal History, tend to be unchanging and cannot respond to treatment. Dynamic risk factors, also known as criminogenic needs, are changeable and responsive to treatment." See

Eugenie Jackson and Christine Mendoza, *Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not*, (2020) [Jackson and Mendoza]. *Harvard Data Science Review*, *2*(1), online: **https://doi.org/10.1162/99608f92.1b3dadaa.**

59 For a sample list of the COMPAS assessment questions, see Appendix A: COMPAS-Probation Documents Full COMPAS Assessment Instrument in Sharon Lansing, *New York State COMPAS-Probation Risk and Need Assessment Study: Examining the Recidivism Scale's Effectiveness and Predictive Accuracy*, New York State Division of Criminal Justice Services Office of Justice Research and Performance, online: **https://www.criminaljustice.ny.gov/crimnet/ojsa/opca/compas_probation_report_2012.pdf.**

60 Human Rights Watch.

61 David Robinson and Logan Keopke, *Civil Rights and Pretrial Risk Assessment Instruments,* Upturn Inc., Washington (December 2019) [Civil Rights and Risk Assessments] at 4, online: **http://www.safetyandjusticechallenge.org/wp-content/uploads/2019/12/Robinson-Koepke-Civil-Rights-Critical-Issue-Brief.pdf.**

62 This phrase is taken from an article by Sandra Mayson. See generally Sandra Gabriel Mayson, *Bias In, Bias Out* (September 28, 2018). 128 Yale Law Journal 2218 (2019), University of Georgia School of Law Legal Studies Research Paper No. 2018-35 [Mayson], online: **https://ssrn.com/abstract=3257004.**

63 See generally, Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan and Cass Sunstein, *Discrimination in the Age of Algorithms*, *Journal of Legal Analysis*, Volume 10, 2018 [Kleinberg et al] at 113, online: **https://doi.org/10.1093/jla/laz001.**

64 Ontario Human Rights Commission, *A Disparate Impact: Second Interim Report on the Inquiry into Racial Profiling and Racial Discrimination of Black Persons by the Toronto Police Service,* (August 2020) [OHRC Disparate Impact], online: **http://www.ohrc.on.ca/sites/default/files/A%20Disparate%20Impact%20Second%20interim%20report%20on%20the%20TPS%20inquiry%20executive%20summary.pdf#overlay-context=en/disparate-impact-second-interim-report-inquiry-racial-profiling-and-racial-discrimination-black.**

65 *Ibid* at 2.

66 Dr. Ivan Zinger, Correctional Investigator of Canada, "Indigenous People in Federal Custody Surpasses 30% Correctional Investigator Issues Statement and Challenge" Public Safety Canada (January 21, 2020), online at **https://www.canada.ca/en/public-safety-canada/news/2020/01/indigenous-people-in-federal-custody-surpasses-30-correctional-investigator-issues-statement-and-challenge.html.**

67 Citizen Lab at 15-18.

68 A good discussion of these and related concepts framing AI in the Canadian legal context took place the LCO's Criminal Justice Roundtable [LCO Criminal Justice Roundtable] in March 2019. Roundtable materials are available online: **https://www.lco-cdo.org/wp-content/uploads/2019/11/LCO-Crim-AI-Background-Package.pdf.**

69 One notable exception being Citizen Lab 101-123 within the context of predictive policing.

70 For an interesting analysis of AI and disability in the United States, see Meredith Whittaker et al, *Disability, Bias, and AI,* AI Now Institute (November 2019), online: **https://ainowinstitute.org/disabilitybiasai-2019.pdf.**

71 Joshua Kroll, Joanna Huey, Solon Barocas, Edward Felten, Joel Reidenberg, David Robinson, and Harlan Yu, *Accountable Algorithms* (March 2, 2016) University of Pennsylvania Law Review, Vol. 165, 2017, Fordham Law Legal Studies Research Paper No. 2765268 [Kroll et al] at 41, online: **https://ssrn.com/abstract=2765268.**

72 ProPublica.

73 More specifically, the COMPAS controversy demonstrated the differential impact of emphasizing false positives versus false negatives as a statistical measure of fairness. The distinction between the two can be confusing but is significant: *Accuracy in risk assessment can be defined in terms of error rates. Risk assessment tools can make two kinds of errors.* ***False positives occur when people are misclassified as high risk****. When these types of errors occur, arrested individuals and their families primarily the bear the costs because people classified as high risk are more likely to be detained...****False negatives occur when people are misclassified as low risk.*** *Victims and communities primarily bear the costs of this kind of error because people classified as low risk are*

more likely to be released and therefore have the opportunity to commit crimes in the community. [Emphasis added] Pretrial Risk Assessments in California at 14. According to the Shared Statement, "…the design of any tools should give far greater weight to the avoidance of false positives than false negatives, as the harms of detaining the wrong person are far greater than erring on the side of release." Shared Statement at 4.

74 Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns and Aaron Roth, *Fairness in Criminal Justice Risk Assessments: The State of the Art*. Sociological Methods & Research (2017) [Berk et al] at 12-15, online: **https://www.researchgate.net/publication/315667137_Fairness_in_Criminal_Justice_Risk_Assessments_The_State_of_the_Art.**

75 Mayson at 2223.

76 Huq, Aziz Z., *Racial Equity in Algorithmic Criminal Justice* (June 20, 2018). Duke Law Journal, Vol. 68, 2019, U of Chicago, Public Law Working Paper No. 663 [Huq] at 1053, online: **https://ssrn.com/abstract=3144831.**

77 See generally, Kehl et al at 28-32.

78 Automated Decisions Systems Task Force, *Automated Decision Systems Task Force Report,* New York City, (November 2019) [NYC AI Task Force], online; **https://www1.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf.**

79 Rashida Richardson, ed., *Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force*, AI Now Institute, December 4, 2019 [AI Now Shadow Report], online: **https:// ainowinstitute.org/ads-shadowreport-2019.html.**

80 Christopher Bavitz, Sam Bookman, Jonathan Eubank, Kira Hessekiel, and Vivek Krishnamurthy, *Assessing the Assessments: Lessons from Early State Experiences In the Procurement and Implementation of Risk Assessment Tools* (December 2018) Berkman Klein Center Research Publication No. 2018-8 [Berkman Klein Early Lessons], online: **https://ssrn.com/abstract=3297135.** See generally, **https://cyber.harvard.edu/projects/ai-algorithms-and-justice.**

81 See, for example, Wendy Gillis, "Critics of new carding policy: "Destroy the data," *Toronto Star* November 17, 2016), online: **https://www.thestar.com/news/gta/2016/11/17/critics-of-new-carding-policy-destroy-the-data.html** and Donovan Vincent, "Sidewalk Labs' urban data trust is "problematic," says Ontario privacy commissioner," *Toronto Star,* (September 26, 2019), online: **https://www.thestar.com/news/gta/2019/09/26/sidewalk-labs-urban-data-trust-is-problematic-says-ontario-privacy-commissioner.html.**

82 PAI at 14-15.

83 Danger Ahead at 1755.

84 *Ibid.*

85 PAI at 15. See PAI Requirement 1: Training datasets must measure the intended variables.

86 *Ibid* at 16. See PAI Requirement 2: Bias in statistical models must be measured and mitigated.

87 *Ibid* at 22. See PAI Requirement 3: Tools must not conflate multiple distinct predictions.

88 *Ibid* at 30. Requirement 9: Tools must support data retention and reproducibility to enable meaningful contestation and challenges.

89 See Danielle Keats Citron and Frank Pasquale, *The Scored Society: Due Process for Automated Predictions* (2014). Washington Law Review, Vol. 89, 2014, U of Maryland Legal Studies Research Paper No. 2014-8 [Scored Society] at 1, online: **https://ssrn.com/abstract=2376209.**

90 According to Koepke and Robinson, the probability of *success* (i.e. not being rearrested prior to an individual's trial) ranges from 83.5% (Federal Pretrial Risk Assessment Instrument) to 74% (PSA) to 57.9% (COMPAS). Civil Rights Risk Assessments at 7-8.

91 Shared Statement at 8.

92 Danger Ahead at 1804.

93 PAI at 23.

94 *Ibid* at 24.

95 See the discussion in section Six of this paper and LCO's forthcoming Issue Papers, *Regulating AI: An International Survey* and *AI, Algorithms and Government Decision-Making*.

96 Shared Statement.

97 Dillon Reisman, Jason Schultz, Kate Crawford and Meredith Whittaker, *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability,* AI Now Institute (April 2018) [AI Now Impact Assessment], online: **https://ainowinstitute.org/aiareport2018.pdf.**

98 Berkman Klein Early Lessons.

99 Electronic Frontier Foundation, *Artificial Intelligence and Algorithmic Tools: Policy Guide for Judges and Judicial Officers* (2018) [EFF], online: **https://www.eff.org/files/2018/12/21/ai_policy_issues_handout.pdf.**

100 NLADA Joint Statement.

101 The Toronto Declaration.

102 Institute of Electrical and Electronics Engineers, *Ethically Aligned Design* (1st Edition, 2019) [IEEE], online: **https://standards.ieee.org/industry-connections/ec/autonomous-systems.html.**

103 NZ AI and Law.

104 PAI at 15.

105 *Ibid* at 16-20.

106 *Ibid* at 22.

107 *Ibid* at 24.

108 *Ibid* at 25.

109 *Ibid* at 26.

110 *Ibid* at 27-28.

111 *Ibid* at 29.

112 *Ibid* at 30.

113 *Ibid* at 31-33.

114 Law Commission of Ontario, *LCO / Mozilla Foundation Roundtable Report on Digital Rights and Digital Society* (May 2018) [LCO/Mozilla Roundtable], online: **https://www.lco-cdo.org/wp-content/uploads/2018/08/LCO-Mozilla-Roundtable-Final-Report-EN.pdf.**

115 NY AI Task Force.

116 AI Now Shadow Report.

117 Shared Statement.

118 See "Policing Black Bodies II: Race and Pretrial Practices", a virtual panel discussion with various stakeholders in pretrial justice, hosted by the National Association of Criminal Defense Lawyers (July 2020) [PJI Webcast], online: **https://www.nacdl.org/Content/Race-and-the-Criminal-Justice-System-Series?mc_cid=2ba1d628e5&mc_eid=8b246f8b29.** The PJI's "racial equity journey" involved extensive community engagement, changing the ethnic and racial composition of PJI's leadership and staff members, and hiring a racial healing psychotherapist. See generally, Pretrial Justice Institute, *A Racial Equity Transformation: PJI's Rationale*, (July 2019), online at: **https://university.pretrial.org/viewdocument/a-racial-equity-transformation-pji.**

119 PJI Webcast.

120 *Ibid.*

121 Angela Hattery and Earl Smith, *Policing Black Bodies, (2017).*

122 PJI Webcast.

123 *Ibid.*

124 See, for example, the AI Now Accountability Toolkit; The Toronto Declaration; the AI: Algorithms and Justice project at the Berkman Klein Center for Internet and Society, Harvard University; Electronic Frontier Foundation, *Artificial Intelligence and Algorithmic Tools: Policy Guide for Judges and Judicial Officers* (2018), online: **https://www.eff.org/files/2018/12/21/ai_policy_issues_handout.pdf;** Kira Hessekiel, Kim Eliot, James Tierney, Jonathan Yang, and Christopher T. Bavitz, *AGTech Forum Briefing Book: State Attorneys General and Artificial Intelligence*, May 8-9, 2018, Harvard Law School. Berkman Klein Center for Internet & Society, online: **https://cyber.harvard.edu/publications/2018/05/AGTech.**

125 Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008) [Citron], online:: **https://openscholarship.wustl.edu/law_lawreview/vol85/iss6/2.**

126 Kroll et al.

127 See generally Kehl et al at 32 and Scored Society.

128 Desai and Kroll at 8 (footnotes omitted).

129 See generally, NYC AI Task Force at 27-28.

130 See generally, Taylor R. Moore, *Trade Secrets and Algorithms as Barriers to Social Justice*, Center for Democracy and Technology (August 2017), online: **https://cdt.org/wp-content/uploads/2017/08/2017-07-31-Trade-Secret-Algorithms-as-Barriers-to-Social-Justice.pdf**

131 PAI at 31.

132 AI Now Impact Assessment.

133 The US Constitution's Equal Protection Clause is included in section 1 of the 14th Amendment. This section reads, in part
"*No State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any State deprive any person of life, liberty, or property, without due process of law;* ***nor deny to any person within its jurisdiction the equal protection of the laws.*** *[Emphasis added]."*

134 Huq at 31. For an extensive discussion of algorithmic bias and equal protection under the US Constitution, see Huq at 31-70.

135 See, for example, Finale Doshi-Velez and Kortz Finale, *Accountability of AI Under the Law: The Role of Explanation.* Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper (2017), online: **https://dash.harvard.edu/bitstream/handle/1/34372584/2017-11_aiexplainability-1.pdf?sequence=3&isAllowed=y.**

136 For a discussion of AI, algorithms and their potential impact on access to justice and legal aid, see the LCO's paper on this subject delivered to the International Legal Aid Group: Ryan Fritsch and Nye Thomas, *AI and Automated Decision-Making: Impact on Access to Justice and Legal Aid,* Law Commission of Ontario (2019), online: **https://www.lco-cdo.org/wp-content/uploads/2019/06/LCO-ILAG-Paper-AI-Legal-Aid-and-Access-to-Justice-June-3-2019.pdf.**

137 The Due Process Clause is included in section 1 of the 14th Amendment. This section reads, in part
"*No State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States;* ***nor shall any State deprive any person of life, liberty, or property, without due process of law****; nor deny to any person within its jurisdiction the equal protection of the laws. [Emphasis added].*
The US Constitution's 5th Amendment contains similar language applied to the US Federal Government.

138 *State v. Loomis,* 881 N.W.2d 749 (Wisc. 2016).

139 The *Loomis* decision has been criticized for providing insufficient procedural protections to criminal defendants. The case is also illustrative of the limits of litigation as a means to ensure due process protections in cases involving AI and algorithms in the criminal justice system. The LCO will discussed these issues further in the discussion of Issue #10 – The Limits of Litigation.

140 ss in a litigation context, see AI Now Institute, *Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems.*

141 Desai and Kroll at 4. In the authors view, part of the misunderstanding "may flow in part from the religious, devotional culture around algorithms, where algorithms might as well be God." *Ibid.*

142 *Ibid* at 5.

143 Kroll et al at 656-677.

144 *Ibid* at 678-695.

145 See *Ewart v. Canada, 2018 SCC 30* for an early Canadian case.

146 *State v. Loomis,* 881 N.W.2d 749 (Wisc. 2016) [*Loomis*].

147 For more on *Loomis*, see Berkman Klein Early Lessons at 6-8; Kehl et al at 18-21; Lauren Kirchner, *Wisconsin Court: Warning Labels Are Needed for Scores Rating Defendants' Risk of Future Crime*, ProPublica (Jul. 14, 2016), online: **https://www.propublica.org/article/wisconsin-court-warning-labels-neededscores-rating-risk-future-crime.**

148 *Loomis* at para 34.

149 For a summary of these issues, see the sources listed in note 147.

150 See generally, AI Now Institute, *Litigating Algorithms,* (September 2018), online at: **https://ainowinstitute.org/litigatingalgorithms.pdf;** Rashida Richardson, Jason Schultz and Vincent Southerland, *Litigating Algorithms 2019 US Report,* AI Now Institute, (September 2019), online: **https://ainowinstitute.org/litigatingalgorithms-2019-us.pdf;** and Gilman.

151 Jesse Beatson, Gerald Chan, and Jill Presser (eds), *Litigating AI*, (Toronto: Emond Publishing, forthcoming 2021).

152 Government of Canada, *Directive on Automated Decision-Making,* February 5, 2019 [Canada AI Directive], online: *Government of Canada* **https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592**

153 *Ibid,* s. 4.1

154 *Ibid*, s 4.2

155 *Responsible use of artificial intelligence (AI): Exploring the future of responsible AI in government*, September 9, 2019, [Canada AI White Paper], online: *Government of Canada* < **https://www.canada.ca/en/government/system/digital-government/modern-emerging-technologies/responsible-use-ai.html#toc2>.**

156 Canada AI Directive, s. 5.

157 Canada AI White Paper.

158 *Ibid,* s 6.1.

159 *Ibid,* s 6.2.

160 *Ibid,* s 6.3.

161 *Ibid,* s 6.4.

162 *Ibid*, s 6.5.

163 *Ibid*, s. 6.1.

164 *Ibid*, s. 6.

165 *Ibid.*

166 *Ibid*, s. 6.3.1.

167 *Ibid*, s. 6.3.2.

168 *Ibid*, s. 6.4.

169 *Ibid.*

170 **https://www.congress.gov/bill/116th-congress/house-bill/2231/all-info.**

171 Statement by UN High Commissioner for Human Rights Michelle Bachelet, "Smart mix of measures needed to regulate new technologies" (April 24, 2019) online: **https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=24509**

172 Chloe Anderson, Cindy Redcross, Erin Valentine and Luke Miratrix., *Evaluation of Pretrial Justice Systems Reform That Use The Public Safety Assessment,* MDRC Center for Criminal Justice Research, (November 2019) [MDRC Report], online: **https://www.mdrc.org/sites/default/files/PSA_New_Jersey_Report_%231.pdf.**

173 Mayson at 2280.

174 *Ibid.*

175 *Ibid*.

176 *Ibid* at 2282. For a sophisticated description of this argument, see Mayson at 2281-2296.

177 *Ibid* at 2293.