RESEARCH PAPER

Regulating the Safety of Health-Related Artificial Intelligence

Réglementer la sécurité de l'intelligence artificielle dans les soins de santé

MICHAEL DA SILVA, SJD
Lecturer
University of Southampton Law School
Southampton, UK
Senior Fellow in AI and Healthcare
AI + Society Initiative
Centre for Law, Technology and Society
University of Ottawa
Ottawa, ON

COLLEEN M. FLOOD, SJD
Professor
University Research Chair in Health Law & Policy
Director of the Centre for Health Law, Policy and Ethics
Faculty of Law (Common Law Section)
University of Ottawa
Ottawa, ON

ANNA GOLDENBERG, PHD
Senior Scientist
SickKids Research Institute
The Hospital for Sick Children
Associate Professor
Computer Science
University of Toronto
Associate Research Director, Health
Vector Institute
Fellow
Canadian Institute for Advanced Research (CIFAR)
Toronto, ON

DEVIN SINGH, MD
Staff Physician and Lead for Clinical Artificial Intelligence & Machine Learning
Division of Paediatric Emergency Medicine
The Hospital for Sick Children
University of Toronto
Toronto, ON

Abstract

This article analyzes whether Canada's present approach to regulating health-related artificial intelligence (AI) can address relevant safety-related challenges. Focusing primarily on Health Canada's regulation of medical devices with AI, it examines whether the existing regulatory approach can adequately address general safety concerns, as well as those related to algorithmic bias and challenges posed by the intersections of these concerns with privacy and security interests. It identifies several issues and proposes reforms that aim to ensure that Canadians can access beneficial AI while keeping unsafe products off Canadian markets and motivating safe, effective use of AI products for appropriate purposes and populations.

Résumé

Cet article cherche à savoir si l'approche actuelle du Canada en matière de réglementation de l'intelligence artificielle (IA) dans les soins de santé permet de relever les défis pertinents en matière de sécurité. Se concentrant principalement sur la réglementation de Santé Canada pour les dispositifs médicaux qui intègre l'IA, l'article cherche à savoir si l'approche réglementaire en place peut répondre adéquatement aux préoccupations générales en matière de sécurité, ainsi qu'à celles liées aux biais algorithmiques et aux défis sous-jacents en matière de confidentialité et de sécurité. L'article repère plusieurs problèmes et propose des réformes qui visent à garantir l'accès à une IA bénéfique tout en gardant les produits dangereux hors des marchés canadiens et en motivant une utilisation sûre et efficace des produits d'IA pour les groupes appropriés.

Introduction

Artificial intelligence (AI) refers to machines (including software) that perform functions that normally require human cognition without direct human aid (WHO 2021). The use of AI in healthcare settings engenders fierce debates as to its benefits and risks. Health-related AI offers possibilities of more accurate diagnoses, more finely tailored treatment plans, shorter wait times, and greater patient involvement in directing care (CIFAR 2020, 2021; CMA 2020; Reznick et al. 2020). Yet, at the same time, health-related AI raises concerns, for example, about how to ensure that an AI product is both itself safe and used safely by healthcare professionals; how to apportion liability for harm as between manufacturers, professionals, and institutions; what informed consent should entail in AI settings; the threat of algorithmic bias/discrimination; and potential AI-related privacy and security breaches (Blasimme and Vayena 2020; Flood and Régis 2021; Gerke et al. 2020; Murphy et al. 2021). The great challenge, then, concerns how to ensure the benefits of AI while eliminating (or at least limiting) any adverse outcomes of AI use. To meet this challenge, we need sophisticated decision making by humans, particularly with respect to choices as to how to regulate health-related AI.

There are layers of uncoordinated laws that will (often indirectly) impact how AI is used in the Canadian healthcare system (or systems, given provincial legal variations). Regulations include "hard" laws – such as provincial laws regulating hospitals and the operation of safety committees therein, privacy laws and anti-discrimination laws – and "soft" laws – such as hospital accreditation standards, as well as clinical guidelines and educational standards set by professional colleges. There are also incentives flowing from tort law and contract law; over time, judges will develop a body of case law that will help guide how to apportion liability should a patient be harmed as a result of the use (or non-use) of health-related AI (Flood and Régis 2021).

In this paper, we consider one location of regulation of health-related AI, namely, Health Canada. We look at Health Canada's role in approving medical devices as sufficiently safe for sale within Canadian markets. We take this approach as regulation at the federal level

impacts healthcare systems, healthcare providers, patients and AI innovators across Canada: reform here offers an important window of opportunity for national rules/policies. Health Canada also recently adopted a new regulatory pathway to better permit the licensing of health-related AI, including adaptive machine learning (explained further later). As it is presently seeking some input on how best to operationalize this new regulatory pathway (Health Canada 2021, e-mail communication, June 21, 2021), we hope our analysis can inform that process. Likewise, while health-related AI implicates a broad range of important legal concerns, here we focus on product safety – i.e., safety of a given health-related AI tool. However, product safety is entwined with the problem of algorithmic bias and, more indirectly, by privacy/data access issues, so we include these issues within our analysis.

Thus, in what follows, we outline three safety-related concerns (general safety, algorithmic bias and privacy/security) specific to Health Canada's regulation of medical devices. For each, we discuss how existing regulation may address the issue and offer recommendations regarding regulatory reform as needed.

(General) Safety of Health-Related AI

The concern(s)

Existing evidence from research settings suggests that many health-related AI tools outperform their human counterparts (Topol 2019a, 2019b). However, AI will make mistakes. For instance, IBM Watson for Oncology was not trained on real patient data and made erroneous treatment recommendations (Gerke et al. 2020: 302–03). Errors were identified before it was implemented in clinical practice, but developers did not disclose the problems for over a year (Gerke et al. 2020). More recently, the Epic Sepsis Model designed to predict sepsis cases based on electronic health records was found to miss two-thirds of cases (Wong et al. 2021). One hopes that such examples will be rare. But for AI to spark a seismic shift toward improved safety of healthcare, professionals, patients and the public must be able to trust that regulatory systems will weed out unsafe AI.

The last decade saw a rapid expansion of machine learning (ML), a subfield of AI that now powers most AI applications across industry. ML models learn from data and are designed to improve their performance over time from new data without being explicitly (re-) programmed. Adaptive ML that can evolve and continuously learn from real-world health data, as opposed to remaining static, is a potential game changer in terms of safety. It offers a major advantage over other technologies, such as a drug or a hard medical device that does not improve itself over time. Unlike other tools, ML might realize its errors and correct its performance issues. But, at the same time, this aspect of ML creates real challenges: if an ML tool is "approved" as safe prior to entry to the market but then evolves, how can a regulator guarantee that it remains safe for use as it evolves?

Presently, regulation of medical devices assumes that AI software is of lower risk if there is a "human in the loop"; that is, if a healthcare professional oversees the use of AI.

However, many scholars raise concerns that algorithms are "black boxes": healthcare professionals cannot know how a particular diagnosis or treatment recommendation has been reached (Pasquale 2015; Price 2018). Some of these concerns may be overstated given the status quo. Physicians may not, for example, fully understand different clinical interventions (e.g., common medications such as acetaminophen, for which the mechanism of action is not entirely clear [Gerriets et al. 2021]), but they are trained to understand their mechanisms in a general sense and are alert for potential problems. Such general understanding may be all that we should require with respect to AI. Regulators could, in turn, seek transparency on ML decision making; there is an increased emphasis by some on the need for "explainability" tools (e.g., Lundberg et al. 2020; Yap et al. 2021). However, easy fixes in this regard are unlikely. For instance, recent research suggests that this potential technical "fix" can backfire when people put too much trust in "interpretable" models and do not correct their mistakes (Poursabzi-Sangdeh et al. 2021).

Other safety issues relate to how healthcare providers interact with AI tools. There are two sides to this problem: underutilization and overreliance. With respect to underutilization, providers who are uncertain about how AI works may be reluctant to trust and unwilling to adopt AI that could reduce medical errors, which results in an estimated 20,000 Canadian deaths per annum (Baker et al. 2004; Risk Analytica 2017). With respect to overreliance, there are risks if providers rely reflexively on an AI tool without deliberating on, for example, whether it has been trained on unrepresentative data; in such cases, the tool's advice/diagnosis for a particular patient may not be correct (CIFAR 2020). Some experts worry that physicians may even feel compelled to rely on AI against their professional judgment if AI use becomes the standard of care and not using AI results in liability (Froomkin et al. 2019). There are also deeper safety issues inherent in any human–computer interface related to human factors engineering, including how the algorithm presents information and if it does so in a way that minimizes the chance of human error in application.

Existing regulations

Our focus here is, again, on Health Canada, which regulates medical devices, including those with AI, via a licensing regime. This regime is authorized and established under the *Food and Drugs Act* (1985) and *Medical Devices Regulations* (1998). Software that meets the definition of "medical device" under the *Food and Drugs Act* (1985) and *Medical Devices Regulations* (1998) falls under this licensing regime even if it is not part of a hard physical device (Health Canada 2019). Once Health Canada approves a medical device, it is deemed safe for general distribution and sale in Canada and may be adopted into public health plans or sold privately.

There are gaps in this federal regulatory framework (as detailed further in Da Silva et al. 2022). One is that Health Canada's "pre-market" licensing scheme for medical devices only requires that higher-risk devices provide explicit evidence (from studies) of product safety and efficacy. Lower-risk devices can be approved for distribution and sale without providing even

a summary of studies on which safety claims are made. Companies initially self-select the risk category and thus the degree of regulatory oversight. There are no public records as to the extent to which Health Canada reviews this selection process.

A further gap is found in Health Canada's guidance documents interpreting the *Food and Drugs Act's* (1985) and *Medical Devices Regulations'* (1998) application to software. Current guidance largely excludes AI software from licensing requirements if it is "not intended to acquire, process, or analyze a medical image or signal," "intended to display, analyze, or print medical information," "only intended to support" provider decision making, and "not intended to replace … clinical judgment" (Health Canada 2019). While these criteria are not meant to be "determinative" of whether software falls under the *Medical Devices Regulations* (1998), it is again primarily up to innovators themselves to apply these criteria and select their own licensing category. The rationale behind this regulatory carve-out is, presumably, that having a human in the loop can ensure the safety of a given AI tool. But no such exception is made for prescription drugs (i.e., having a physician prescribe the drug does not negate rigorous requirements for testing in the pre-market stage), and it seems unlikely that most healthcare professionals would be able to sufficiently assess any hidden safety issues in AI alone. Furthermore, such an approach would result in underuse of AI by risk-averse physicians, who could not rely on the fact that the product has been cleared as safe by the regulator.

Apart from concerns about gaps in regulatory oversight, ML applications will require ongoing regulatory oversight of a different nature and quality as they evolve over time. Under section 21 of the *Medical Devices Regulations* (1998), all software must be "validated" before entering Canadian markets. This effectively prohibits approval of "adaptive" ML algorithms that learn from new data and change how they operate based on their own real-world performance without human intervention. Under the existing law, this kind of algorithmic change would mean that the application would have to be repeatedly re-licensed. To respond to this, Health Canada is developing a new regulatory pathway for the approval of medical devices with adaptive ML (Health Canada, e-mail communication, June 21, 2021). Yet details on this pathway remain to be seen apart from principles and statements regarding Health Canada plans to work closely with industry partners and "innovate" in the "regulatory sandbox" (Vural et al. 2021).

Recommendations

Health Canada must meet the challenge of how to regulate health-related AI, particularly ML, as safe for use in Canadian markets. To support this task, we offer five brief recommendations.

a. *Transparency:* First, Health Canada will need to be transparent about its plans to regulate and report regularly on performance to garner the trust of patients, healthcare

professionals, the public and even innovators themselves. For example, if AI tools continue to be regulated based on the category of risk that they pose, we suggest that there needs to be transparency in how innovators' choice of licensing category is reviewed. Information should be publicly available on how these choices are audited, how often innovators are required to resubmit to a higher licensing category, and which innovators are required to resubmit. Similarly, as Health Canada develops a new regulatory pathway for ML tools (in the so-called "regulatory sandbox"), it must be transparent regarding what evidence is being relied upon to assess safety prior to market entry and what safety standards will be applied in the post-market stage.

b. *Human in the loop:* Health Canada's current exemption for licensing of health-related AI in some circumstances where there is a healthcare provider "in the loop" should be removed. Providers are not yet trained to assess the risks of health-related AI (Reznick et al. 2020). For the same reason, the mere presence of a human in the loop also should not necessarily trigger lesser scrutiny or a lower risk classification.
c. *Post-market surveillance:* Health Canada should ensure effective post-market surveillance of and transparent reporting on the safety and quality of AI in/as medical devices, particularly for ML. When it comes to post-market surveillance, where risk criteria are set at the pre-market stage, Health Canada should require regular reporting (indexed to risk) within those parameters. Regular random audits by the regulator to ensure that tools are working as intended will be key to successful regulation.
d. *Strong pre-market and post-market oversight:* There may be pressure from some innovators to follow in the footsteps of the US Food and Drug Administration's pilot project, under which "excellent" companies could receive "pre-certification" of new software technologies in exchange for post-market surveillance (US FDA 2020). However, while post-market surveillance is critical for ML tools, this should not be at the expense of appropriate pre-market review. A focus on the track record of the innovator, rather than the product, may be appealing to innovators in terms of having, at least at first blush, a lighter regulatory impact. However, this kind of regulatory oversight could backfire by disadvantaging smaller start-ups without a significant track record but with an excellent product.

    Furthermore, as we discuss below with respect to the problem of algorithmic bias, the initial design of the ML device is critical to how it will perform over time, and appropriate pre-market licensing requirements can ensure that high standards are met in terms of design. Where one unsafe AI tool could produce significant public pushback (Terry 2018), innovators too should support pre-market review – addressing traditional safety concerns and algorithmic bias and cybersecurity issues discussed below – to weed out bad products before they come to market and cause harm to patients.
e. *Regulatory burden assessment:* While it is vital for Health Canada to appropriately regulate the safety of health-related AI, they must also evaluate the regulatory burden that falls upon AI innovators, particularly for start-ups and small companies, so that they are not deterred from bringing excellent products to the market. Constant attention must be

paid to achieving the goal of patient safety while minimizing unnecessary regulatory barriers. After all, unduly keeping otherwise safe health-related AI with huge potential for improving health outcomes off the market will not improve patient safety.

Algorithmic Bias and Safety Worries

The concern(s)

Health-related AI is designed by humans, who have explicit and implicit biases. The algorithms that humans develop could accordingly be inadvertently biased against marginalized groups/patients. AI innovators make many design choices that may increase or decrease algorithmic bias risks, such as decisions about whether or not to pursue and build upon data sets that are more representative but may be more costly or otherwise difficult to obtain. Even representative data sets may prove problematic if, for example, the data were originally collected in ways that encode researchers' bias or are input into an algorithm that itself encodes biases and accordingly has differential impacts (Cirillo et al. 2020).

Beyond the biases/limitations of the programmers/developers themselves, there is the question of whether data sets are themselves representative. There is already discrimination against different groups (e.g., women, racialized populations, sexual and gender minorities) in healthcare (Ayhan et al. 2020; Ben et al. 2017; Dusenbery 2018). Health-related AI might help by providing recommendations that could cause providers to "check" their inherent biases/judgments. Yet AI trained on data that themselves systematically under-represent groups because of discrimination already within the system may not cure, but replicate, existing discrimination problems, resulting in AI recommendations that are subpar for patients from marginalized populations (McCradden et al. 2020; Obermeyer et al. 2019). This could entrench or exacerbate existing inequalities. Consider, for example, the well-known under-inclusion of pregnant women in psychiatric trials (Cirillo et al. 2020; WHO 2009) or the under-treatment of both women and black individuals for cardiovascular disease (Vaccarino et al. 2005). Not all recorded differences in the data are bad biases. Where there are clinical differences scientifically grounded in gender, sex, age or race, this specificity is welcome for it leads to better diagnoses, treatment plans and ultimately health outcomes. But the risks of wrongful bias demand regulatory attention.

Existing regulations

Present medical device regulations in Canada do not specifically address the problem of algorithmic bias. One could look to federal and provincial human rights laws to combat discrimination that results from undesirable algorithmic bias in healthcare settings. However, unlike upstream device regulations, human rights laws will only have impact after discriminatory action occurs, and it will be difficult for individual patients to show causation between upstream decisions made in AI innovation and the harms that they experience (Henderson et al. 2022).

The scope of human rights laws presents further limits. Canada's supreme human rights law, the constitutional *Canadian Charter of Rights and Freedoms* (1982), only applies to governmental action so its protections cannot directly apply to developers or (normally) healthcare professionals. Provincial human rights laws, like the Saskatchewan Human Rights Code (https://saskatchewanhumanrights.ca/your-rights/saskatchewan-human-rights-code/), do apply to private actors and prohibit discrimination on grounds such as race and gender. So, at least theoretically, an AI tool that is not trained on an appropriately representative database may violate the provincial laws. But claimants would face steep evidentiary burdens (Henderson et al. 2022), and we have yet to see judgments requiring collection of "representative" data or, otherwise, clearly protecting against algorithmic bias (Krishnamurthy 2021).

Recommendations

Health Canada alone likely cannot eliminate the threat of algorithmic bias stemming from health-related AI, but we have recommendations on how the federal government can use its powers to at least minimize the algorithmic bias-based risks most likely to have an impact on patient safety.

a. *Assess risk of algorithmic bias in pre-market review:* Health Canada should assess the risk of algorithmic bias as part of safety review. Health Canada should, for instance, require that developers of all health-related AI applications, regardless of risk classification, take all reasonable steps to collect and use training data that are representative of the populations for whom an AI tool will be used. It should also carefully review its present expedited processes for adoption of devices approved by regulators in foreign jurisdictions to ensure that this does not mean that adoption of health-related AI is inapt for Canadian needs. This is particularly so if international regulators try to speed up or lighten regulatory oversight to "support" their own local AI innovators. For Canadian innovators, Health Canada could assess which other mechanisms exist to spur better data inclusiveness, such as whether the Tri-Council (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada and Social Sciences and Humanities Research Council 2018) guidelines on research ethics sufficiently speak to this issue.

   In some cases, using representative data may not be feasible, given the unavailability or prohibitive cost of acquiring it. Race-specific data are not even collected in some jurisdictions, preventing some biases from being assessed. The salience of, for example, race or gender to diagnosis or treatment may vary by clinical context, and risks posed by algorithmic bias may vary accordingly. It is thus essential that in every circumstance, significant effort is put into understanding for whom this algorithm will work by properly characterizing the population on which it was trained and thus providing information as to its limitations for use in clinical settings. While Health Canada may already attend

to these issues (and representatives suggest that this is the case), an explicit and transparent requirement is needed to engender trust on the part of healthcare professionals and patients.

b. *Assess algorithmic bias as part of post-market review:* Beyond the initial point of licensure, Health Canada should also monitor the extent to which AI tools may be perpetuating or exacerbating existing inequalities via life-cycle surveillance of AI in the real world (e.g., examining if any adverse errors disproportionately impact those underrepresented in training data sets).
c. *The federal government should fund data set development:* As we discuss further in the context of privacy issues, Health Canada should also spearhead with other federal departments and agencies the building and liberalization of representative data sets to support the development of health-related AI.

Safety, Privacy and Security

The concern(s)

Many worry that "big data" collection by governments or multinational corporations (e.g., Google, IBM, Facebook) will inevitably produce privacy violations. Users worry that their personal health data might be used to, for example, deny insurance or risk-rate premiums or adversely affect employment opportunities. High-profile personal information breaches, such as an agreement between Google DeepMind and the National Health Service that gave Google 1.5 million patients' records in 2015, which Google sought to bring to the US, underscore these concerns (BBC News 2021). The UK Privacy Commission found that there were insufficient privacy protections in that case, which is now the subject of a class action suit (BBC News 2021). However, large, representative data sets remain necessary to build safe, effective AI and to mitigate risks of algorithmic bias.

Privacy concerns are further complicated by related cybersecurity issues. Even where data holders and users acquire data in a justified manner, they must ensure that they have sufficient security safeguards to protect the data from falling into the hands of persons without legitimate claims to access them. Canada's healthcare systems are, in practice, operated by a plethora of different actors and entities (non-profit hospitals, private clinics, small general practices, large pharmacy chains) with varying capacity to ensure secure systems.

From a safety lens, privacy concerns may make it more difficult for AI innovations to be built from aggregate data that are representative of populations they will serve. Members of Indigenous, Black and other communities that have been marginalized may oppose sharing data because of historical and present-day abuses of their data by governments and private actors. Yet there is a risk of error in diagnosis and treatment if data are not included from such populations. Data "de-identification," in which identifying information is stripped away, reduces privacy risks but is not always possible. There is a risk too, with large data sets involved in most health-related AI projects, of "re-identification" and misuse by less

scrupulous actors (Cohen and Mello 2018). International actors and other jurisdictions are developing best practice standards for protecting data used in AI, and even basic cybersecurity standards can help minimize these risks. But the risks remain. Innovators may also face challenges acquiring representative data from excessive privacy restrictions and/or the multiple layers of restrictions operating at federal, provincial and sub-provincial (e.g., hospital) levels.

Existing regulations

Across Canada, there is a lattice of federal and provincial law relating to privacy of health information. At the heart of federal and provincial privacy protections are requirements that Canadians normally must give consent to the disclosure and sharing of their personal health information. Access to data is, however, difficult across Canada, and given the large data sets needed to create AI tools, it is more likely that AI will seek out extant data sets, even if not sourced in Canada. This is almost always viewed as more feasible than building data sets themselves, avoiding the need to obtain patient consent. Indeed, many AI tools rely on the same data sets from foreign jurisdictions, like US-based MIMIC data sets (Nagaraj et al. 2020).

Canadian privacy laws also do not generally require consent for the use of de-identified data, but triangulation of data and AI could be used to re-identify persons. Recent amendments to Ontario's *Personal Health Information Protection Act* (*PHIPA*) (2004), prohibit re-identification of health information and penalize rule-breakers, and *PHIPA* sets some minimal cybersecurity standards that different types of medical tools must meet. But these kinds of protections are not uniform across Canada.

Recommendations

Health Canada also cannot address issues regarding safety, privacy and security on its own, but there are at least three ways that the federal government, likely including Health Canada, can minimize related concerns.

a. *International security standards*: First, as part of pre-market clearance and ongoing post-market surveillance of ML applications, Health Canada should require compliance with international cybersecurity standards, similar to what regulators have proposed in Germany (FIDMD 2021). For instance, it could require completion of a checklist by the applicant "on data protection, information security and quality," including whether the manufacturer has met international (specifically ISO) and domestic standards with respect to information security and management (FIDMD 2021). These kinds of basic requirements are, of course, not a panacea. Yet they will minimize the chances of at least some kinds of privacy violations, which is a good outcome on its own and should help foster the development of more inclusive data sets.

b. *Punishment for re-identification:* Second, to deliver on the promise of safer healthcare for groups who have been marginalized by society, AI requires large representative data sets, and yet the same historically marginalized groups who may benefit from AI built on more representative data do not necessarily trust our governance systems. To address this problem, the federal government of Canada should assess whether existing federal and provincial laws adequately prevent sharing of identifiable information and reconstituting of previously de-identified information and work with others to ensure an adequate framework exists. Importantly, particularly as the technological capacity of re-identification becomes more feasible with ML, there must be a review of whether penalties for re-identification are sufficient to deter bad actors. Presently, there are many ethical AI innovators who cannot access the data they need because of re-identification concerns. All else being equal, it would seem better to heavily punish bad actors as opposed to inhibiting access to data for all.

   This broader problem is likely one beyond Health Canada's jurisdiction and points to broader systemic concerns – hence the move away from our specific focus and discussion regarding "the federal government," rather than Health Canada, in the preceding paragraphs – but Health Canada should consider whether it can require as part of both pre-market and post-market oversight that patients' and users' privacy is sufficiently protected, given the connections among robust privacy protections, data representativeness and safety. These issues may simply stem from a more general failure to protect data outside Health Canada's jurisdiction, but that should not stop Health Canada from examining if/how they can contribute to better protections. Insofar as Health Canada is not enabled to address these issues, in turn, the federal government may consider providing it with the authority to consider them within safety analyses. Regardless of what one thinks of that proposal, it is clear that the federal government must ensure that someone is attending to these important concerns.

c. *A pan-Canadian strategy on data creation and protection:* Finally, we agree with the Public Health Agency of Canada (2021) Expert Advisory Group on the Pan-Canadian Health Data Strategy's recent statement on the need to enable safe, national pooling of representative data sets. This will require the federal government to use its convening power to lead all provinces and territories into common standards for data sharing and support Indigenous governments and communities that have been marginalized in data sharing (Black Health Equity Working Group 2021).

   Once again, Health Canada may not be able to resolve these issues alone, but the federal government could support and fund Indigenous peoples, Black and other visible minority groups for "better data collection to ensure fair and inclusive AI in medicine" (Pasquale 2020: 39). Health Canada could consider funding partnerships between AI innovators and marginalized groups – provided it is more than a tokenistic exercise – to foster data inclusivity.

Conclusion

AI promises a safer future for Canadian patients, overall, but to realize that future, legal governance of health-related AI must support safe innovations and weed out poor products and unethical innovators. In other words, Health Canada as the regulator must have as its primary goal protection of patient safety and with it a fundamental commitment to transparency, particularly as it experiments with different kinds of regulatory approaches and standards of evidence for AI products that evolve over time.

The regulatory challenge here is significant for health-related AI (particularly ML) is not static: Health Canada as the regulator must be nimble but, at the same time, relentlessly transparent with its successes and failures to garner trust in the safety of health-related AI. Of course, the goal should be to ensure that the regulatory touch consistent with patient safety does not bury innovations in red tape, which will not help Canadian patients. However, this should not be an excuse to aim for minimal or ineffective regulation in a bid to foster Canadian innovation and support Canadian innovators. Furthering innovation consistent with protecting patient safety, Health Canada's primary goal, requires that Health Canada conduct both pre- and post-market review of medical devices with AI that attends to the safety-related risks, including those relating to algorithmic bias, privacy and cybersecurity. Health Canada's resources will need to be bolstered to achieve these goals, so that the resulting regulatory framework ensures that the safety and quality of health-related AI products are maintained over time. This will help foster trust on the part of all stakeholders and the use of beneficial health-related AI that rightly makes it onto Canadian markets.

Funding

This paper is part of the Canadian Institutes of Health Research–funded (grant #155390) project *Machine MD: How Should We Regulate AI in Health Care?* co-led by Colleen M. Flood, Anna Goldenberg, Catherine Régis and Teresa Scassa. Michael Da Silva received post-doctoral funding from the Alex Trebek Forum for Dialogue when working on this piece.

Acknowledgements

The authors are grateful for insights developed at research team meetings and for feedback, both substantive and editorial, from Bryan Thomas, Pascal Thibeault and Catherine Régis.

Declaration

Michael Da Silva was a non-remunerated member of Health Canada's External Reference Group on Adaptive Machine Learning–Enabled Medical Devices. Devin Singh is the co-founder and chief executive officer of Hero AI, a healthcare technology start-up company with a focus on integrating clinical automation solutions into hospital workflows. There are no actual or other potentially perceived conflicts to report.

Correspondence may be directed to: Colleen M. Flood, University of Ottawa, Faculty of Law, Fauteux Hall, 57 Louis-Pasteur Private, Room BRS 331, Ottawa, ON K1N 6N5. Colleen can be reached by phone at 613-562-5800 x 8791 or by e-mail at colleen.flood@uottawa.ca.

References

Ayhan, C.H.B., H. Bilgin, O.T. Uluman, O. Sukut, S. Yilmaz and S. Buzlu. 2020. A Systematic Review of the Discrimination against Sexual and Gender Minority in Health Care Settings. *International Journal of Health Services* 50(1): 44–61. doi:10.1177/0020731419885093.

Baker, G.R., P.G. Norton, V. Flintoft, R. Blais, A. Brown, J. Cox et al. 2004. The Canadian Adverse Events Study: The Incidence of Adverse Events among Hospital Patients in Canada. *CMAJ* 170(11): 1678–86. doi:10.1503/cmaj.1040498.

BBC News. 2021, October 1. DeepMind Faces Legal Action over NHS Data Use. Retrieved January 19, 2022. <https://www.bbc.com/news/technology-58761324>.

Ben, J., D. Cormack, R. Harris and Y. Paradies. 2017. Racism and Health Service Utilisation: A Systematic Review and Meta-Analysis. *PLoS ONE* 12(12): e0189900. doi:10.1371/journal.pone.0189900.

Black Health Equity Working Group. 2021. *Engagement, Governance, Access, and Protection [EGAP]: A Data Governance Framework for Health Data Collected from Black Communities in Ontario*. Retrieved January 19, 2022. <https://blackhealthequity.ca/wp-content/uploads/2021/03/Report_EGAP_framework.pdf>.

Blasimme, A. and E. Vayena. 2020. The Ethics of AI in Biomedical Research, Patient Care, and Public Health. In M.D. Dubber, F. Pasquale and S. Das, eds., *The Oxford Handbook of Ethics of AI* (pp. 703–18). Oxford University Press.

Canadian Charter of Rights and Freedoms, Part 1 of the *Constitution Act*, 1982, Being Schedule B to the Canada Act 1982 (UK), 1982, c 11. Retrieved January 19, 2022. <https://laws-lois.justice.gc.ca/eng/const/page-12.html>.

Canadian Institute for Advanced Research (CIFAR). 2020, July. *Building a Learning Health System for Canadians: Report of the Artificial Intelligence for Health Task Force*. Retrieved January 19, 2022. <https://cifar.ca/wp-content/uploads/2020/11/AI4Health-report-ENG-10-F.pdf>.

Canadian Institute for Advanced Research (CIFAR). 2021, February. *AI & Healthcare: A Fusion of Law & Science: An Introduction to the Issues*. CIFAR. Retrieved January 19, 2022. <https://cifar.ca/wp-content/uploads/2021/03/210218-ai-and-health-care-law-and-science-v8-AODA.pdf>.

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada and Social Sciences and Humanities Research Council. 2018. *Tri-Council Policy Statement Ethical Conduct for Research Involving Humans. TCPS2 2018*. Government of Canada. Retrieved April 25, 2022. <https://ethics.gc.ca/eng/documents/tcps2-2018-en-interactive-final.pdf>.

Canadian Medical Association (CMA). 2020. *The Future of Technology in Health and Health Care: A Primer*. Authors.

Cirillo, D., S. Catuara-Solarz, C. Morey, E. Guney, L. Subirats, S. Mellino et al. 2020. Sex and Gender Differences and Biases in Artificial Intelligence for Biomedicine and Healthcare. *npj Digital Medicine* 3: 81. doi:10.1038/s41746-020-0288-5.

Cohen, I.G. and M. Mello. 2018. HIPAA and Protecting Health Information in the 21st Century. *JAMA* 320(3): 231–32. doi:10.1001/jama.2018.5630.

Da Silva, M., C.M. Flood and M. Herder. 2022. Regulation of Health-Related Artificial Intelligence in Medical Devices: The Canadian Story [In press]. *UBC Law Review* 55(3).

Dusenbery, M. 2018. *Doing Harm: The Truth about How Bad Medicine and Lazy Science Leave Women Dismissed, Misdiagnosed, and Sick*. HarperCollins.

Federal Institute for Drugs and Medical Devices (FIDMD). 2021. *The Fast-Track Process for Digital Health Applications (DiGA) According to Section 139e SGB V: A Guide for Manufacturers, Service Providers and Users*. Retrieved April 25, 2022. <https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.pdf;jsessionid=9A51146423DC56047E367CB8C58323AF.intranet251?__blob=publicationFile>.

Flood, C.M. and C. Régis. 2021. AI & Health Law in Canada. In F. Bariteau-Martin and T. Scassa, eds., *Artificial Intelligence and the Law in Canada* (pp. 203–28). LexisNexis.

Food and Drugs Act, R.S.C., 1985, c. F-27. Retrieved January 19, 2022. <https://laws-lois.justice.gc.ca/eng/acts/f-27/FullText.html>.

Froomkin, A.M., I.R. Kerr and J. Pineau. 2019, February 20. When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning. *Arizona LR* 61: 33–99. doi:10.2139/ssrn.3114347.

Gerke, S., T. Minssen and G. Cohen. 2020. Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare. In B. Bohr and K. Memarzadeh, eds., *Artificial Intelligence in Healthcare* (pp. 295–336). Academic Press.

Gerriets, V., J. Anderson and T.M. Nappe. 2021. *Acetaminophen*. StatPearls Publishing.

Health Canada. 2019, December 18. *Guidance Document: Software as a Medical Device (SaMD): Definition and Classification*. Retrieved January 19, 2022. <https://www.canada.ca/content/dam/hc-sc/documents/services/drugs-health-products/medical-devices/application-information/guidance-documents/software-medical-device-guidance-document/software-medical-device-guidance-document.pdf>.

Henderson, B., C.M. Flood and T. Scassa. 2022. Artificial Intelligence in Canadian Healthcare: Will the Law Protect Us from Algorithmic Bias Resulting in Discrimination?" *Canadian Journal of Law and Technology* 19(2): 475–504.

Krishnamurthy, K. 2021. AI and Human Rights Law. In F. Bariteau-Martin and T. Scassa, eds., *Artificial Intelligence and the Law in Canada* (pp. 229–41). LexisNexis.

Lundberg, S.M., G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair et al. 2020. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nature Machine Intelligence* 2: 56–67. doi:10.1038/s42256-019-0138-9.

McCradden, M.D., S. Joshi, M. Mazwi and J.A Anderson. 2020. Ethical Limitations of Algorithmic Fairness Solutions in Health Care Machine Learning. *The Lancet Digital Health* 2(5): E221–23. doi:10.1016/S2589-7500(20)30065-0.

Medical Devices Regulations (SOR/98-282). Retrieved January 19, 2022. <https://laws-lois.justice.gc.ca/eng/regulations/sor-98-282/FullText.html>.

Murphy, K., E. Di Ruggiero, R. Upshur, D.J. Willison, N. Malhotra, J. Ce Cai et al. 2021. Artificial Intelligence for Good Health: A Scoping Review of the Ethics Literature. *BMC Medical Ethics* 22: 14. doi:10.1186/s12910-021-00577-8.

Nagaraj, K., V. Harish, L.G. McCoy, F. Morgado, I. Stedman, S. Lu et al. 2020. From Clinic to Computer and Back Again: Practical Considerations When Designing and Implementing Machine Learning Solutions for Pediatrics. *Current Treatment Options in Pediatrics* 6(4): 336–49. doi:10.1007/s40746-020-00205-4.

Obermeyer, Z., B. Powers, C. Vogeli and S. Mullainathan. 2019. Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. *Science* 366(6464): 447–53. doi:10.1126/science.aax2342.

Pasquale, F. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

Pasquale, F. 2020. *New Laws of Robotics: Defending Human Expertise in the Age of AI*. Harvard University Press.

Personal Health Information Protection Act, 2004, S.O. 2004, c. 3, Sched. A. Retrieved May 9, 2022. <https://www.ontario.ca/laws/statute/04p03>.

Poursabzi-Sangdeh, F., D.G. Goldstein, J.M. Hofman, J.W. Vaughan and H.M. Wallach. 2021. Manipulating and Measuring Model Interpretability. *arXiv*. doi:10.48550/arXiv.1802.07810.

Price, W.N. 2018. Medical Malpractice and Black-Box Medicine. In I.G. Cohen, H. Lynch, E. Vayena and U. Gasser, eds., *Big Data, Health Law and Bioethics* (pp. 295–306). Cambridge University Press.

Public Health Agency of Canada. 2021, November. *Expert Advisory Group Report 2: Building Canada's Health Data Foundation*. Retrieved January 19, 2022. <https://www.canada.ca/en/public-health/corporate/mandate/about-agency/external-advisory-bodies/list/pan-canadian-health-data-strategy-reports-summaries/expert-advisory-group-report-02-building-canada-health-data-foundation.html>.

Reznick, R.K., K. Harris, T. Horsley and M.S. Hassani with Council Task Force on Artificial Intelligence and Emerging Digital Technologies. 2020, February. *Task Force Report on Artificial Intelligence and Emerging Digital Technologies*. Royal College of Physicians and Surgeons of Canada. Retrieved January 19, 2022. <https://www.royalcollege.ca/rcsite/health-policy/initiatives/ai-task-force-e>.

Risk Analytica. 2017, August. The Case for Investing in Patient Safety in Canada. Canadian Patient Safety Institute. Retrieved January 19, 2022. <https://www.patientsafetyinstitute.ca/en/About/Documents/The%20Case%20for%20Investing%20in%20Patient%20Safety.pdf>.

Terry, N. 2018. Appification, AI, and Healthcare's New Iron Triangle. *Journal of Health Care Law & Policy* 21(2): 117–82. doi:10.2139/ssrn.3020784.

Topol, E. 2019a. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books.

Topol, E.J. 2019b. High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nature Medicine* 25: 44–56. doi:10.1038/s41591-018-0300-7.

U.S. Food and Drug Administration (US FDA). 2020, September. *Developing the Software Precertification Program: Summary of Learnings and Ongoing Activities*. Retrieved January 19, 2022. <https://www.fda.gov/media/142107/download>.

Vaccarino, V., S.S. Rathore, N.K. Wenger, P.D. Frederick, J.L. Abramson, H.V. Barron et al. 2005. Sex and Racial Differences in the Management of Acute Myocardial Infarction, 1994 through 2002. *The New England Journal of Medicine* 353: 671–82. doi:10.1056/NEJMsa032214.

Vural, I.E., M. Herder and J.E. Graham. 2021. From Sandbox to Pandemic: Agile Reform of Canadian Drug Regulation. *Health Policy* 125(9): 1115–120. doi:10.1016/j.healthpol.2021.04.018.

Wong, A., E. Otles, J.P. Donnelly, A. Krumm, J. McCullough, O. DeTroyer-Cooley et al. 2021. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Internal Medicine* 181(8): 1065–70. doi:10.1001/jamainternmed.2021.2626.

World Health Organization (WHO). 2009. *Mental Health Aspects of Women's Reproductive Health: A Global Review of the Literature*. Retrieved January 19, 2022. <http://apps.who.int/iris/bitstream/handle/10665/43846/9789241563567_eng.pdf;jsessionid=F69A34D660ED21863D1D62F2BCBB14D5?sequence=1>.

World Health Organization (WHO). 2021, June 28. *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. Retrieved January 19, 2022. <https://www.who.int/publications/i/item/9789240029200>.

Yap, M., R.L. Johnston, H. Foley, S. MacDonald, O. Kondrashova, K.A. Tran et al. 2021. Verifying Explainability of a Deep Learning Issue Classifier Trained on RNA-Seq Data. *Scientific Reports* 11: 2641. doi:10.1038/s41598-021-81773-9.