Canadian Journal of Law and Technology

Volume 19
Number 2 *Technology-facilitated Gender-based Violence*

Article 10

1-2022

Artificial Intelligence in Canadian Healthcare: Will the Law Protect Us from Algorithmic Bias Resulting in Discrimination?

Bradley Henderson
University of Ottawa, Faculty of Law

Colleen M. Flood
University of Ottawa, Faculty of Law

Teresa Scassa
University of Ottawa, Faculty of Law

Artificial Intelligence in Canadian Healthcare: Will the Law Protect Us from Algorithmic Bias Resulting in Discrimination?

Bradley Henderson, Colleen M. Flood & Teresa Scassa[*]

INTRODUCTION

Many hope that artificial intelligence (AI) will transform healthcare systems by improving clinical outcomes, efficiency, safety, and quality, as well as access to care (e.g., reducing wait times, improving the speed of diagnoses, and permitting the delivery of highly specialized care like robotic surgery in remote and rural areas).[1] AI may also help improve upon existing high rates of avoidable medical errors[2] and help healthcare providers automate many administrative tasks that distract them from patient care. However, the introduction of AI in healthcare has also raised concerns, with perhaps the most pressing among them being that AI could perpetuate or exacerbate existing bias and discrimination problems. On the face of it, one might hope that the introduction of AI algorithms would *enhance* objectivity in clinical decision-making by counteracting human biases known to result in differential treatment on prohibited grounds of discrimination, including on the basis of race, sex, and sexual orientation.[3] However, because AI is built by innovators with their own biases and/or upon data sets that may reflect pre-existing biases and inequalities,[4] there is a danger that AI, at least in some cases, will instead perpetuate or even

---

[*] Bradley Henderson is a JD Candidate at the University of Ottawa Faculty of Law. He is also a salaried employee of Health Canada. However, Health Canada played no role whatsoever in the conception, design, analysis, or writing of this article. This article reflects the views of the authors, and not those of Health Canada. Colleen M Flood is a Professor, University of Ottawa Research Chair, and Director of the uOttawa Centre for Health Law, Policy & Ethics. Teresa Scassa is a Professor and Canada Research Chair in Information Law and Policy at the University of Ottawa. We would like to warmly thank Michael Da Silva and Bryan Thomas for their insightful comments on various parts of this article.

1 See Royal College of Physicians and Surgeons of Canada, *Task Force Report on Artificial Intelligence and Emerging Digital Technologies* (Ottawa: Royal College of Physicians and Surgeons of Canada, 2020) at 34 [Task Force].

2 See G Ross Baker et al, "The Canadian Adverse Events Study: The Incidence of Adverse Events Among Hospital Patients in Canada" (2004) 170:11 CMAJ 1678 at 1683-85. See also Fei Jiang et al, "Artificial Intelligence in Healthcare: Past, Present and Future" (2017) 2:4 Stroke & Vascular Neurology 230 at 230 [Jiang].

3 See Ravi B Parikh, Stephanie Teeple & Amol S Navathe, "Addressing Bias in Artificial Intelligence in Health Care" (2019) 322:24 JAMA 2377 at 2378 [Parikh].

4 See Robert Challen et al, "Artificial Intelligence, Bias and Clinical Safety" (2019) 28:3 BMJ Quality & Safety 231 at 232-35.

worsen existing discrimination within healthcare. Addressing this concern is essential so that the benefits of healthcare-related AI are not realized at the expense of perpetuating or increasing discrimination (whether direct or indirect) against marginalized groups.

The term "algorithmic bias" is generally used in the literature to refer to an unwarranted skewing of outputs because of a problem in the algorithmic design. It is also closely related to "data bias," where the data that the algorithm is built upon or uses is skewed in an unwarranted way. In this article we are concerned with algorithmic and data bias resulting in healthcare-related AI applications that treat groups or individuals differently on the basis of arbitrary traits, such as race or sex, with no clinical, moral, or legal justification. So understood, algorithmic and data bias result in the legal and moral concept of wrongful discrimination. This form of injustice can be intentional or unintentional, and its effects can be individualized or systemic. Unintentional and systemic discrimination may occur where an AI application makes recommendations based on arbitrary factors such as race or gender. These recommendations can have adverse effects for patients and/or exacerbate existing inequalities.[5]

Legal and moral responsibility for addressing these risks extends beyond the AI innovators who create AI algorithms. Before licensing their use, as we discuss below, regulators have an obligation to scrutinize healthcare-related AI health applications and the data sources upon which they are trained, to detect their limitations or embedded biases and, at a minimum, ensure that concerns are communicated to health professionals and affected patients.

In this article, we canvas *why* AI may perpetuate or exacerbate extant discrimination through a review of the training, development, and implementation of healthcare-related AI applications and set out policy options to militate against such discrimination. The article is divided into eight short parts including this introduction. Part II focuses on explaining AI, some of its basic functions and processes, and its relevance to healthcare. In Part III, we define and explain the difference and relationship between algorithmic bias and data bias, both of which can result in discrimination in healthcare settings, and provide some prominent examples of healthcare-related AI applications that have resulted in discrimination or have produced discriminatory outputs. Part IV explains in more detail differences between algorithmic bias and data bias, with a focus on data bias and data governance, including the non-representativeness of data sets used in training AI. From this point we turn to look at possible legal responses to the problem of algorithmic discrimination, and, in Part V, we demonstrate the insufficiency of existing *ex post* legal protections (i.e., legal protections that offer redress after someone has suffered harm), including claims in negligence, under human rights legislation, and under the *Charter of Rights and Freedoms*.[6] Part VI explores possibilities within the Canadian *ex ante* legal

---

[5] See the discussion on adverse impact discrimination in *Fraser v. Canada (Attorney General)*, 2020 SCC 28, 2020 CarswellNat 4333, 2020 CarswellNat 4334 (S.C.C.) at paras. 30, 50.

landscape (i.e., the regulation of AI applications before they become available for use in healthcare settings), notably through federal regulation of medical devices, and identifies gaps in oversight. Finally, in Part VII we provide recommendations for federal and provincial governments and innovators as to the appropriate governance and regulatory approach to counter algorithmic and data bias that results in discrimination in healthcare-related AI, before concluding in Part VIII.

1. HEALTHCARE-RELATED AI APPLICATIONS[7]

AI has been broadly defined as "intelligent computer programs" capable of undertaking higher-order tasks associated with intelligent beings, such as making complex predictions or recommendations.[8] For the purpose of this article, we are concerned with "narrow" AI — that is to say, AI that performs specific tasks or seeks to achieve specific goals within a set of pre-determined parameters. For example, narrow AI could support healthcare practitioners or system managers in solving a specific problem based on certain inputs (e.g., diagnostic images) or in identifying new patterns.[9] "General AI," or AI that can define and seek to achieve its own tasks and goals beyond pre-defined parameters, similar to the human mind, does not yet exist.[10] This type of AI could include, for example, a fully-automated healthcare provider capable of identifying novel problems, categorizing them, connecting them to relevant information, and applying its decision-making processes to implement solutions.

In healthcare, "machine learning" (ML) is a common form of narrow AI that may, for example, seek to optimize workflow or predict the results of diagnostic imaging.[11] AI involves mathematical and statistical analysis.[12] ML

---

6 *Canadian Charter of Rights and Freedoms,* s 15, Part I of the *Constitution Act, 1982,* being Schedule B to the *Canada Act 1982* (UK), 1982, c 11 [*Charter*].

7 AI applications do and will have a much broader reach beyond healthcare and into public health, but for reasons of scope and the length of this article we primarily confine our analyses to issues of algorithmic/data bias and discrimination resulting therefrom.

8 See World Health Organization, *Ethics and governance for artificial intelligence for health: WHO guidance* (Geneva: World Health Organization, 2021) at 4 [WHO Guidance]; see also Task Force, *supra* note 1 at 14; see also Timo Minssen et al, "Regulatory Responses to Medical Machine Learning" (2020) 7:1 JL & Biosciences 1 at 2 [Minssen].

9 See Bertalan Mesko, "The Role of Artificial Intelligence in Precision Medicine" (2017) 2:5 Expert Rev Precision Medicine & Drug Development 239 at 240.

10 See Erik R Ranschaert et al, "Advantages, Challenges, and Risks of Artificial Intelligence for Radiologists" in Erik R Ranschaert, Sergey Morozov & Paul Algra, eds, *Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks* (Switzerland: Springer, 2019) 329 at 330.

11 Sarah Gerke, Timo Minssen & Glenn Cohen, "Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare" (2020) Academic Press 295 at 295 [Gerke]. See also, generally, Weicheng Kuo et al "Expert-Level Detection of Acute Intracranial Hemorrhage on Head Computed Tomography Using Deep Learning" (2019) 116:45

takes this one step further — it often involves neural networks and deep learning such that AI can "learn" from its training data. Neural networks are an interconnected array of nodes consisting of, at a minimum, an input layer and an output layer. Each node is, generally, assigned a weight, which mathematically and statistically analyzes an input to convey an output.[13] The many connections between these nodes creates the "network." Deep learning involves a neural network with multiple hidden layers between the input and output layers for more extensive analysis,[14] rendering the algorithm's functions less transparent. Complex, highly obscure algorithms are often referred to as "black boxes"[15] and raise questions about the extent to which humans are able to supervise or oversee their operations.

AI developers, through a process called supervised learning,[16] use large data sets to "train" an algorithm by mapping the training data (i.e., inputs) to specific "successful" outputs or goals. During this process, the developer defines a list of features for the AI application to analyze, and they may assign weights to those features, or leave the initial assignment of weights for some or all features to random chance. The AIML application may then continuously adjust and optimize, including by adjusting its weighting of features, by learning from its training data. Deep learning, again, takes this one step further: the algorithm defines the features where they are less certain. Overall, this process generates the algorithm's map or "label,"[17] (i.e., its set of rules that will analyze new inputs and predict outputs or outcomes). Notably, a developer may not always fully understand an algorithm's rules, especially where the algorithm is a "black box," since they are a combination of the developer's programming (e.g., feature

---

Proceedings National Academy Sciences 22737, as well as Andre Esteva et al, "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks" (2017) 542:7639 Nature 115, for examples of AI applications in clinical practice, including software that assists healthcare professionals in more accurately diagnosing skin cancer and strokes.

12 Hooman H Rashidi et al, "Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods" (2019) 6 Academic Pathology 1 at 9 [Rashidi].

13 Task Force, *supra* note 1 at 14.

14 See Yann LeCunn, Yoshua Bengio & Geoffrey Hinton, "Deep Learning" (2015) 521:7553 Nature 436 at 436-38.

15 See Frank Pasquale, *The Black Box Society* (Cambridge, Massachusetts; London, England: Harvard University Press, 2015) [Pasquale].

16 In addition to supervised learning, AI developers may use unsupervised learning (clustering) to train an algorithm to identify commonalities in unorganized data. AI developers may also use reinforcement learning, whereby an algorithm is trained to learn a sequence of events based on its environment. See Rashidi, *supra* note 12 at 8.

17 In the context of AI, we refer to a "label" as meaning an AI application's "process map" between inputs and outputs, as derived through its training and labelling of data. For example, an AIML application trained to detect skin cancer will create a process map to predict the probability of a cancerous lesion when it is provided a new input image.

selection and defined goals) and the algorithm's mathematical and statistical processes as derived from its training. Importantly, after training, validating, and testing an algorithm, a developer may "lock" the algorithm, or they may leave it "unlocked" such that it can continually "adapt." Adaptive AIML is capable of learning not only from its training data, but from new input data in real-world settings, within its parameters, to continuously learn and self-optimize.[18]

An adaptive AIML application's continuous learning and self-optimizing may improve its safety and efficacy for some patients, possibly by reducing algorithmic bias if it is exposed to a diversity of patients, but the opposite may also hold true — especially for patients or patient populations on whom an AI application is used less frequently. Notably, developers may see feedback loop effects whereby an adaptive AIML application becomes increasingly optimized for data that it is exposed to at greater frequencies.[19] This raises complex bioethical questions including whether highly beneficial AIML applications should be allowed to learn in real time where their training data is limited. In such cases, patients who are not well represented in an adaptive AIML application's training data would likely bear greater safety risks and risk of discrimination.

While adaptive AIML applications in healthcare are under development, many non-adaptive AI and AIML applications are being tested or have been implemented in healthcare. Prominent types of healthcare-related AI applications include those that identify and classify skin cancers, restore the control of movement in patients with neurological conditions, and diagnose cardiac conditions.[20] The importance of early diagnosis and treatment in these therapeutic areas to prevent death and disease deterioration[21] has also likely spurred development of AI applications that apply at multiple stages of care, such as in detecting and diagnosing strokes, predicting the effectiveness of treatments, and predicting disease prognosis.[22] Healthcare-related AI applications also extend beyond clinician tools and include use for setting priorities within healthcare systems and other patient-level and data analytics applications.[23]

While healthcare-related AI applications are rapidly expanding and advancing, the features used as inputs often include common patient traits, such as symptoms, test results, comorbidities, age, gender, and race.[24] Generally,

---

[18] See Minssen, *supra* note 8 at 2.

[19] See James Zou & Londa Schiebinger, "AI Can Be Sexist and Racist — It's Time to Make It Fair" (2018) 559 Nature 324 at 325 [Zou].

[20] See Jiang, *supra* note 2 at 231.

[21] *Ibid.*

[22] *Ibid.* at 239-40.

[23] See Eric J Topol, "High-Performance Medicine: The Convergence of Human and Artificial Intelligence" (2019) 25:1 Nature Medicine 44 [Topol].

[24] See Jiang, *supra* note 2 at 232.

the greater the number of patient traits, and the larger the quantity of other available information, the greater the ability of a developer to train an algorithm to recognize concealed patterns, learn from outliers, and tailor itself to a greater number of subsets of individuals or patient groups (therefore reducing the likelihood of discrimination caused by the algorithm itself). However, training data is often fraught with bias, which, left unmitigated, could result in discrimination in, for example, treatment decisions vis-à-vis marginalized patients. And importantly, many patients' traits map onto prohibited grounds of discrimination under federal and provincial human rights legislation.[25] Below, we explore whether and to what extent the problem of algorithmic bias, resulting in discrimination, can be counteracted — whether *ex ante* (e.g., by working upstream to mitigate bias in training data) or *ex post* (e.g., through litigation by individuals and groups who have experienced discrimination).

2. MULTIPLE SOURCES OF BIAS ARE LIKELY TO ARISE IN THE DEVELOPMENT AND USE OF AI APPLICATIONS IN HEALTHCARE

To better understand healthcare-related AI bias, we can distinguish between two often interrelated kinds of bias with two subsets therein. The first is algorithmic bias, which includes coding bias that is introduced, even if unconsciously, by a *developer* during an AI application's development, as well as AIML that itself "learns" over time to be biased. The second is biased data with two sources of bias of most concern: (a) training on non-representative data (i.e., data that is not properly representative of the range of possible patients), and (b) data bias resulting from the fact that available data is derived from systems that are biased in terms of access, diagnosis, and treatment of marginalized groups. We discuss these further below but also note here that beyond the scope of this article are other sources of bias that could result in healthcare-related AI discrimination such as inherent bias from new inputs (e.g., inaccuracies in existing medical devices for specific populations)[26] and the interpretation of an AI application's output(s) (e.g., giving too much deference to an AI application's output, which is known as automation complacency or automation bias).[27]

(a) Algorithmic Bias (Coding Bias & Machine Learning Bias)

Coding bias may result from a developer's weighting of an AI application's features or in their defining the AI application's definition of "success."[28] These

---

[25] See e.g. *Canadian Human Rights Act,* RSC 1985, c H-6; see also e.g. *Human Rights Code,* RSO 1990, c H.19.

[26] See Michael W Sjoding et al, "Racial Bias in Pulse Oximetry Measurement" (2020) 383:25 New Eng J Med 2477.

[27] See Parikh, *supra* note 3 at 2377.

[28] See Minssen, *supra* note 8 at 16. See also Davide Cirillo et al, "Sex and Gender

sources of bias may result from developers of AI applications inappropriately selecting proxy features during their development. Proxy features are those that an AI application uses as a surrogate to predict its "goal," for example, where an AI application uses a patient's socio-economic class to predict their risk of future cardiomyopathy. In the case of adaptive AIML, sometimes the algorithm is coded fine, but it makes its own inferences while it learns in order to produce an output (i.e., a social or medical construct, or "definition of success"), such as creditworthiness, risk of future cardiac disease based on genetic predispositions, or healthcare need. Algorithmic bias arises where we expect an AI application to tell us one thing, but it is *actually* telling us something else, especially for different groups.[29] For example, this could occur where an AI application uses postal codes to infer socio-economic class and predict patients' risk of future cardiomyopathy: the AI application may score two patients with identical postal codes the same, even though they may have drastically different healthcare needs and risk factors. This form of algorithmic bias then may result in discrimination in the diagnosis and treatment of marginalized patients. Other sources of algorithmic bias that can result in discrimination include choices as to how the algorithm is tested and validated.

(b) Data Bias

(i) Non-representative Training Data

Non-representative training data are overly homogenous data sets that do not adequately reflect the intended or likely patient population that will be treated using the AI application, or data sets that deliberately or inadvertently exclude significant subpopulations or relevant traits of an intended or likely patient population. For example, homogenous data sets may introduce contextual bias: if an AI application is designed using data from a high-income area, it may be less accurate when used in a low-income area[30] and fail to account for contextual factors such as significant differences in the rates of prescription drug coverage or access to paid sick leave. Contextual bias may also be introduced during the validation and testing of an AI application — to which some have suggested requiring cross-contextual validation and testing in multiple settings (e.g., different geographic regions with varying demographics).[31] Further, where an AI application is trained on data from a specific

---

Differences and Biases in Artificial Intelligence for Biomedicine and Healthcare" (2020) 3:1 NPJ Digital Medicine 1 at 4 [Cirillo].

[29] See Ziad Obermeyer et al, "Algorithmic Bias Playbook" (2021) at 1, online (pdf): *Chicago Booth: The Centre for Applied Artificial Intelligence* <www.chicagobooth.edu/-/media/project/chicago-booth/centers/caai/docs/algorithmic-bias-playbook-june-2021.pdf> [Obermeyer, "Algorithmic Bias Playbook"].

[30] See Minssen, *supra* note 8 at 17.

[31] See W Nicholson Price II, "Medical AI and Contextual Bias" (2019) 33:1 Harv JL & Tech 66 at 109-10.

subpopulation, for example, data that is predominantly White and male, the AI may be less accurate for other subpopulations with distinct considerations, contextual factors, or healthcare needs.

It follows that high quality and accurate AI applications in healthcare require robust, reliable, valid, and heterogeneous training data.[32] Unfortunately, many accessible data sets currently used to train AI applications in healthcare are relatively racially, geographically, or socio-economically homogenous[33] and therefore fail to adequately represent many vulnerable and marginalized patient populations. For example, MIMIC-III is a large, commonly used, publicly available database for the development of AI applications in adult critical care.[34] It provides critical care data from a large tertiary care hospital in Boston, Massachusetts.[35] In this example, an AI application derived from MIMIC-III data could be biased in favour of the socio-economically advantaged who were able to access such care, and against other patient subpopulations. We need to consider the generalizability of AI applications trained on its data to Canadian patients in rural areas,[36] Indigenous populations, racially diverse communities, lower income communities, or other communities (or patient subpopulations).

Narrow, homogenous data increases the likelihood that an AI application will better optimize successful outputs for well represented subpopulations compared to less represented subpopulations, including marginalized and vulnerable individuals. Prominent examples include some image-analyzing algorithms in dermatology that produce less accurate results for Black patients compared to White patients,[37] potentially compounding existing inequalities in skin cancer survival rates between these patient groups.[38] Similarly, some thoracic imaging-analyzing algorithms produce less accurate results for women — whose sex was underrepresented in the AI application's training data,

---

[32] See Gerke, *supra* note 11 at 302.

[33] See Sujay Nagaraj et al, "From Clinic to Computer and Back Again: Practical Considerations When Designing and Implementing Machine Learning Solutions for Pediatrics" (2020) 6 Current Treatment Options in Pediatrics 336 at 339 [Nagaraj]. Further, genetics research, for example, has historically been carried out on populations of European decent; see also Amy R Bentley, Shawneequa Callier & Charles N Rotimi, "Diversity and Inclusion in Genomic Research: Why the Uneven Progress?" (2017) 8:4 J Community Genetics 255 at 259. Biomedical research has, historically, failed to adequately include female patients and to account for gender-related physiological differences, see Cirillo, *supra* note 28 at 2.

[34] See Nagaraj, *supra* note 33 at 339. MIMIC-III has been used in the development of over 1300 AI applications in adult critical care.

[35] See Alistair EW Johnson et al, "MIMIC-III, a Freely Accessible Critical Care Database" (2016) 3:1 Scientific Data 1.

[36] See Nagaraj, *supra* note 33 at 344.

[37] See Adewole S Adamson & Avery Smith, "Machine Learning and Health Care Disparities in Dermatology" (2018) 154:11 J American Medical Assoc 1247 [Adamson].

[38] *Ibid*; see also Janice N Cormier et al, "Ethnic Differences Among Patients with Cutaneous Melanoma" (2006) 166 Archives Internal Medicine 1907.

compounding existing inequalities in cardio-thoracic diagnoses and treatments between sexes.[39] More recently, COVID-19 prompted a flood of new AI applications to provide insights into its clinical management, prevention, and disease severity likelihoods for certain patient populations, e.g., those with specific comorbidities.[40] Yet, one study that examined 226 COVID-19 AI models found that 97% were a high risk of bias, and with 42% of those the risk sprang from the fact that the training data was not representative of the intended patient population.[41]

Even assuming that representative data is available in principle, there is a separate risk that it might be curated[42] to such an extent that it becomes non-representative of certain patient subpopulations and therefore fails to account for real-world variables.[43] This is a concern, for example, where patients' traits in training data are limited to age, gender, and postal code, possibly due to their availability or accessibility. Race or ethnicity data may not be readily available but is needed to increase the accuracy of AIML applications.

(c) Upstream Data Bias

The masking of upstream data bias, and its subsequent incorporation into an algorithm's "label,"[44] may result from ongoing structural inequalities or from reliance on data that reflects outdated or discriminatory norms.[45] Consider, for example, a natural language processing application trained on two decades of physician notes, which have since been structured for the development of an AIML application. Such an application would likely produce less accurate results for non-binary and gender non-conforming individuals, considering the dominant historical use of the pronouns "he" and "she," as well as their use on, for example, patient intake forms. Also consider an AI application that uses

---

[39] Agostina J Larrazabal et al, "Gender Imbalance in Medical Imaging Datasets Produces Biased Classifiers for Computer-Aided Diagnosis" (2020) 117:23 Proceedings National Academy Sciences 12592.

[40] Casey Ross, "Machine Learning is Booming in Medicine. It's also Facing a Credibility Crisis" (2 June 2021), online: *Stat News* < www.statnews.com/2021/06/02/machine-learning-ai-methodology-research-flaws/ > .

[41] Laure Wynants et al, "Prediction Models for Diagnosis and Prognosis of Covid-19: Systematic Review and Critical Appraisal" (2021) 369 BMJ 1 at 6.

[42] Curating data involves the formatting, organizing, labelling, etc., of data. Highly curated data may advertently or inadvertently exclude important information such as certain patient features or other patient data that is incomplete. Further, the curating process itself can also introduce bias through, for example, the inappropriate labelling of data.

[43] See Rashidi, *supra* note 12 at 2.

[44] See Ziad Obermeyer et al, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations" (2019) 366:6464 Science 447 at 453 [Obermeyer, "Dissecting Racial Bias"].

[45] See Noel Sharkey, "The Impact of Gender and Race Bias in AI" (28 August 2018), online (blog): *Humanitarian Law & Policy* < blogs.icrc.org/law-and-policy/2018/08/28/impact-gender-race-bias-ai/ > .

frequency of healthcare visits as a proxy for disease progression in different patient populations. Without other important information, disparities in healthcare accessibility between groups that are entrenched in data sets would likely result in biased outputs.

Some data sets used to train healthcare-related AI applications have reflected existing systemic issues in healthcare, thus replicating these problems in those AI applications' outputs. For example, in the US, Epic Systems Inc. developed a predictive "no-show" model whereby facilities could double book patients deemed a high "no-show" risk to maximize clinical bookings.[46] However, if both patients showed up for their appointments, then each patient would be more likely to receive less care, in terms of time. This model generated predictive "no-show" values using information from patients' electronic health records, including ethnicity and socioeconomic class. And even when the implementing facility excluded these variables, the model still generated biased predictive values based solely on the levels of past healthcare access. The model failed to account for implicit discrimination in access to healthcare by different groups and, therefore, was more likely to double book already marginalized patients and further reduce their level of care.

Another US study, on a separate AI application in healthcare, found similar outcomes.[47] Like Epic's "no-show" model, this AI application's algorithm failed to account for unequal access to healthcare and compounded this inequality between White and Black patients.[48] The study examined racial bias in predictive health risk assessment scores, which health networks and insurers frequently use to identify patients that require additional care for better health outcomes and lower overall future expenditures. This study found that, comparatively, Black patients had more comorbidities than White patients, which should have resulted in increased interventions for this patient population, on average. Yet the algorithm incorporated bias by predicting these scores using patient-level healthcare costs rather than actual illness, without considering additional barriers to care for Black patients.

These sources of bias and the above examples of AI applications that have introduced, perpetuated, or exacerbated bias have resulted in discrimination, or may have resulted in discrimination if they were introduced into clinical practice, causing preventable risks to dignity and, in some cases, risks to patient access, health, and safety.

However, having raised concerns about algorithmic and data bias, it is important to underscore that appropriate incorporation into AI of data based on personal characteristics (such as sex, gender identity, race, and age) is both justifiable and *essential* for healthcare-related AI to serve all patients.[49] Consider,

---

[46] Sara G Murray, Robert M Wachter & Russell J Cucina, "Discrimination by Artificial Intelligence in a Commercial Electronic Health Record — a Case Study" (2020) 10 Health Affairs Blog.

[47] See Obermeyer, "Dissecting Racial Bias", *supra* note 44.

[48] *Ibid*. at 449-51.

for example, physiological or clinically meaningful differences between genders, such as known differences in response to cardiovascular treatments or differences in autism spectrum disorder symptomology between men and women.[50] In considering how best to prevent biased algorithms that result in *unwarranted* discrimination from a regulatory perspective, we must allow that AI applications *should* track clinically relevant connections between sensitive characteristics and appropriate care. The key, therefore, is for an AI application in healthcare to track sensitive characteristics where this is warranted and to thwart attempts to track these characteristics where it is not.[51] This generally requires significant interdisciplinary input into the algorithm's development (e.g., clinical, socio-economic, and epidemiological, to prompt the algorithm to distinguish between undesirable bias and distinctions that are clinically relevant)[52] and large amounts of representative and feature-rich data (i.e., of the intended or likely users).

As noted, many of these issues already plague human decision-making. But a single AI application in healthcare will have extended reach; it will be subject to a degree of automation complacency, and practitioners or other decision-makers may not always understand its underlying reasoning.[53] Further, such applications often advertently or inadvertently replace multiple human decision-makers, which may enable them to perpetuate or exacerbate bias at a greater frequency and magnitude compared to the status quo.

Even where developers design AI applications under ethical frameworks to reduce bias from methodological choices during an AI application's development, and where interdisciplinary teams account for and prompt an algorithm to distinguish between undesirable bias and distinctions that are clinically relevant, accurate and high-quality AI applications in healthcare may still be limited by non-representative (including incomplete) data. Ultimately, no amount of regulatory oversight or manipulation of data will correct for the basic problem of incomplete and non-representative data. This points to some fundamental questions, addressed in the next section: Why are so many AI applications in healthcare trained on non-representative data, and why are those sources non-representative in the first place?

---

[49] See Cirillo, *supra* note 28 at 2.

[50] *Ibid*. at 4.

[51] We acknowledge that this distinction is not always clear. Whether the tracking of characteristics by a healthcare-related AI application is inappropriate or clinically relevant will depend on several contextual factors. The hope is that interdisciplinary teams can develop algorithms that produce as close to objective outputs as possible. In many cases, however, this may be difficult to assess as current clinical care — AI's comparator or standard against which it is generally assessed — is often itself biased.

[52] See Cirillo, *supra* note 28 at 7.

[53] See Pasquale, *supra* note 15.

3. DATA DATA EVERYWHERE?

Most AI applications in healthcare are developed using data sets not primarily generated for AI research and development.[54] Sources often include electronic health records, healthcare utilization data, and biomedical research, which may be incomplete or non-representative due to data governance or systemic data issues, increasing the risk of healthcare-related AI bias resulting in discrimination.

(a) Data Governance

Researchers and developers often access healthcare data sets through secondary-use research frameworks, which generally focus on, among other protections, *minimizing* the scope of data shared between the data steward and the researcher, rather than *protecting* patients while enabling broader access.[55] The outdated ethos of *minimizing* data sharing is an obstacle to developing accurate and high-quality AI applications in healthcare — it belies the principle that more data will result in less biased and more accurate algorithms.[56] A different approach to data governance could liberate the collection, use, and disclosure of such data while engaging stronger accompanying protections against and penalties for inappropriate uses and the re-identification of de-identified data, which would allow for more exploratory uses and healthcare innovation while protecting patient privacy.[57]

Further, even though patient consent is generally not required under secondary-use frameworks across Canada if the data is de-identified,[58] barriers continue to exist when accessing such data. Privacy laws and their interpretation across Canadian provinces and territories vary to a degree.[59] Data stewards' interpretations of privacy laws may also be overly cautious (i.e., erring on the side of restricting access) to ensure their actions remain lawful.[60] Further, accessing high quality data can be challenging because different data stewards exist within jurisdictions, data standardization is inconsistent between

---

[54] See Melissa D McCradden, Elizabeth A Stephenson & James A Anderson "Clinical Research Underlies Ethical Integration of Healthcare Artificial Intelligence" (2020) 26:9 Nature Medicine 1325 at 1325 [McCradden].

[55] *Ibid.*

[56] See Gerke, *supra* note 11 at 302.

[57] See generally OECD, OECD Health Policy Studies, *Health Data Governance: Privacy, Monitoring and Research* (Paris: OECD, 2015) [OECD].

[58] See Council of Canadian Academies, "Accessing Health and Health-Related Data in Canada: The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation" (2015) at 80 [CCA].

[59] *Ibid.* at 30, 79, 135. See also Parminder S Raina et al, "Accessing Health Care Utilization Databases for Health Research: A Canadian Longitudinal Study on Aging Feasibility Study" 28:3 Can J on Aging 287 at 293 [Raina].

[60] See CCA, *supra* note 58 at 30.

jurisdictions,[61] special considerations exist for certain populations,[62] and the process to obtain data may be lengthy, complex,[63] and, therefore, costly.[64] A researcher seeking data from multiple provinces would have to comply with each province's requirements, and the data may still not be fully interoperable. Multi-jurisdictional data would better enable pan-Canadian AI applications, as it would not be specific to any one province's statistical distribution. Yet, while multi-jurisdictional portals are being developed,[65] interoperability challenges and incomplete data issues persist. A further question, and one outside the scope of this article, is if patients' data is being used to develop AI applications by commercial entities, are they entitled to some of the benefits?[66]

Within a province, such as Ontario, a researcher must apply to a data custodian for access to patient data, which must be accompanied by a research plan and research ethics board approval.[67] The *Personal Health Information Protection Act, 2004* also requires research ethics boards to consider relevant matters, including adequate privacy safeguards and the preservation of confidentiality,[68] and it requires the researcher to enter into a data sharing agreement with the data custodian[69] and to comply with any terms and

---

[61] See *ibid.* at 59.

[62] Including cultural sensitivities, as well as additional legal requirements when accessing Indigenous peoples' health care data. See, for example, culturally sensitive ethical data-gathering frameworks, such as the Black Data Governance Framework: Black Health Equity Working Group, "Engagement, Governance, Access, and Protection (EGAP): A Data Governance Framework for Health Data Collected from Black Communities" (2021), online (pdf): *Black Health Equity Working Group* <blackhealthequity.ca/wp-content/uploads/2021/03/Report_EGAP_framework.pdf>. Further, accessing Indigenous health data transcends Western "privacy" rights and requires unique governance frameworks, in accordance with Canadian and international law, as a necessary component of self-determination. Through these frameworks, a person seeking to collect or use Indigenous health data will have to work closely with the specific Indigenous community and its leaders. See First Nations Health Authority, the British Columbia Ministry of Health, and Indigenous Services Canada, "Data and Information Governance: Case Study Report" (2019) at 6, online (pdf): *First Nations Health Authority, the British Columbia Ministry of Health, and Indigenous Services Canada* <www.fnha.ca/Documents/FNHA-BC-Tripartite-Agreement-Case-Study-Data-and-Information-Governance.pdf>.

[63] Raina, *supra* note 59 at 291-93.

[64] See generally CCA, *supra* note 58.

[65] See e.g. the Health Data Research Network's Data Access Support Hub (DASH): <www.hdrn.ca/en/dash>, which aggregates harmonized data from multiple provincial and national sources for distributed analytics.

[66] Jacob L Jaremko et al, "Canadian Association of Radiologists White Paper on Ethical and Legal Issues Related to Artificial Intelligence in Radiology" (2019) 70:2 Can Assoc Radiologists J 107 at 113-14 [Jaremko].

[67] *Personal Health Information Protection Act, 2004*, SO 2004, c 3, Sched A, s 44 [*PHIPA*].

[68] *Ibid.*, s 44(3).

[69] *Ibid.*, s 44(5).

restrictions set out in that agreement.[70] This process by itself is not necessarily problematic, but where data is sought from multiple jurisdictions, compliance may become challenging. Regarding standardization, Canadian jurisdictions have a long history with challenges in adopting and implementing interoperable and consistent electronic health records.[71] Inconsistently structured health records exacerbate already subjective, unstructured information with a high propensity to capture bias.[72]

Further, there has been a shift towards collecting and aggregating more de-identified personal health information. However, governance in some areas, such as sensitive primary care settings, remains less clear. For example, healthcare providers, as data custodians, may be reluctant to release especially sensitive primary care records, even if de-identified. In response, some countries are implementing "opt out" structures, presuming patient consent.[73] While this model raises classic "ownership" issues, it also challenges common law privacy rights of intrusion upon seclusion[74] into patients' sensitive information, as methods of re-identification become more advanced. And in certain areas, such as radiology, inadequately processed digital imaging can easily be re-identified.[75]

Most provinces do prohibit the improper use of personal health data, which would include re-identifying individuals. Recently, Ontario[76] enacted a new statutory provision explicitly prohibiting the re-identifying of de-identified personal health data,[77] with fines of up to $200,000 or imprisonment for up to one year for natural persons and up to $1,000,000 for any non-natural persons.[78] However, a conviction requires the requisite *mens rea* (i.e., a *wilful* contravention), so this type of provision may not be sufficient to deal with

---

[70] *Ibid*., s 44(6).

[71] See generally Feng Chang & Nishi Gupta, "Progress in Electronic Medical Record Adoption in Canada" (2015) 61:12 Can Family Physician 1076; see also Amanda L Terry et al, "Gaps in Primary Healthcare Electronic Medical Record Research and Knowledge: Findings of a Pan-Canadian study" (2014) 10:1 Healthcare Policy 46.

[72] See Irene Y Chen, Peter Szolovits & Marzyeh Ghassemi, "Can AI Help Reduce Disparities in General Medical and Mental Health Care?" (2019) 21:2 AMA J Ethics 167.

[73] See Madhumita Murgia, "England's NHS Plans to Share Patient Records with Third Parties" (26 May 2021), online: *Financial Times* < www.ft.com/content/9fee812f-6975-49ce-915c-aeb25d3dd748 > [Murgia].

[74] See *Oliveira v. Aviva Canada Inc*., 2018 ONCA 321, 2018 CarswellOnt 4855 (Ont. C.A.), affirming *Oliveira v. Aviva Canada Inc. et al*, 2017 ONSC 6161, 2017 CarswellOnt 16264 (Ont. S.C.J.).

[75] Jaremko, *supra* note 66 at 112.

[76] Much of the discussion in Parts V and VIII regarding privacy and data governance legislation and initiatives primarily relates to Ontario; while we recognize that this is an Ontario-centric approach, Ontario is leading much of the law reform in this area in Canada.

[77] *PHIPA*, *supra* note 67, s 11.2 [amended by SO 2019, c 15, Sched 30, s 3].

[78] *PHIPHA*, *supra* note 67, s 72(1)(b.1) [enacted by SO 2019, c 15, Sched 30, s 7]; *PHIPHA*, *supra* note 67, ss 72(2)(a)-(b) [amended by SO 2020, c 5, Sched 6, s 23].

failures to implement best practices in AI design.[79] Further, public sector legislation, such as that in Ontario, also prohibits persons or entities, including inter- and extra-ministerial data integration units or ministries, from using or attempting to use information that has been de-identified in accordance with the *Freedom of Information and Protection of Privacy Act*.[80] However, the fines for a wilful breach of this prohibition, upon conviction, are very low, not exceeding $5,000.[81]

(b) Data Gaps

As discussed in Part III, even when developers can access large data sets, these may be non-representative or incomplete. For example, Canadian administrative data typically has limited race and ethnicity data.[82] Even where this data does exist it may be inconsistent. For example, a healthcare provider may use country of origin or religion to infer race or ethnicity, if they record it at all. Further, healthcare providers may discriminatorily misclassify patient interactions with the health system. For example, they may dismiss the concerns of Indigenous persons, inappropriately classifying an emergency room visit as alcohol abuse, as was the case with Brian Sinclair and many other Indigenous persons since.[83]

---

[79] Ontario also recently enacted provisions enabling the Commissioner to impose administrative monetary penalties to encourage compliance with the *Personal Health Information Protection Act, 2004* and its regulations or to "[prevent] a person from deriving, directly or indirectly, any economic benefit as a result of a contravention of [the] Act or its regulations." See *PHIPA*, *supra* note 67, s 61.1 [enacted by SO 2020, c 5, Sched 6, s 17].

[80] *Freedom of Information and Protection of Privacy Act*, RSO 1990, c F.31, s 49.8 [*FIPPA*] [enacted by SO 2019, c 7, Sched 31, s 6.)

[81] *Ibid*., s 61(1)(b.1) [enacted by SO 2019, c 7, Sched 31, s 8]; *ibi.d* , s 61(2).

[82] The COVID-19 pandemic brought significant attention to this issue. See Kwame McKenzie, "Socio-Demographic Data Collection and Equity in COVID-19 in Toronto" (2021) 34 EClinicalMedicine; see also Emily Thompson et al, "COVID-19: A Case for the Collection of Race Data in Canada and Abroad" (2021) 47 Canada Communicable Disease Report 300. Further, the most recent federal budget, Budget 2021, included funding to facilitate the collection of representative data: see Canada, Department of Finance, *Budget 2021* (Ottawa: Department of Finance Canada, 2021) at 230-31 [Budget 2021].

[83] See Jane Gerster, "A man was ignored to death in an ER 10 years ago. It could happen again", *Global News* (21 September 2018), online: < globalnews.ca/news/4445582/brian-sinclair-health-care-racism/ > ; see also Darren Bernhardt, "Sick Indigenous elder accused of being drunk at Winnipeg hospital, family says", *CBC News* (16 October 2020), online: < www.cbc.ca/news/canada/manitoba/racism-health-care-hospital-winnipeg-1.5765103 > ; see also Rhianna Schmunk, "B.C. investigating allegations ER staff played 'game' to guess blood-alcohol level of Indigenous patients", *CBC News* (19 June 2020), online: < www.cbc.ca/news/canada/british-columbia/racism-in-bc-healthcare-health-minister-adrian-dix-1.5619245 > .

Those designing AI applications for healthcare must be able to ethically access robust, diverse data to avoid overfitting[84] smaller data sets that may be non-representative or underfitting[85] incomplete data sets: data must be better collected from vulnerable and marginalized populations;[86] biomedical research must continue to increase its diversity; Canada needs standardized, multi-jurisdictional electronic data sources; and new approaches are needed for collecting, using, and sharing health data that balance privacy risks and promote sharing, with reinforced protections for patients. Trust, communication, transparency, and safeguards will play a major role achieving such a goal. Creating and enabling access to robust, representative data will require a cooperative, whole-of-government approach at both the provincial and federal level. Yet, as discussed, this raises one further key consideration: larger data sets, or the use of multiple data sets, increases the risk of "data triangulation,"[87] which could lead to re-identification.[88]

4. CANADIAN EX POST LEGAL MECHANISMS ARE INSUFFICIENT TO MITIGATE AI-RELATED DISCRIMINATION

An act of discrimination in a healthcare setting may ground a claim at common law in negligence, under federal and provincial human rights legislation, or under the *Charter of Rights and Freedoms* where the act involves government action or a private entity, such as a hospital, implementing or furthering a specific governmental policy or program.[89] While these mechanisms play important roles in deterring and compensating acts of discrimination, they constitute reactive "wait and see" approaches where, beyond their deterrent effects, the harms are likely to have already occurred. They also depend on a potential claimant's willingness to pursue such an action (and of course whether

---

[84] Overfitting may occur where an AI application becomes highly trained and, therefore, highly accurate for a specific patient subpopulation, all of whom are well represented in the training data — so much so that the AI application is not generalizable to patients underrepresented in the training data.

[85] Underfitting may occur where a lack of data leads to insufficient training and an inaccurate model for the specific patient subpopulation for whom the AI application is being trained.

[86] See Gerke, *supra* note 11 at 304.

[87] *Ibid.* at 317; see also David Thesmar et al, "Combining the Power of Artificial Intelligence with the Richness of Healthcare Claims Data: Opportunities and Challenges" (2019) 37:6 PharmacoEconomics 745 at 749.

[88] In *Gordon v. Canada (Minister of Health)*, 2008 FC 258, 2008 CarswellNat 522, 2008 CarswellNat 6510 (F.C.) at para. 34 the Federal Court agreed with the Intervenor that the test for identifying an individual (and, therefore, for anonymization) is where there is "a serious possibility that an individual could be identified through the use of [the] information, alone or in combination with other available information."

[89] *Eldridge v. British Columbia (Attorney General)*, 1997 CarswellBC 1939, 1997 CarswellBC 1940, [1997] 3 S.C.R. 624 (S.C.C.) at paras. 42-44.

they have the resources to do so), whether they are privy to the legal arrangement or have standing, and whether they can meet evidentiary bars.

In the context of healthcare-related AI, negligence may give rise to discrimination-based remedies. At common law, a person subject to discrimination that causes injury (e.g., a failure to properly diagnose or properly treat) may have a claim in negligence against the developer/manufacturer, a healthcare institution, or a healthcare provider. The risk of these kind of lawsuits may spur innovators to do their best not to bring biased healthcare-related AI applications to market. However, the claimant bears the evidentiary burden of proving that the defendant breached a standard of care, where a duty of care exists, as well as factual and legal causation, which involve significant time, resources, and risks — including judicial uncertainty.[90] Any claimant will face difficulties in establishing a breach of a standard of care, as well as factual and legal causation, where an algorithm is a "black box," (i.e., where its processes and risks are opaque to a practitioner, facility, or, even, the developer).[91] Product liability claims involving computer software are already challenging for claimants to establish:[92] for negligent design claims, claimants have to prove a defect in a software's design, the presence of an unreasonable risk for which a reasonable alternative design was available that would have reduced or eliminated the risk at the time, as well as causation between that defect and the claimant's injury.[93] In determining whether a reasonable alternative design existed at the time, courts will consider the risk-utility test,[94] which involves a number of factors — many of which require a significant understanding of the specific technology and its industry. This would be especially challenging for complex or state-of-the-art AI, AIML, and adaptive AIML applications.

A person subject to discrimination in a healthcare setting may also launch a complaint under human rights legislation.[95] However, human rights tribunals

---

[90] See Colleen M Flood & Bryan Thomas, "Canadian Medical Malpractice Law in 2011: Missing the Mark on Patient Safety" (2011) 86:3 Chicago-Kent L Rev 1053 at 1068-72.

[91] See Ian Kerr, Jason Millar & Noel Corriveau, "Robots and Artificial Intelligence in Health Care" in Joanna Erdman, Vanessa Gruben & Erin Nelson, eds, *Canadian Health Law and Policy*, 5th ed (Toronto: LexisNexis Canada, 2017) 257 at 272.

[92] See Jaremko, *supra* note 66 at 115.

[93] See George R Wray, Max Jarvie & Samantha Bonanno, "Artificial Intelligence and Product Liability: Catching up with the Future" (11 March 2021), online: *BLG* <www.blg.com/en/insights/2021/03/artificial-intelligence-and-product-liability>; see also Lisa R Lifshitz, "It's hard to sue a robot: product liability considerations and AI in Canada" (17 September 2018), online (blog): *Canadian Lawyer* <www.canadianlawyermag.com/news/opinion/its-hard-to-sue-a-robot-product-liability-considerations-and-ai-in-canada/275459>.

[94] See *Ragoonanan Estate v. Imperial Tobacco Canada Ltd.*, 2000 CarswellOnt 4613 (Ont. S.C.J.) at para. 103.

[95] For example, where an AI application discriminates in the provisions of services in Ontario based on a prohibited ground listed in section 1 of the *Ontario Human Rights Code,* RSO 1990, c H.19, the person may launch a claim under section 34 of the Code.

often involve delays for claimants seeking redress (generally because the system is backlogged due to insufficient resources),[96] low success rates in finding systemic discrimination,[97] low damage awards,[98] and evidentiary challenges.[99] A claimant bears the onus of establishing *prima facie* discrimination in, for example, the denial of services based on a prohibited ground. They must establish that discrimination was, more probably than not, a factor in the respondent's decision.[100] While courts may rely on factual inferences and social context, evidence must still be adduced that tangibly links discrimination based on a prohibited ground to the alleged conduct.[101] And claimants already face barriers; depending on the strength of the inference, courts may not take judicial notice of data indicating systemic discrimination — which an AI application may exacerbate as an adverse effect — such as census data or general research studies.[102] Regardless, if an AI application in healthcare were to rely on a sensitive attribute in making a decision without clinical justification, resulting in potential discrimination, it remains uncertain as to what level of influence this attribute must have on the AI application's output to ground a claim of discrimination. This uncertainty may arise in a situation where, for example, an AI application denied a patient referral, admission, or the provision or funding of care. Like negligence claims, it would be difficult for a claimant to prove that a facility, person, or organization discriminated in the provision of services based on a prohibited ground when using an AI application, especially where the algorithm is a "black box." Proper analysis of an algorithm's training data, development, and statistical analysis to generate sufficient evidence to ground a

---

96 The Canadian Human Rights Tribunal, for example, had 345 active complaints as of December 31, 2020. While the Tribunal experienced a net reduction in active complaints in 2016 and 2017 (outstanding complaints plus new complaints less closed complaints), it experienced net increases in active complaints in 2018, 2019, and 2020. See Canada, Canadian Human Rights Tribunal, *Annual Report 2020*, online (pdf): <www.chrt-tcdp.gc.ca/transparency/AnnualReports/2020-ar/2020-ar-en.pdf>.

97 See Jean-Simon Schoenholz, "Opening the Doors of Justice: Group Litigation and Claims of Systemic Discrimination" (2017) 48:2 Ottawa L Rev 687 at 713 [Schoenholz].

98 See Jeffrey Radnoff & Pamela Foy, "The Tort of Discrimination" (2002) 26 Advocates Q 309 at 314; see also Otto Ranalli & Bruce Ryder, "Undercompensating for Discrimination: An Empirical Study of General Damages Awards Issued by the Human Rights Tribunal of Ontario, 2000-2015" (2017) 13 JL & Equity 91 at 102-19.

99 See Schoenholz, *supra* note 97 at 702.

100 See *Pieters v. Peel Law Assn.*, 2013 ONCA 396, 2013 CarswellOnt 7881 (Ont. C.A.) at para. 83.

101 See Colleen Sheppard & Mary Louise Chabot, "Obstacles to Crossing the Discrimination Threshold: Connecting Individual Exclusion to Group-Based Inequalities" (2018) 96:1 Can Bar Rev 1 at 14-28 [Sheppard]. See also *Québec (Commission des droits de la personne et des droits de la jeunesse) c. Bombardier Inc. (Bombardier Aéronautique Centre de formation)*, 2015 SCC 39, 2015 CarswellQue 6297, 2015 CarswellQue 6298 (S.C.C.) at para. 88.

102 See Sheppard, *supra* note 101 at 30. See also *Kahkewistahaw First Nation v. Taypotat*, 2015 SCC 30, 2015 CarswellNat 1585, 2015 CarswellNat 1586 (S.C.C.) at para. 31.

successful claim may be an insurmountable task. Further, the availability of some of this information may be limited by private contracts, such as the terms within data sharing agreements.[103]

Discrimination claims under the *Charter of Rights and Freedoms* are also unlikely to provide claimants with sufficient redress. Like claims in negligence, section 15 equality claims require significant time and resources to litigate. Further, the claimant risks no redress or further loss (e.g., incurred costs), and they face significant evidentiary hurdles in proving a *prima facie* infringement.[104] And, again, evidentiary challenges would compound where an algorithm is a "black box."

As can be seen, then, ex-post legal mechanisms have significant limitations as the black box problem will make it difficult for plaintiffs to prove discrimination or negligence. Further, the barriers in accessing *ex post facto* justice to remedy AI-based discrimination would compound systemic bias and discrimination against many marginalized groups.

5. THE EXISTING *EX ANTE* LEGAL LANDSCAPE IN CANADA IS EVOLVING, BUT GAPS STILL PERSIST

In addition to laws that may provide a remedy after an injury has occurred, and which operate to deter the conduct leading to those harms, other laws attempt to directly regulate behaviour to prevent injuries *before* a harm occurs (*ex ante* laws). The existing *ex ante* landscape in Canada consists of a patchwork of federal and provincial laws and policies that are not tailored specifically to AI. While provinces have traditionally had jurisdiction over healthcare delivery, the Supreme Court of Canada has held that "health" is an amorphous topic subject to overlapping jurisdiction depending on the scope of the legislation.[105] The complexity and reach of AI in healthcare means that most AI applications will be, directly or indirectly, subject to a web of federal and provincial legislation and institutional policies and practices.

---

[103] Consider, for example, where a data custodian shares health data with a researcher under an agreement that explicitly prohibits the disclosure of data to any other party.

[104] See *Ewert v. Canada*, 2018 SCC 30, 2018 CarswellNat 2804, 2018 CarswellNat 2805 (S.C.C.) at para. 79, where a plaintiff's section 15 claim that an actuarial assessment tool used by Correctional Services Canada was discriminatory, based on an increased risk of inaccuracy when used for Indigenous persons compared to non-Indigenous persons, was dismissed. The Court noted that a mere increase in the risk of a tool's inaccuracy for a specific group is insufficient to constitute discrimination, and that a claimant would have to prove that the tool in fact overestimates (or underestimates) risk leading to a different outcome for the specific group.

[105] See *RJR-Macdonald Inc. c. Canada (Procureur général)*, 1995 CarswellQue 119, 1995 CarswellQue 119F, (*sub nom.* RJR-MacDonald Inc. v. Canada (Attorney General)) [1995] 3 S.C.R. 199 (S.C.C.) at para. 32, citing *Schneider v. British Columbia*, 1982 CarswellBC 241, 1982 CarswellBC 741, [1982] 2 S.C.R. 112 (S.C.C.) at 141-142 [S.C.R.].

Provincially, self-regulating entities, such as the College of Physicians and Surgeons of Ontario, establish licensing requirements and processes for membership therein, as well as ethical codes of conduct by which their members must abide. As a form of indirect *ex ante* regulation for the safe use of AI applications in healthcare, self-regulating (health practitioner) colleges could require a certain degree of digital health literacy as a fundamental competency within their licensing frameworks.[106] In addition, a myriad of provincial laws regulate public and private healthcare facilities. However, the direct *ex ante* regulation of many AI applications in healthcare falls within federal jurisdiction. The scope of this jurisdiction, and the extent to which it is relied upon in federal laws, dictates whether certain AI applications fall within Health Canada's *ex ante* oversight.

Federally, the *Food and Drugs Act*[107] (the Act) and the subordinate *Medical Devices Regulations*[108] (the Regulations), enacted under Parliament's criminal law power,[109] prohibit the selling of unsafe medical devices,[110] as well as the selling, labelling, packaging, treating, processing, and advertising of deceptive devices.[111] The Regulations further prohibit a person from selling or importing a medical device unless the following conditions are met: the manufacturer complies with the established *ex ante* product licensing scheme;[112] the person, if they are not the manufacturer, complies with the *ex ante* establishment licensing scheme;[113] the person complies with the Regulations' labelling requirements[114] and various post-market requirements as conditions of the licensing schemes. The existing *ex ante* product licensing scheme requires a manufacturer to submit, among other things, information demonstrating a medical device's safety and effectiveness (i.e., that it meets sections 10 to 20 of the Regulations) as a precondition for Health Canada to authorize the sale and importation of that device.[115] Historically, post-market requirements have included mandatory incident reporting by manufacturers and importers, who rely on voluntary reporting from patients, practitioners, and other persons. However, recent

---

[106] See Task Force, *supra* note 1 at 30, where the Task Force recommends digital health literacy as a fundamental competency in physician training.

[107] *Food and Drugs Act*, RSC 1985, c F-27 [*FDA*].

[108] *Ibid.*; *Medical Devices Regulations,* SOR /98-282 [*MDR*].

[109] *Constitution Act, 1867* (UK), 30 & 31 Vict, c 3, s 91(27), reprinted in RSC 1985, Appendix II, No 5.

[110] See *FDA*, *supra* note 107, s 19.

[111] *Ibid.*, s. 20.

[112] See *MDR*, *supra* note 108, ss 26-42. Except for Class I medical devices, which are still subject to, *inter alia*, quality standards prescribed by sections 10 to 20 of the Regulations.

[113] See *MDR*, *supra* note 108, ss 44-51.1. Establishment licences are generally required for persons selling or importing medical devices who are not the manufacturer and who do not possess a medical device licence; exceptions are set out in subsection 44(2).

[114] *Ibid.*, ss 21-23.

[115] See *MDR*, *supra* note 108, ss 32(2)(c), 32(3)(f), and 32(4)(i).

amendments to the Regulations require hospitals to report serious incidents[116] and authorization holders to report serious risks identified in foreign jurisdictions,[117] and they also enable the Minister to order an authorization holder to gather and submit information where other regulatory requirements are insufficient to manage significant uncertainties[118] — a shift towards more on-market oversight, or a "life-cycle" approach to regulation.

Importantly, the Act and the Regulations only capture medical devices that meet the definition of "device" in section 2 of the Act,[119] and the Regulations only prohibit the importing, advertising, and sale[120] of medical devices, unless the manufacturer holds a licence in respect of a medical device.[121] The definition of a device in the Act includes any of its components or accessories, and it captures devices intended to diagnose, treat, mitigate, or prevent a disease, disorder, or abnormal physical state or to restore, modify, or correct the body structure or the functioning of any body part. Health Canada currently interprets this definition to exclude from its oversight AI applications that merely relate symptoms and test results to best practices, or AI applications that support practitioner decision-making but stop short of near-term diagnosing or treating.[122] The Act and the Regulations do not contain any AI-specific provisions; however, "device" captures medical devices (hardware) with software and "software as a medical device," even if it is not associated with any hardware.[123] The Regulations provide a general *ex ante* safety and effectiveness framework, including for many AI and AIML-containing medical devices, but the rigidity of the rules presently seem to preclude the approval of adaptive AIML applications that continuously learn and change.[124] At the time of this writing, Health Canada has not yet approved any adaptive AIML applications

---

[116] See *FDA*, *supra* note 107, s 21.8 [enacted by SC 2014, c 24, s 5]; see also *MDR*, *supra* note 108, s 62 [enacted by SOR /2019-191, s 1].

[117] See *MDR*, *supra* note 108, s 61.2 [enacted by SOR /2020-262, s 16].

[118] See *FDA*, *supra* note 107, s 21.32 [enacted by SC 2014, c 24, s 4]; see also *MDR*, *supra* note 108, s 62.2 [enacted by SOR /2020-262, s 17].

[119] See the definition of "sell," *FDA*, *supra* note 107, s 2 [amended by SC 2019, c 29, s 163].

[120] See *FDA*, *supra* note 107, s 2 [amended by SC 2019, c 29, s 163].

[121] See *MDR*, *supra* note 108, ss 26, 27, 44, and 80(1).

[122] See Canada, *Guidance Document: Software as a Medical Device (SaMD): Definition and Classification* (Guidance Document), online: < www.canada.ca/en/health-canada/services/drugs-health-products/medical-devices/application-information/guidance-documents/software-medical-device-guidance-document.html > (Ottawa: Health Canada, 2019).

[123] *Ibid*.

[124] An adaptive AIML application would not likely meet current validation standards (as required by section 20 of the *Medical Devices Regulations*), nor would a manufacturer know with certainty when an adaptive AIML application changes significantly, requiring an amendment (as required by section 34 of the *Medical Devices Regulations*). Significant changes are changes that "could reasonably be expected to affect the safety or effectiveness of a medical device," as defined in section 2 of the *Medical Devices*

— but the question remains, should they approve such applications, and how should they approach the challenge of regulating technology which will change over time? It should also be noted that the definition of "sell" in the Act was recently extended to include any lease arrangement,[125] which has important implications for AI because software is not always sold (distributed), but sometimes is leased. However, more importantly, because the prohibitions in the Regulations are on the "sale" of a medical device, there is presently no prohibition on practitioners' developing adaptive AIML applications in-house and using them on their patients, absent any kind of selling, including distributing or leasing, or conducting a "clinical trial" involving human subjects.[126]

In 2019, the Act was amended, providing for an Advanced Therapeutic Products (ATP) pathway.[127] An "advanced therapeutic product" is a "therapeutic product" or "class of therapeutic products" described in Schedule G of the Act — to which, or from which, the Minister may add or remove ATPs or classes of ATPs by order.[128] Overall, if leveraged, the pathway would provide regulatory oversight for healthcare-related adaptive AIML that meets the definition of "device,"[129] but it would also allow for considerably more discretion for the regulator in how to treat new and novel technologies seeking licensure.[130] Health Canada is now considering the use of the ATP pathway to create a regulatory pathway for adaptive AIML, the prohibitions of which would extend beyond sale or lease and would include, for example, the manufacture or testing of an adaptive diagnostic imaging AIML application for use[131] within a particular hospital for its own purposes.

Health Canada could use the ATP pathway (or medical device licensing requirements more broadly) to help address (some) problems of algorithmic bias. For example, it could require as part of the licensing process evidence from innovators as to the representativeness and appropriateness of the training data. However, given the difficulties AI innovators have in accessing data that we

---

Regulations, and include changes to its performance characteristics, principles of operation and specifications of materials, and software.

[125] See *FDA*, *supra* note 107, s 2 [amended by SC 2019, c 29, s 163].

[126] *Ibid*., s 2, see the definition of "clinical trial," and s 3.1 [enacted by SC 2019, c 29, s 166].

[127] *Ibid*., ss 21.9-21.96 [enacted by SC 2019, c 29, s 169].

[128] *Ibid*., ss 21.91(1) and (3) [enacted by SC 2019, c 29, s 169].

[129] *Ibid*., s 2 [amended by SC 2019, c 29, s 163].

[130] The ATP pathway in the Act provides a flexible oversight mechanism for "novel, complex, and distinct" therapeutic products, including those that are personalized, developed at the point of care, and manufactured in 'non-traditional' ways. See Canada, "Regulatory innovation for health products: Enabling advanced therapeutic products" (02 February 2021), online: *Health Canada* <www.canada.ca/en/health-canada/corporate/about-health-canada/activities-responsibilities/strategies-initiatives/health-products-food-regulatory-modernization/advanced-therapeutic-products.html>.

[131] See *FDA*, *supra* note 107, s 21.9(1) and (2) [enacted by SC 2019, c 29, s 169].

highlighted earlier, it may not always be possible to demand of innovators that training data be representative. In these circumstances Health Canada could consider the use of advisory product labels for healthcare providers, warning them of this risk. This is of course far from a perfect solution but will at least put the issue squarely before the healthcare professional rather than leaving them to assume that the AI innovation must be safe/appropriate for all patients because it has been licensed.

More generally, Health Canada is considering how best to further move towards a life-cycle approach for regulation, monitoring medical devices in the real world through, for example, the imposition of terms and conditions on medical device licences, which may obligate licence holders to submit certain information periodically.[132] This is of course essential if it is to provide regulatory oversight of adaptive AIML, which by its very definition is designed to evolve over time. A life-cycle approach provides a platform for Health Canada to collect and respond to evidence of algorithmic bias from real-world experience, at least bias that results in safety concerns for patients. Again, we do not wish to convey this is a perfect solution, as it may be difficult for providers and patients to detect and, therefore, report problems of algorithmic bias with non-adaptive or adaptive AIML applications. In this regard, the regulator may find help for post-market oversight from employing its own secondary software or algorithms where data could be aggregated and analyzed in near real time.

We also note the need for coordination by federal and provincial governments with efforts made at the federal level to regulate to prevent algorithmic bias in healthcare-related AI. Health Canada's jurisdictional authority would not protect against bias and discrimination among AI applications in healthcare that harm dignity without compromising safety, nor would it cover AI applications in healthcare that fall outside of Health Canada's purview, such as patient scheduling applications, system-level insight or pattern recognition applications, or insurer risk-assessment applications. The provinces and territories may have to address these issues, in some capacity, by modernizing various statutes and regulations — both *ex post* and possibly novel *ex ante* legal mechanisms — to better protect people from AI-related bias and possible discrimination in healthcare and in private healthcare-related contractual settings.

6. RECOMMENDATIONS AND NEXT STEPS

Accurate and high-quality AI applications in healthcare require: (1) robust, complete, and representative training data wherever possible; (2) algorithms that

---

[132] See Canada, Health Canada, *Forward Regulatory Plan 2021-2023: Regulations amending the Medical Devices Regulations (Agile Regulations)* (last modified 6 April 2021), online: < www.canada.ca/en/health-canada/corporate/about-health-canada/legislation-guidelines/acts-regulations/forward-regulatory-plan/plan/exemptions-certain-digital-health-software.html > .

do not enable or support bias on the basis of arbitrary traits, such as race and sex, with no clinical, moral or legal justification; and (3) strong development standards to protect against bias throughout the entire lifecycle of an AI application (i.e., the application's conception, design, and implementation, with appropriate feedback mechanisms at each stage). To achieve these objectives, governments and innovators should take several steps.

(a) Federal Safety Regulation of Algorithmic Bias

Due to the pervasiveness and potential effects of algorithmic and data biases — such as inaccurate predictions or interventions — Health Canada should treat potential bias in healthcare-related AI applications as a safety issue.[133] Regulators should require as part of licensure the disclosure of an algorithm's training data, objective function, and prediction methodology,[134] as well as specific analysis as to the representativeness of an AI application's training data. Where developers are unable to access robust, complete, and representative training data, they should demonstrate serious due diligence in trying to access such data before relying on synthetic approaches to fill any such gaps. Consideration could be given to requiring algorithmic audits from certified third parties as a precondition for approval, with or without the assignment of a performance classification, which should include its risk of bias, and which should appear on the product label. Where disclosure regarding the AIML's algorithm (e.g., a "black-box") or its data sources is not possible, Health Canada could rely on established clinical research ethics frameworks (i.e., clinical trials) to fill certain evidentiary gaps.[135]

Where there are algorithmic or related data bias risks, Health Canada should consider providing warnings as to the risks of unreflective application by a healthcare professional of healthcare-related AI to an at-risk population. Our suggestion for product label warnings would we hope nudge developers to critically examine the representativeness of an AI application's training data and the fairness of an algorithm itself. However, if too liberally employed, such warnings could become largely meaningless and provide developers with an "out" to download responsibility and liability for off-label use on practitioners or facilities.

Movements towards a lifecycle regulatory approach, essential for appropriate governance of adaptive AIML, also offers the prospect of collecting better real-world data and opportunities to identify and respond to manifestations of algorithmic bias. We do not discount the enormous regulatory challenge here of requiring a highly iterative lifecycle "validation" process,[136] or

---

[133] See Gerke, *supra* note 11 at 304.

[134] See Obermeyer, "Dissecting Racial Bias", *supra* note 44 at 447.

[135] See Gerke, *supra* note 11 at 303; see also McCradden, *supra* note 54 at 1325.

[136] Effy Vayena, Alessandro Blasimme & I Glenn Cohen, "Machine Learning in Medicine: Addressing Ethical Challenges" (2018) 15:11 PLoS Medicine 1 at 3.

even the need for use by the regulator of secondary software to monitor continuous statistical outcome metrics, flag issues, and monitor safety-related bias and other safety and efficacy issues.

Overall, by viewing bias as a safety concern, Health Canada should push innovators to use robust, complete, and representative training data and to establish strong internal frameworks for developer-led bias analysis.

(b) Provincial and Additional Federal Mechanisms to Mitigate Bias

In addition to incorporating digital literacy into practitioner licensing and competency frameworks, practitioners should familiarize themselves with any available audits of AI applications they intend to use in their practice. Further, hospitals and other purchasers should require and review such audits in advance of purchasing and reimbursement decisions. For any healthcare-related AI application that falls outside of Health Canada's purview, payers or users should demand robust third-party analysis of the algorithm and its training data. Certification standards and third-party audits could be leveraged for AI devices and applications that fall within or outside of Health Canada's jurisdiction.

Provincial governments, as well as the federal government, should also modernize discrimination-based human rights legislation to better remedy AI-related bias and its resulting harms, especially where complex algorithms, including "black box" algorithms and adaptive AIML, will pose additional evidentiary hurdles for such claims. However, modernized legislation should be accompanied by significant government funding increases into already underfunded human rights commissions and agencies. Such funding increases could modernize oversight, including for research and advice functions. Provincial and federal governments should also clearly articulate best de-identification practices and support organizations seeking to develop such practices.[137]

Recently, Ontario released a White Paper inviting consultation on its plans to introduce a new private sector data protection law for Ontario. The White Paper specifically addresses automated decision systems. It indicates that one option is to "prohibit the use of AI and automated decision-making systems when they could cause harm to citizens."[138] It also indicates that a goal of reforms could be "to inform Ontarians when and how their data is used by these technologies, and empower them with a right to object to these uses, or at least to contest them."[139] Quebec's *Act to modernize legislative provisions as regards the protection of personal information*,[140] which has reformed that province's public

---

[137] For example, see CANON | Canadian Anonymization Network, online: < deidentify.ca/ > .

[138] Ontario, *Modernizing Privacy in Ontario: Empowering Ontarians and Enabling the Digital Economy* (White Paper) (Toronto: Government of Canada, 2021), online: < www.ontariocanada.com/registry/showAttachment.do?postingId = 37468&attachmentId = 49462 > at 11.

[139] *Ibid.*

and private sector data protection laws, also includes certain rights with respect to automated decision systems, including a version of a right to an explanation of the decisional system. It also provides a right to submit observations to someone within the organization who is in a position to review the decision, although it stops short of an actual right to demand that the decision be reviewed by a human. Bill C-11,[1410] the federal bill to reform Canada's private sector data protection law also provided for a right to an explanation of automated decision systems, including a right to be informed of how the data used to make the decision was obtained. It should be noted that the concept of a decision in Bill C-11 was quite broad and included "any technology that assists or replaces the judgment of human decision-makers using techniques such as rules-based systems, regression analysis, predictive analytics, machine learning, deep learning and neural nets."[142] Although these are examples of actual or proposed reforms to private sector data protection laws, and not personal health information protection laws, they do demonstrate the extent to which rights to an explanation of automated decision-systems may become a key part of data protection legislation not just in the private sector but in the public and health sectors as well.

Finally, we note the possibility of the Federal government using its criminal law powers to provide penalties for discrimination resulting from biased healthcare-related AI products (or indeed more broadly).[143] Such a proposal obviously requires an in-depth treatment of division of powers questions as well as a deeper interrogation of the interaction of such a law with other provincial and territorial laws and policies (including the recent initiatives underway in Ontario and Quebec), and we plan to address this possibility in a sibling paper.

(c) Innovators: Transparency & Trust

Regardless of specific regulatory requirements, innovators — best positioned to mitigate algorithmic bias — should publicly disclose the kinds of training data they use, any real and potential gaps in data, and actual or potential bias of an AI application.[144] Innovators should also include highly interdisciplinary teams in the development of an algorithm, to strive to discern between, and prompt an

---

[140] See *An Act to modernize legislative provisions as regards the protection of personal information*, CQLR 2021, c 25. Note that the coming into effect of the provisions of this new law is phased in over a two-year period.

[141] See Bill C-11, *An Act to enact the Consumer Privacy Protection Act and the Personal Information and Data Protection Tribunal Act and to make consequential and related amendments to other Acts*, 2nd Sess, 43rd Parl, 2020, cl 62(2)(c) and 63(3) (first reading 17 November 2020). Note that this Bill was not passed before the call of the 2021 federal election, and it is unclear whether it will be reintroduced as is, revised, or replaced by any incoming government.

[142] *Ibid.,* see the definition of "automated decision system" in cl 2.

[143] See *Reference Re Genetic Non-Discrimination Act*, 2020 SCC 17.

[144] See Gerke, *supra* note 11 at 302-03.

algorithm to recognize, undesirable bias and distinctions that are clinically relevant. Innovators must also maintain rigorous internal development processes that include ongoing audits and checks for algorithmic bias.[145] Such testing must be conducted using sensitive-feature comparator groups.[146] But, again, this would require sufficient representative data for these groups, including the sensitive feature or patient trait in question.

Innovators will bear a high degree of moral responsible over the next few decades of AI development and implementation. Further breaches of the public's confidence, in the context of AI in healthcare, could significantly undermine the gathering and accessibility of robust, representative, and complete data. And high-profile safety issues, which may arise from algorithmic bias, could create major barriers for the development and implementation of AI applications in healthcare, potentially through strict, reactive regulations by both federal and provincial governments.

(d) Addressing Non-Representative Data and the Inaccessibility of Large, Diverse Data Sets

In order to ensure that there is sufficient high-quality data for AI innovation in healthcare, provincial and federal governments must implement new mechanisms to incentivize the generation and aggregation of health data for sharing purposes. Without effective data gathering and sharing mechanisms, Canada will fall behind other nations. Patients will not realize the benefits of AI applications in healthcare, they will be more likely to suffer additional bias-related harms, and Canada's overall economic prosperity will also suffer. While many significant concerns have been raised regarding the collecting of race-based data,[147] if we do not collect this type of data, we risk allowing bias and its harms to propagate in AI applications in healthcare, and we also risk limiting the benefits to already inequitably advantaged individuals. There are some signs that governments in Canada are beginning to address issues around access to high quality data. The most recent federal budget, for example, allocated funding to enhance the creation of data standards through the Standards Council of Canada. It also contained money to establish the role of Data Commissioner.[148] Although the exact nature of this role remains unclear, it could be similar to Australia's Data Commissioner[149] or New Zealand's Data Steward.[150]

---

[145] See Obermeyer, "Algorithmic Bias Playbook", *supra* note 29 at 19.

[146] *Ibid*. at 5.

[147] See LLana James, "Race-based COVID-19 data may be used to discriminate against racialized communities" (14 September 2020), online: *The Conversation* < theconversation.com/race-based-covid-19-data-may-be-used-to-discriminate-against-racialized-communities-138372 > .

[148] Budget 2021, *supra* note 82 at 230-31.

[149] See Australia, Office of the National Data Commissioner, "The Data Availability and Transparency Bill has been introduced to the Australian Parliament", online: < pmc.gov.au/public-data/national-data-commissioner > .

Leadership of this kind to develop the appropriate frameworks for data sharing would be invaluable.

The provinces should also act with respect to their own stores of data. In Ontario, the government's Digital Strategy[151] signals the intention to create a new Data Authority, which will be charged with "building modern data infrastructure to support economic and social growth at scale, while ensuring that data is private, secure, anonymous and cannot identify people individually."[152] In addition, it seems likely that Ontario will build out its first extra ministerial data integration unit, the Ontario Health Data Platform (OHDP), giving it greater scope. The OHDP was created following amendments to *PHIPA* during the COVID-19 crisis.[153] It was built to serve as a data sharing facility that could combine data from multiple sources, including health, public, and private sectors. The OHDP may well see its mandate extended beyond the pandemic as a new health data sharing infrastructure for the province.[154]

Federal and provincial governments must also continue to work closely with Indigenous governments and communities to advance ethical data sharing structures that respect Indigenous data sovereignty and Indigenous communities' privacy rights and cultural sensitivities. In doing so, governments must ensure that AI applications in healthcare developed from such data specifically benefit Indigenous communities, especially those that elect to share their health data — reciprocity is key. Governments should also continue to develop ethical data governance frameworks for other culturally sensitive populations.

If governments consider more proactive approaches in gathering and releasing large healthcare data sets for AI development, including for commercial use,[155] they should also consider enhanced protections for patients. Federal and provincial privacy legislation could provide additional protections for patients by explicitly prohibiting the re-identification of healthcare data and backing such a prohibition with significant fines and penalties, where legislatures and governments have not already done so.[156] Governments could also leverage public-private partnerships, and greater

---

[150] See New Zealand, "Government Chief Data Steward" (last updated 17 September 2020), online: < www.digital.govt.nz/digital-government/leadership/government-functional-leads/government-chief-data-steward-gcds/ > .

[151] See Ontario, "Building a Digital Ontario" (30 April 2021, last modified 4 June 2021), online: < www.ontario.ca/page/building-digital-ontario > .

[152] *Ibid.*

[153] Ontario Health Data Platform, "Ontario Health Data Platform COVID-19" (last modified 17 September 2020), online: < ohdp.ca > .

[154] Ontario, "Ontario Appoints Special Advisor to Develop Health Data Platform" (4 June 2020), online: < news.ontario.ca/en/release/57110/ontario-appoints-special-advisor-to-develop-health-data-platform > .

[155] England plans to share patient records with third parties, including commercial organizations. See Murgia, *supra* note 73.

[156] See the discussion on data governance in Part III.

federal-provincial-territorial cooperation on standards,[157] access, interoperability, etc., to build large databases for AI development, in a transparent manner, to build trust while providing assurances or some public commitment regarding the benefits of any resulting AI applications.

CONCLUSION

In this article we have considered the problems of bias and the discrimination that can result from healthcare-related AI. These are complex problems that range from flaws in algorithm design to problems of insufficient or deficient data. We have explained how these problems of bias can arise, and how they might result in discriminatory treatment or exacerbate existing discrimination within the healthcare system. Yet AI applications in healthcare will continue to advance, and we must take advantage of their significant benefits. To drive the ethical development of AI applications in healthcare, to minimize harms, and to ensure the just distribution of their benefits, we must respond to these many significant problems through broad, multi-dimensional, concerted approaches.

An important solution at the federal and provincial level involves ensuring that innovators have access to high quality, robust, and representative data sets. We have highlighted some of the existing challenges with access to data within the Canadian healthcare system, as well as some of the initiatives underway to address these deficiencies. Much work remains to be done in this area, as initiatives are currently in the very early stages, and they only exist in some jurisdictions and not others (leading to further fragmentation). Solutions involving access to data must take into account privacy considerations, as well as those relating to Indigenous data sovereignty. In Canada, fragmented jurisdiction over healthcare issues combined with a history of poor cross-jurisdictional standardization has created a context in which the available data are often difficult to access and poorly interoperable. Much work needs to be done to enhance data quality, data interoperability, and access to high-quality data through innovative data sharing frameworks.

---

[157] For data standards for data sharing, see Michael Girard, "Global Standards for Digital Cooperation" (28 October 2019), online: *Centre for International Governance Innovation* www.cigionline.org/articles/global-standards-digital-cooperation/>; for a Canadian framework for data reuse, see Michael Girard, "A Canadian Framework for Data Reuse" (26 April 2021), online: *Centre for International Governance Innovation* <www.cigionline.org/publications/canadian-framework-data-reuse/>; for a data standards task force, see Michael Girard, "Data Standards Task Force for Digital Cooperation" (11 August 2020), online: *Centre for International Governance Innovation* <www.cigionline.org/publications/data-standards-task-force-digital-cooperation/>; for standards for the digital economy, see Michael Girard, "Standards for the Digital Economy: Creating an Architecture for Data Collection, Access and Analytics" (4 September 2019), online: *Centre for International Governance Innovation* <www.cigionline.org/publications/standards-digital-economy-creating-architecture-data-collection-access-and-analytics/>.

We have also looked at existing legal frameworks for *ex post* and *ex ante* responses to bias and discrimination in AI applications in healthcare. *Ex post* mechanisms are frequently unsatisfactory because they activate only after harm has occurred, and they require considerable effort and resources on the part of individuals who have suffered harms. Further, litigating algorithmic bias and discrimination will be challenging because of causation problems and difficulties in accessing algorithms and their training data. In our view, *ex ante* solutions (ones that prevent the harm) are more promising than *ex post* mechanisms, both for patients and innovators. However, they will require existing laws (such as those regulating medical devices) to be better tailored to the AIML and adaptive AIML context. A system for licensing healthcare-related AI should take into account the problems of bias and discrimination in order to prevent these issues from manifesting in the first place. Further, provinces may have to consider novel *ex ante* approaches to prevent AI-based bias and discrimination in healthcare for AI applications that that fall outside federal jurisdiction.

In some regards, Canada is well positioned to lead effective and equitable AI development and implementation in healthcare, such as its expertise among AI researchers and research institutions and its diverse society.[158] However, in order to do so, federal and provincial governments must take steps — ideally with some degree of cooperation — to address the significant and intersecting problems that we have identified. Properly developed AI applications have the potential to mitigate extant bias and discrimination, but such development requires intervention both to create optimal conditions for development and to ensure, through a combination of regulation and recourse, that high expectations are met. We cannot risk creating new sources of bias or perpetuating or exacerbating extant, and often concealed, discrimination in healthcare.

---

[158] See Adam Kassam & Naila Kassam, "Artificial Intelligence in Healthcare: A Canadian Context" (2020) 33 Healthcare Management Forum 5 at 7.