Health Care AI and the Governance of De-identified and Anonymized Data: Balancing AI's Big Data Needs with Strong Privacy Protections

*Bradley Henderson**

Public and private sector actors are developing increasingly complex artificial intelligence (AI) tools in health care, for a wide range of health care-related purposes. To develop safe, effective, and equitable health care AI tools, developers require access to large quantities of high-quality health care data, often in the form of de-identified or anonymized data. However, increased access to robust health care data exacerbates informational privacy risks, especially in the health care sector because of the sensitivity of health information. Most prominently, health care AI tools may increase the risks of an individual or group being re-identified

Les acteurs des secteurs public et privé développent des outils d'intelligence artificielle (IA) en santé de plus en plus complexes pour une grande variété de buts en lien avec les soins de santé. Pour développer des outils d'IA sécuritaires, efficaces et équitables, les développeurs ont besoin d'accéder à de grandes quantités de données de santé de haute qualité, souvent sous forme de données dépersonnalisées ou anonymisées. Cependant, un accès accru à de robustes données de santé exacerbe les risques liés à la confidentialité informationnelle, surtout dans le secteur des soins de santé en raison de la sensibilité de l'information en santé.

---

* Bradley Henderson is a recent JD graduate of the University of Ottawa Faculty of Law. When this paper was accepted for publication, he was a salaried employee of Health Canada; however, Health Canada played no role whatsoever in the conception, design, analysis, or writing of this article. This paper reflects the views of the author, and not those of Health Canada. The original version of this paper was written in fall 2023 as a directed research project under the supervision of Professor Teresa Scassa.

Citation: Bradley Henderson, "Health Care AI and the Governance of De-identified and Anonymized Data: Balancing AI's Big Data Needs with Strong Privacy Protections" (2025) 17:1 McGill JL & Health 40.

being re-identified from de-identified or anonymized health care data, significantly enhance cybersecurity threats, or enable unwanted health-related inferences about individuals and groups, all of which could result in serious psychological harm. Across Canada, constitutional constraints have led to a patchwork of federal and provincial data protection laws, which have also failed to keep up with the pace of AI innovation and its resulting informational privacy risks. Existing laws and law reform efforts incentivize developers' use of anonymized and, in some cases, de-identified data for AI innovation, yet these laws and law reform efforts provide for inadequate governance of de-identified and anonymized health care data, especially because of AI-related re-identification risks. This paper recommends several policy approaches moving forward that would mitigate health care AI-related privacy risks, primarily concerning de-identified and anonymized health care data, in a context-dependent manner. These policy approaches would better protect privacy by promoting the appropriate use of individuals' and groups' health care information, while still supporting health care AI innovation.

Les outils d'IA en santé peuvent notamment augmenter le risque que des individus ou des groupes soient réidentifiés à partir de données dépersonnalisées ou anonymisées, exacerber considérablement les cybermenaces, ou encore permettre des inférences non désirées sur la santé d'individus ou de groupes, ce qui peut entraîner d'importants préjudices psychologiques. À travers le Canada, les contraintes constitutionnelles ont mené à un ensemble disparate de lois fédérales et provinciales sur la protection des données, lesquelles n'ont pu suivre le rythme d'innovation de l'IA et de ses risques émergents pour la confidentialité. Les lois existantes et les efforts de réforme législative encouragent l'utilisation des données anonymisées et, dans certains cas, des données dépersonnalisées par les développeurs d'IA, mais ces lois et réformes législatives offrent une gouvernance inadéquate de telles données, surtout à cause des risques de réidentification liés à l'IA. Cet article recommande plusieurs approches politiques à adopter pour atténuer les risques liés à la protection de la vie privée dans le domaine de l'IA en santé, principalement en ce qui concerne les données de santé dépersonnalisées et anonymisées, et ce, d'une manière contextuelle. Ces approches politiques protègeraient mieux la confidentialité en promouvant une utilisation appropriée des renseignements de santé des individus et des groupes, tout en soutenant l'innovation de l'IA dans le secteur de la santé.

Acknowledgement: *The author thanks Professor Teresa Scassa for her guidance, comments, and feedback on various versions of this paper, Dean Colleen M. Flood for her guidance on early ideas for this paper generally, and the anonymous peer reviewers for their helpful feedback.*

Glossary

"Personal information" generally means identifying information about an individual, i.e., where a "serious possibility [exists] that an individual could be identified through the use of [the] information, alone or in combination with other available information."[1]

"De-identified information" generally means personal information from which an individual cannot be directly identified, but from which they could still be indirectly identified (i.e., a risk of re-identification remains).[2]

"Anonymized data" generally means information (in relation to personal information) from which, at all times, reasonably foreseeable in the circumstances, an individual cannot (irreversibly) be directly or indirectly identified.[3]

"Health information" generally means information that relates to an individual's health or health care.[4]

"Data" generally means, for the purpose of this paper, information—and vice versa.

---

1 See *Gordon v Canada (Minister of Health)*, 2008 FC 258 at para 34.

2 See e.g. the definition of "de-identify" in Bill C-27, *An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts*, 1st Sess, 44th Parl, 2022, cl 2, s 2(1) (second reading 24 April 2023) [*Bill C-27*]. See also *Act Respecting the Protection of Personal Information in the Private Sector*, CQLR c P-39.1, s 12 [*Quebec Private Sector Personal Information Act*]. For example, a data set may exclude an individual's name (direct identifier), yet other information in the data set about the individual may allow other persons to identify them, such as the individual's age, sex, gender, occupation, postal code, and height (indirect identifiers).

3 See e.g. *Quebec Private Sector Personal Information Act*, *supra* note 2, s 23. Also, to note, definitions of personal information, de-identified information, and anonymized data sometimes differ between jurisdictions, depending on the applicable legislation and case law. For example, some legislation defines "de-identified information" to mean what this paper and the broader privacy community now considers "anonymized data", see e.g. the definition of "de-identify" in Ontario's *Personal Health Information Protection Act, 2004,* SO 2004, c 3, Schedule A, s 2 [*PHIPA*].

4 See e.g. *PHIPA*, *supra* note 3, s 4.

I. Introduction

The digitization of health care data and advances in computing power have enabled public and private sector actors to develop increasingly complex artificial intelligence (AI) tools in health care.[5] These tools may, for example, generate novel insights into complex diseases, predict the likelihood of certain conditions, recommend treatment options, determine resource allocation, increase health care efficiency, or mitigate extant problems, such as misdiagnoses and discrimination in health care.[6] However, safe, effective, and equitable health care AI tools depend on large quantities of high-quality health care data and developers' access to that data—often in the form of de-identified or anonymized data.

In turn, vast amounts of health care data and increased access to that data, including by commercial AI developers, have exacerbated informational privacy risks, especially for individuals and groups represented in such data. These privacy risks are particularly salient in the health care sector because of the sensitivity of health care information, such as a person's genetic predispositions, HIV status, or other physical or mental states associated with significant social stigma and historic or ongoing discrimination. Most prominently, health care AI tools may increase risks of being re-identified from de-identified or anonymized health care data, significantly enhance cybersecurity threats, or enable unwanted health-related inferences about individuals and groups. Breaches of individuals' and groups' informational privacy in the health care context can result in serious psychological harm.

Across Canada, constitutional constraints have led to a patchwork of federal and provincial laws that protect individuals' control over their personal information, including their personal health information. However, many of these laws have failed to keep up with the pace of health care AI innovation and its resulting informational privacy risks. Most personal information and personal health information laws do not apply to anonymized data, and ongoing law reform would reduce oversight for de-identified data in certain circumstances. The rationale behind these policy choices is that

---

5 See Jennifer S Winter & Elizabeth Davidson, "Governance of Artificial Intelligence and Personal Health Information" (2019) 21:3 Digital Policy, Regulation and Governance 280 at 281.

6 David W Bates et al, "The Potential of Artificial Intelligence to Improve Patient Safety: A Scoping Review" (2021) 4:1 NPJ Digital Med 1 at 1.

less stringent oversight over these types of data would incentivize entities to anonymize and de-identify data—thereby protecting privacy—and support the data's use and disclosure, including for AI-related innovation. While individuals and groups generally support the use or disclosure of their health information for health-research purposes, including in the form of anonymized and de-identified data, their support is often contingent on strong privacy protections and penalties for non-compliance.[7]

This paper is divided into eight short parts, including this introduction. Part II explains health care AI's big data needs and Part III describes in greater detail the predominant informational privacy risks exacerbated by health care AI. Part IV then elaborates on the patchwork of data protection legislation across Canada and Part V discusses the constitutional constraints that have limited and continue to limit federal AI and data protection legislative reform. In Part VI, this paper argues that ongoing federal law reform efforts and existing laws provide for inadequate governance of de-identified and anonymized health care data, especially because of AI-related re-identification risks. While remaining cognizant of the constitutional constraints described in part V, Part VII of this paper then recommends several policy approaches moving forward that would mitigate health care AI-related privacy risks—primarily concerning de-identified and anonymized health care data, before concluding in Part VIII. These contextual policy approaches range from light-touch notification requirements to stronger quasi-approval processes depending on the sensitivity of the health care information, the nature of the actors involved, and the re-identification risks. Proper oversight, in a context dependent manner, would promote the appropriate use of individuals' and groups' health care information, while still supporting health care AI innovation and improving Canadian health care.

---

7 P Alison Paprica, Magda Nunes de Melo & Michael J Schull, "Social Licence and the General Public's Attitudes Toward Research Based on Linked Administrative Health Data: A Qualitative Study" (2019) 7:1 CMAJ Open E40 at E44.

II. Developing Safe, Effective, and Equitable Health Care AI Requires Robust and Accessible Data

AI tools, in particular machine learning (ML) tools, generally require vast amounts of training data to learn from that data how best to achieve certain objectives or tasks, with varying degrees of human guidance.[8] In some cases, these tools may also learn from new input data in real-time to continuously learn and self-optimize.[9] Beyond training data and new input data, AI developers need access to additional health care data to test and validate those AI tools before deploying them in, for example, clinical settings. Yet AI's data needs do not stop here. Developers need even more data throughout those tools' lifecycles to retrain, retest, and revalidate them, so that they remain accurate, or even improve, over time.[10] Of particular relevance to health care, the advances in computing power and the digitization of health care-related records, mentioned above, have enabled a subtype of ML known as deep learning.[11] Deep learning allows machine-based tools to transform and optimize themselves through mathematical and statistical analyses based on large amounts of data, including raw (unstructured) data and different types of data.[12] While deep learning tools may become complex and opaque, their many "layers" of analytical nodes produce outputs that assist us in identifying hidden patterns within data, to generate novel insights and improve health care and health outcomes.[13] It follows that better health care and health outcomes in Canada require large amounts of health care data.

Individual and group-related health care data comes in many forms and from many different sources. Health care-related personal information

---

8 Andrew L Beam & Isaac S Kohane, "Big Data and Machine Learning in Health Care" (2018) 319:13 JAMA 1317.

9 See Timo Minssen et al, "Regulatory Responses to Medical Machine Learning" (2020) 7:1 JL Biosci 1 at 2.

10 See Jianxing He et al, "The Practical Implementation of Artificial Intelligence Technologies in Medicine" (2019) 25:1 Nat Med 30 at 31.

11 See Andre Esteva et al, "A Guide to Deep Learning in Healthcare" (2019) 25:1 Nat Med 24 at 24.

12 C David Naylor, "On the Prospects for a (Deep) Learning Health Care System" (2018) 320:11 JAMA 1099 at 1099–100.

13 *Ibid*. See also Esteva et al, *supra* note 11 at 24–25.

may include patients' medical histories, health care utilization, prescription drug usage, medical device outputs, insurance claims, and other clinical information,[14] such as diagnostic images, lab results, and genetic tests.[15] Relatedly, sources of health care-related personal information include governments, pharmacies, commercial organizations, testing laboratories, hospitals, physician practices, nursing homes, third-party payers (such as insurers), clinical trial development organizations, information-technology vendors, and other similar entities.[16] In addition, while beyond the scope of this paper, various other sources of health-related data include social media, genealogical test results, and wearable technology companies, which gather vast amounts of consumer-related personal health information and consumer-related personal information that may become personal health information.[17] For example, an individual's internet search history and geolocation data may seem innocuous, but an entity may use this data, alone or in combination with other data, to lawfully or unlawfully infer (collect) new health information about that individual.[18]

Health care AI developers seek out large amounts of health care data, often at the patient-level, about thousands of individuals, and for as many distinct data points about those individuals as possible. These data points often include patient features such as an individual's date of birth, ethnicity, postal code, sex, gender, income, occupation, diagnoses, conditions, medications, and diagnostic images. Vast feature-rich data is especially important for health care AI, because health care is highly contextual; we often do not know which feature-variables (inputs) are causing or contributing to

---

14 See W Nicholson Price II, "Problematic Interactions Between AI and Health Privacy" (2021) 2021:4 Utah L Rev 925 at 930. See also Winter & Davidson, *supra* note 5 at 282.

15 See He et al, *supra* note 10 at 30.

16 See Winter & Davidson, *supra* note 5 at 283.

17 See Minssen et al, *supra* note 9 at 15.

18 See Winter & Davidson, *supra* note 5 at 282. For example, a company might infer that an individual has suffered a myocardial infarction where that individual browsed "chest pain" on the company's mobile web-browsing application, subsequently presented to a hospital, and then visited a cardiac rehabilitation centre three times weekly for several months—as recorded in the mobile web-browsing application's geolocation data log.

specific health care-related outcomes (outputs).[19] Because complex AI tools seek to identify hidden patterns within data or hidden factors that contribute to a specific health care-related output, developers may not know in advance of developing an AI tool which types of data, or which features within data, will be useful for training, testing, and validating that tool; and AI tools' data needs may change over time.[20] It follows that AI developers generally support a maxim of "the more data, the better" (or data maximization) in the health care sector for AI innovation.[21]

Problematically, most data custodians, such as insurers and health care providers, operate in silos, and the data within those organizations often contain gaps. Data custodians within jurisdictions commonly operate on different data infrastructure, such as poorly interoperable electronic health record systems, and data stewards within those organizations commonly manage data in accordance with different standards. These differences are especially pronounced between jurisdictions, such as between provinces or countries, primarily due to different personal information or health sector legislation.[22] Further, data custodians generally collect health care data primarily for the purposes of patient care or improving that care. Consequently, health care data may be limited in terms of patient features (which data custodians may have considered unnecessary to collect at the time of care). For example, during a specific physician-patient interaction, a physician would likely only record patient features relevant to that interaction, insofar as those features assist the physician in delivering safe and effective care to that patient, during or after the interaction.

---

19 See Bradley Henderson, Colleen M Flood & Teresa Scassa, "Artificial Intelligence in Canadian Healthcare: Will the Law Protect Us from Algorithmic Bias Resulting in Discrimination?" (2022) 19:2 CJLT 475 at 484–85.

20 See Blake Murdoch, Allison Jandura & Timothy Caulfield, "Privacy Considerations in the Canadian Regulation of Commercially-Operated Healthcare Artificial Intelligence" (2022) 5:4 Can J Bioethics 44 at 45 [Murdoch, Jandura & Caulfield, "Privacy Considerations"]. See also Roger Allan Ford & W Nicholson Price II, "Privacy and Accountability in Black-Box Medicine" (2016) 23:1 Mich Telecomm & Tech L Rev 1 at 32; Charlotte A Tschider, "The Healthcare Privacy-Artificial Intelligence Impasse" (2020) 36:4 Santa Clara High Tech LJ 439 at 440.

21 See Tschider, *supra* note 20 at 441.

22 See Part IV, below, for a broader discussion on this point.

Other sources of data pose similar challenges. Biomedical research data, for example, often captures health care data (or patient features) relevant to a specific clinical study, such as the anticipated endpoints of a clinical trial, adverse events, date of birth, sex, comorbidities, and concomitant medications. In fact, most developers create health care AI tools using data not originally intended for AI research and development.[23] Where developers rely on such data as a secondary use to train an AI tool, incomplete and patchy data risks creating AI tools that may underfit future input data and thereby produce less accurate outputs—especially in respect of patient-feature gaps.[24] Overall, these interoperability issues and data gaps pose barriers that impede the development of large, comprehensive data sets.[25] However, large, comprehensive data sets alone are not sufficient to develop safe and effective health care AI.

Safe and effective health care AI tools also require large, comprehensive data sets that are heterogeneous,[26] including for historically marginalized and underrepresented groups, hereafter referred to as "robust data".[27] AI tools trained on ethnically and racially homogeneous health care data, for example, risk producing inaccurate, discriminatory, ineffective, and unsafe outputs for patients underrepresented in that data.[28] These AI tools are more likely to overfit homogeneous data, increasing their accuracy for well-represented patient subpopulations, while decreasing their accuracy for underrepresented patient subpopulations,[29] potentially leading to biased

---

23 See Melissa D McCradden, Elizabeth A Stephenson & James A Anderson, "Clinical Research Underlies Ethical Integration of Healthcare Artificial Intelligence" (2020) 26:9 Nat Med 1325 at 1325.

24 See Henderson, Flood & Scassa, *supra* note 19 at 489.

25 See Winter & Davidson, *supra* note 5 at 282–83.

26 See Henderson, Flood & Scassa, *supra* note 19 at 482.

27 See Ford & Price, *supra* note 20 at 9. See also W Nicholson Price II & I Glenn Cohen, "Privacy in the Age of Medical Big Data" (2019) 25:1 Nat Med 37 at 37.

28 See Michael Da Silva et al, "Regulating the Safety of Health-Related Artificial Intelligence" (2022) 17:4 Healthc Policy 63 at 71 [Da Silva et al, "Regulating Safety"].

29 See Henderson, Flood & Scassa, *supra* note 19 at 490.

outputs in favour of one subpopulation to the detriment of another.[30] For example, a training data set containing thousands of patients, with dozens of patient-level features, from Boston, Massachusetts is likely to generate an AI tool that produces less accurate outputs for, as an example, Indigenous peoples living in rural, remote, and northern Canadian communities.[31] Unacceptably, Indigenous peoples continue to experience structural racism in Canadian health care, leading to significant adverse consequences such as differential access to primary and tertiary care[32] and different lifetime risks of developing certain conditions.[33] Biased AI tools are likely to exacerbate, rather than mitigate, these unacceptable structural inequities.[34] In general, health care data used to train AI tools must include diverse patient subpopulations—including diverse ethnic and racial subpopulations and different geographical subpopulations within Canada—with individual-level feature-rich data for those subpopulations in a range of clinical contexts. Only then can we create health care AI tools for diverse populations that will equitably benefit those populations across Canada.[35]

We also need robust, aggregated, standardized data from several sources and regions for strong AI innovation. Aggregated data from several sources generally improves the quality of data by filling in gaps, including

---

30 Algorithmic bias generally refers to skewed outputs without legal, clinical, or moral justification. *Ibid* at 476.

31 For example, MIMIC-III is a large publicly available healthcare dataset, intended for the development of healthcare-related AI, from a tertiary care hospital in Boston, see Alistair EW Johnson et al, "MIMIC-III, A Freely Accessible Critical Care Database" (2016) 3:1 Sci Data 1 at 2.

32 See Brenda L Gunn, "Ignored to Death: Systemic Racism in the Canadian Healthcare System" (2016) at 4, online (pdf): <ohchr.org> [perma.cc/H4FW-ZW5Q]. See also Patrick McLane et al, "Impacts of Racism on First Nations Patients' Emergency Care: Results of a Thematic Analysis of Healthcare Provider Interviews in Alberta, Canada" (2022) 22:804 BMC Health Serv Res 1 at 7.

33 See e.g. Tanvir Chowdhury Turin et al, "Lifetime Risk of Diabetes Among First Nations and Non-First Nations People" (2016) 188:16 CMAJ 1147 at 1147.

34 See Marieke Bak et al, "You Can't Have AI Both Ways: Balancing Health Data Privacy and Access Fairly" (2022) 13:1 Front Genet 1 at 3.

35 See Henderson, Flood & Scassa, *supra* note 19 at 479.

for individual patients and patient-specific features.[36] Further, aggregated data from different regions generally increases the resulting datasets' ethnic and racial diversity. For example, aggregating data from wealthier urban areas, lower-income rural areas, geographic regions with higher proportions of Black individuals, and geographic regions with higher proportions of other historically marginalized communities would yield more robust datasets.

However, in doing so, we must also aggregate data in ways that respect the self-determination and data sovereignty of marginalized communities, such as Black individuals and Indigenous peoples.[37] Data self-determination and data sovereignty are interrelated concepts that generally refer to a right for historically marginalized groups to own, control, access, and steward their data, in large part due to historic and ongoing data-gathering and data-sharing injustices perpetrated by both governments and private actors against those groups.[38] Although, Indigenous data sovereignty is grounded in Indigenous groups' and communities' *sui generis* legal rights to self-determination that are not available to other minority groups. Notably, data governance frameworks are emerging that seek to ensure that others' collection or use of Black communities' data occurs in an ethical and appropriate manner, to advance health equity, return tangible benefits to those communities, and dismantle structural racism.[39] For Indigenous communities specifically, data self-determination and data sovereignty also apply to data relating to those communities' lands and cultures. These concepts clarify that those communities are partners in data-related research, rather than subjects, and that any information management and data collection must align with those communities' culture and practices.[40]

In response to data gaps and interoperability issues, various actors have begun developing standardized, multiregional and/or multijurisdictional

---

36 See Price & Cohen, *supra* note 27 at 42.

37 See e.g. Black Health Equity Working Group, *Engagement, Governance, Access, and Protection (EGAP): A Data Governance Framework for Health Data Collected from Black Communities* (Toronto: Black Health Equity Working Group, 2021) at 26. See also Keshav Makunda, "Indigenous Data Sovereignty" (last modified 2 August 2024), online: <lib.sfu.ca> [perma.cc/5LDW-QP8C].

38 See Da Silva et al, "Regulating Safety", *supra* note 28 at 71.

39 See Black Health Equity Working Group, *supra* note 37 at 16, 25–26.

40 See Makunda, *supra* note 37.

data portals. For example, the Public Health Agency of Canada's Pan-Canadian Health Data Strategy Expert Advisory Group is seeking to improve the national creation, collection, storage, and use of health care-related data by, among other things, modernizing data interoperability, updating approaches concerning privacy and access to digital data, and clarifying governance and accountability mechanisms for government partners.[41] In Ontario, the provincial government intends to create a new Data Authority to build "modern data infrastructure to support economic growth at scale."[42] During the COVID-19 pandemic, the Government of Ontario also created the Ontario Health Data Platform to combine data and facilitate data sharing from public and private sector sources.[43] Further, the Health Data and Research Network Canada, a Canadian non-profit organization, has created a portal for researchers seeking multiregional health and health-related administrative data in Canada, which also streamlines data requests and access across Canada.[44]

AI developers must be able to access robust, aggregated data for health care AI to reach its full potential.[45] Yet, existing secondary data use frameworks for personal health information, including those for health care AI research and development, as discussed in greater detail below, generally focus on minimizing the scope of health data shared between data custodians (or stewards) and researchers. The entire process of collecting, aggregating, and sharing healthcare data—from a patient's first interaction with a healthcare system to a not-for-profit company aggregating and providing access to multijurisdictional data, and every step in between and thereafter—raises many significant privacy concerns for both individuals and groups.

---

41 See Government of Canada, "Overview of the Former Expert Advisory Group for the Pan-Canadian Health Data Strategy" (last modified 20 March 2024), online: <canada.ca> [perma.cc/5PWG-CUU4].

42 See Ontario, ''Building a Digital Ontario" (30 April 2021), online: <ontario.ca> [perma.cc/9VEF-QC9M]. See also Henderson, Flood & Scassa, *supra* note 19 at 502.

43 See Ontario Health Data Platform, ''Ontario Health Data Platform COVID-19" (last modified 20 April 2023), online: <ohdp.ca> [perma.cc/RG5K-78AS]. See also Henderson, Flood & Scassa, *supra* note 19 at 490.

44 See Health Data Research Network Canada, "Data Access Support Hub", (last visited 23 November 2023), online: <hdrn.ca> [perma.cc/8YSQ-WUWC].

45 See Bak et al, *supra* note 34 at 4.

III. Health care AI Exacerbates Informational Privacy Risks

Privacy is linked to an individual's autonomy, personal identity, well-being, and dignity.[46] Health-related privacy generally concerns informational privacy, that is, "the use and control over one's personal health information."[47] Various legal instruments aim to protect privacy rights or, more aptly, to protect individuals and identifiable groups from losing control over their personal health information. Each of these legal instruments has slightly different yet related scopes and purposes, such as fundamental international human rights agreements to which Canada is a party,[48] the *Canadian Charter of Rights and Freedoms*,[49] and various federal and provincial laws with quasi-constitutional status.[50] Further, the related concept of data protection generally captures both individual privacy (freedom from unwanted interference or intrusion) and the protection of personal information (informational privacy).[51] Privacy breaches can cause serious harm to individuals and groups in terms of their autonomy and decisional privacy.[52]

Privacy risks may be consequential or deontological.[53] Consequential risks include privacy breaches that result in tangible or intangible harms,

---

46 See Murdoch, Jandura & Caulfield, "Privacy Considerations", *supra* note 20 at 45.

47 *Ibid* at 45; See also *R v Tessling*, 2004 SCC 67 at para 23.

48 See *Universal Declaration of Human Rights,* UNGA, 3rd Sess, UN Doc A/RES/3/217A (1948) GA Res 217A (III), art 12.

49 *Canadian Charter of Rights and* Freedoms, Part I of the *Constitution Act,* 1982, being Schedule B to the *Canada Act* 1982 (UK), 1982, c 11, s 8.

50 See e.g. *Personal Information Protection and Electronic Documents Act*, SC 200, c 5 [*PIPEDA*]. See also *PHIPA*, *supra* note 3; *Alberta v United Food Commercial Workers, Local 401*, 2013 SCC 63 at para 19; *Lavigne v Canada (Office of the Commissioner of Official Languages)*, 2002 SCC 53 at para 24.

51 See Tschider, *supra* note 20 at 440–41.

52 See Ford & Price, *supra* note 20 at 21. To note, decisional privacy generally refers to one's right against unwanted interference in their decisions, as a component of their autonomy, see generally Beate Roessler, *The Value of Privacy* (Cambridge: Polity Press, 2005).

53 See Teresa Scassa, "The Unduly Narrow Scope for 'Harm' and 'Biased Output' Under the AIDA" (22 August 2022), online (blog): <teresascassa.ca> [perma.cc/6NFB-7TZ2] [Scassa, "Harm and Biased Output"].

such as higher insurance premiums, discrimination, exploitation, deception, or emotional and psychological harm arising from the inappropriate collection, use, or disclosure of one's sensitive personal health information.[54] This harm may also be severe. More than a century ago, Samuel Warren and Louis Brandeis recognized that privacy-related harms could manifest as psychological suffering or distress greater than the harm associated with mere bodily injury or injury to one's feelings.[55] For example, the unlawful and unwanted disclosure of a person's recent terminal illness diagnosis could severely affect that person's mental health. Overall, these consequential outcomes can impact an individual's choices and opportunities.[56] For example, patients may avoid important care for fear of being denied future employment. On the other hand, deontological concerns relate more to one's loss of control over their personal health information and the resulting moral or ethical rights-based infringement, even where no tangible or intangible harm flows from such infringement.[57] In addition, breaches of an identifiable group's privacy may, problematically, manifest in severe collective harms while also constituting quantifiable, *de minimus,* individual-level harms, creating challenges for successful individual and class action lawsuits.[58]

Health care AI specifically exacerbates many privacy and data protection-related risks. Robust data "increases the number of individuals that could be affected, the severity of effect, and the difficulty for aggrieved individuals to engage in preventative measures" concerning privacy risks and

---

54 See Price & Cohen, *supra* note 27 at 38.

55 Paul Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization" (2010) 57 UCLA L Rev 1701 at 1732–33. See also Samuel D Warren & Louis D Brandeis, "The Right to Privacy" (1890) 4:5 Harv L Rev 193 at 196–97, 216.

56 Teresa Scassa, "AI and Data Protection Law" in Florian Martin-Bariteau & Teresa Scassa, eds, *Artificial Intelligence and the Law in Canada* (Toronto: LexisNexis Canada, 2021) 123 [Scassa, "AI and Data Protection Law"].

57 See Price & Cohen, *supra* note 27 at 38.

58 See Scassa, "Harm and Biased Output", *supra* note 53. For example, a privacy breach could reveal sensitive public health information about a community, which may lead to significant stigma, discrimination, or other adverse consequences in relation to that community, yet, on an individual level, the resulting harm may be considered trivial for the purpose of being actionable at law.

resulting harms.[59] The AI development process also means that health care data may change control, possibly several times, across multiple institutions and even internationally,[60] rendering such data more susceptible to being intercepted and misused, including by malicious actors.[61] While AI-related privacy risks are numerous and expanding, this paper focuses on those risks most prominent in the health care context, which are predominantly re-identification risks from de-identified or anonymized data and, to a lesser extent and as they relate to re-identification risks, inferences about an individual's or group's health, and cybersecurity breaches.

A. Re-identifying Individuals Using De-identified or Anonymized Data

Aggregating large, de-identified or anonymized data sets from multiple sources increases re-identification risks. As noted, aggregated health care-related data often include data from multiple sources, such as insurers, hospitals, electronic health record vendors, and other similar entities. When a person merges two or more data sets, indirect identifiers generally increase for any individual represented in more than one of those data sets.[62] Indirect identifiers will overlap with each other, often yielding a more comprehensive indirect profile of an individual. More individual-level features generate better health care AI; however, more feature rich-data also increases the risk that de-identified or anonymized information will, when aggregated, become more easily identifiable personal health information.[63] This risk is particularly salient where one of the data sets contains direct identifiers about individuals,[64] permitting near-immediate re-identification of an individual,

---

59 See Price & Cohen, *supra* note 27 at 38.

60 See He et al, *supra* note 10 at 31.

61 See Ford & Price, *supra* note 20 at 25.

62 Boris Lubarsky, "Re-Identification of 'Anonymized' Data" (2017) 1:1 Geo L Tech Rev 202 at 203, 211.

63 See William Parker et al, "Canadian Association of Radiologists White Paper on De-Identification of Medical Imaging: Part I, General Principles" (2021) 72:1 Can Assoc Radiologists J 13 at 20.

64 Direct identifiers include things like phone numbers, names, registration numbers, and patient health insurance numbers. Indirect identifiers include things like sex, date of birth, postal code, medical images, and city of residence, and sensitive variables include things like conditions or treatments, see *ibid* at 17.

such as where a person combines health care data with outside information, including non-health care information.[65] In either case, this process of re-identification is commonly referred to as data triangulation.[66] Professor Latanya Sweeney famously illustrated these risks in the mid-1990s.[67] At the time, the Massachusetts Group Insurance Commission released anonymized hospital-level data about its state employees for research purposes. Governor Bill Weld assured the public that the data was anonymized and scrubbed of all identifiable information, such as names and registration numbers. To prove otherwise, Professor Sweeney purchased Cambridge voter registration rolls and linked the two data sets. Once combined, she re-identified Governor Weld and then sent his re-identified hospital records to his office.[68] While Professor Sweeney intended to demonstrate re-identification risks and resulting harms to prompt legislative reform, other actors may not have such virtuous objectives. Personal health information is often more valuable than de-identified and anonymized data when developing health care AI and for other purposes such as commercial marketing or other activities targeted towards individuals, which may incentivize less scrupulous actors or corporations to re-identify individuals.[69]

Re-identification-related privacy risks raise complex questions about anonymized data. For over a decade, experts in the field have been flagging that the concept of "anonymization" is largely illusory.[70] Depending on the

---

65 See Ohm, *supra* note 55 at 1707–08, 1726. See also Winter & Davidson, *supra* note 5 at 284. Unlawful data scraping, i.e., the unlawful harvesting of vast amounts of publicly available personal information and personal health information, including biometric data, further increases re-identification risks. For example, Clearview AI scraped billions of images from the internet without the respective individuals' consent to develop a facial recognition tool, see Office of the Privacy Commissioner of Canada, *Clearview AI ordered to comply with recommendations to stop collecting, sharing images*, (Ottawa: OPC, 2021) online: <priv.gc.ca> [perma.cc/84EZ-52MZ].

66 See Price & Cohen, *supra* note 27 at 40.

67 See Latanya Sweeney, "Simple Demographics Often Identify People Uniquely" (2000) Carnegie Mellon University, Working Paper No 3 at 8, online (pdf): <dataprivacylab.org> [perma.cc/2VJ9-X5LA]. See also Ohm, *supra* note 57 at 1719–20.

68 *Ibid* at 2.

69 See Winter & Davidson, *supra* note 5 at 285.

70 See Ohm, *supra* note 55 at 1704.

study, a person can identify between 63% and 87% of Americans from three indirect identifiers: zip code, birthdate, and sex.[71] Professor Paul Ohm has argued that true anonymization does not exist and that, instead, we can only create strongly de-identified data.[72] He also refers to large, aggregate databases as risking a "database of ruin" through the accretion of information, which could completely undermine individuals' privacy.[73] Further, like other privacy-related harms, re-identification risks extend beyond individuals. Aggregated de-identified and anonymized data also risks allowing persons to re-identify sensitive group and community-level information, which may also permit those persons to infer additional information that may result in collective harms.[74] For example, such harms could include discrimination against an identifiable group, possibly through resource allocation decisions that perpetuate or exacerbate discrimination in health care.[75]

AI increases re-identification risks. AI's big data analytics and predictive abilities allow persons, including malicious actors, to re-identify individuals from de-identified and anonymized data more efficiently and effectively. For example, deep learning tools can link a patient's chest x-rays, with up to 95.5% accuracy, including between images taken up to 10 years apart.[76] In this case, a person with access to an identifiable radiograph could likely link the respective individual to any large, publicly available anonym-

---

71 *Ibid* at 1705.

72 *Ibid* at 1744. In practice, personal information, de-identified information, and anonymized data exist along a spectrum of identifiability, from personal information to anonymized data, with de-identified data falling somewhere in between; and the extent to which data is de-identified, or whether it becomes anonymized, often depends on (1) whether AI is involved and (2) the data-related expertise of the users who access the data. See also Lubarsky, *supra* note 62 at 203.

73 See Ohm, *supra* note 55 at 1747, 1749.

74 See Office of the Privacy Commissioner of Canada, *Policy Proposals for PIPEDA Reform to Address Artificial Intelligence Report*, by Ignacio Cofone, (Ottawa, OPC, 2020) online: <priv.gc.ca> [perma.cc/6V3E-MC3B], s 1.

75 These harms have already manifested in broader social contexts, such as disenfranchising groups of voters and resulting risks to democracy—as was the case in the Cambridge Analytica scandal, see *ibid*, ss 3(c), 3(d).

76 Kai Packhäuser et al, "Deep Learning-Based Patient Re-Identification Is Able to Exploit the Biometric Nature of Medical Chest X-Ray Data" (2022) 12:14851 Scientific Reports 1 at 3.

ized radiograph data set in which that individual is represented, including images that contain sensitive patient information such as health conditions and treatment history.[77] In another study, while specific to consumer-generated health data, Liangyuan Na and colleagues re-identified up to 94.9% of adults and 87.4% of children from aggregated, anonymized physical activity cohort study data using various ML applications.[78] Further, health care images also often have direct or indirect identifiers "burned" into them, creating difficulties for persons to de-identify such data. While AI algorithms may assist persons in obscuring identifiable aspects of such images,[79] other persons may also use AI to try to reverse such obscuring. Like the cybersecurity issues discussed below, AI-related de-identification/anonymization and re-identification represents a technologically complex game of cat and mouse; and time will benefit malicious actors. Once a person has publicly released anonymized data or has otherwise relinquished control of it, that person can never improve the level of anonymization moving forward. It follows that AI-assisted re-identifying of individuals from anonymized data becomes easier over time but can never become more difficult.[80] As Professor Teresa Scassa notes, these time-dependent factors render 'anonymization' a moving target.[81]

In addition, AI tools themselves may inadvertently re-identify individuals from de-identified or anonymized data. Health care AI tools hold information in short-term memory and—as noted—optimize themselves based on training or new input data, leading to plausible risks that an AI tool could re-identify an individual intentionally or unintentionally by "reverse engineering" the data on which the algorithm was trained or to which it was otherwise exposed. The opaque nature of complex AI—or black boxes—exacerbates these privacy risks and transparency issues. We may not know exactly how such systems use or manipulate personal, de-identified, or anonymized

---

77 *Ibid* at 2.

78 Liangyuian Na et al, "Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning" (2018) 1:8 JAMA Netw Open 1 at 7.

79 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 Radiol J 107 at 112.

80 See Ohm, *supra* note 55 at 1757. See also Lubarsky, *supra* note 62 at 212.

81 See Scassa, "AI and Data Protection Law", *supra* note 56 at 133.

health care data.[82] Making matters worse, malicious actors may intentionally manipulate AI tools in this manner to reveal personal, de-identified, or anonymized health information on which those tools were trained or to which they were otherwise exposed, known as a "model inversion attack".[83] These model inversion attacks could lead to direct re-identification of an individual or enable easier re-identification in combination with other data.

B. Inferring Information About Individuals Through Personal Information

AI tools' predictive power may allow them to generate novel health care insights from personal health information or, as noted earlier, seemingly innocuous personal information. Inferences about one's health are clearly valuable for better health care, but only when the individuals for whom those inferences are made consent to such practices and want to know the resulting information. For example, an insurer could use a health care AI tool to predict that a patient is very likely to develop early onset amyotrophic lateral sclerosis, which a younger patient may not want to know. The federal *Genetic Non-Discrimination Act* prohibits, among other things, persons (including insurers) from requiring another person to undergo genetic testing and/or disclose the results of such testing as a condition of entering into a contract[84]—largely because of the predictive nature of genetic testing, the sensitivity of such information, and the resulting tangible and intangible harms arising from the unwanted use and disclosure of such information. However, in the health care AI space, no similar criminal prohibitions exist to protect patients from significant, unwanted health care-related inferences. Interestingly, privacy commissioners are beginning to determine that such inferences (although not necessarily in the health sector) are a "collection" of personal information[85] and, therefore, often require individuals' consent.

---

82 See Murdoch, Jandura & Caulfield, "Privacy Considerations", *supra* note 20 at 44.

83 See e.g. Canadian Centre for Cybersecurity, *Artificial Intelligence – IT-SAP.00.040 (*Ottawa: Can Centre for Cybersecurity, 2022), online: <cyber.gc.ca> [perma.cc/8ER8-PFHU].

84 *Genetic Non-Discrimination Act*, SC 2017, c 3, ss 3(1)(b), 4(1).

85 See Luca Lucarini & Sasha Coutu, "Generating Information with AI May Be Considered a *Collection* or Personal Information Under Privacy Law" (11 April 2023), online: <dentonsdata.com> [perma.cc/AV78-DU89]. See also *En-*

Robust, aggregated data increases these risks. When a person combines data sets, greater patient-level features will often allow a person to make broader health-related inferences about specific patients.[86] These risks are also interrelated with re-identification risks. A person may use new AI-generated information, in combination with other data, to re-identify individuals or identifiable groups, and link those individuals or groups to sensitive health care information.

C. Increased Cybersecurity Concerns and Data Breaches

Health care data breaches are also occurring more frequently in Canada, affecting both public and private sector actors.[87] While persons may unlawfully access individuals' personal health information in ways that amount to, for example, an intentional tort,[88] breaching a contract, or violating a statutory provision, this paper focuses on malicious data breaches: persons conducting cyberattacks against other individuals or organizations and malevolently accessing health care data. Malicious cyberattacks cause widespread privacy-related risks of harm through the loss of control of personal, de-identified, or anonymized health information. For example, in May 2023, a cyberattack on the Better Outcomes Registry & Network, a large network of mostly Ontario health care facilities, caused a breach of 3.4 million individuals' personal health information, including mothers seeking pregnancy care and newborn babies, dating back to 2010.[89] The group responsible for the attack breached MOVEit, the software that the Network used to transfer files. The group was then able to access at least one of the Network's servers. Making matters worse, the Network appears to have stored the informa-

---

quête concernant le Centre de services scolaire du- Val- des- Cerfs, 1020040-S, online: <decisions.cai.gouv.qc.ca> [perma.cc/7VSA-KLT6].

86 See Murdoch, Jandura & Caulfield, "Privacy Considerations", *supra* note 20 at 45.

87 See Vinyas Harish et al, "Cyberattacks on Canadian Health Information Systems" (2023) 195:45 CMAJ 1548 at 1548–50. See also David Burke, "Hospitals 'Overwhelmed' by Cyberattacks Fuelled by Booming Black Market", *CBC News* (2 June 2020), online: <cbc.ca> [perma.cc/D4PX-A3Y2].

88 See e.g. *Jones v Tsige*, 2012 ONCA 32 at paras 70–72, 90 [*Jones*].

89 Katie Dangerfield, "BORN Ontario Data Breach Left Health Data of Millions Exposed. What Went Wrong?", *Global News* (26 September 2023), online: <globalnews.ca> [perma.cc/UXS5-H8BX].

tion in a fully identifiable form, with personal identifiers, such as health card numbers and names, alongside sensitive health care information.[90] In 2019, 48% of all Canadian data breaches occurred in the health sector.[91] Similar to AI-assisted re-identification risks and the reference to a complex game of cat and mouse above, AI offers cybercriminals powerful tools to identify and exploit software vulnerabilities—leading to more efficient attacks.[92] Yet AI may also assist data custodians in better protecting health care data, including through stronger data encryption, breach detection, threat monitoring, and overall system security.[93] Data breaches, especially those concerning personal health information, exacerbate re-identification risks, including where actors use that personal information in combination with publicly available or otherwise easily accessible anonymized health data containing sensitive information.

Each of these interrelated privacy risks raise many significant deontological and consequential concerns, including possible psychological harm to individuals and groups and questions of insurability for individuals.[94] However, we must carefully balance privacy protections, or responses to these risks, with health care AI's big data needs and resulting health care benefits. Excessively strong privacy protections (through data minimization or pursuing 'perfect privacy'[95]) would pose barriers for creating robust, aggregated data sets and limit the availability of those data sets.[96] In turn, this approach would limit the availability of important health care data for secondary uses such as AI development, likely leading to biased, unsafe, and ineffective health care AI tools.[97] For example, developers may instead seek out and use data sets from other jurisdictions, which may not adequately

---

90 *Ibid*.

91 Harish et al, *supra* note 87 at 1548.

92 Gabriele Fiata, "Why Evolving AI Threats Need AI-Powered Cybersecurity", *Forbes* (4 October 2023), online: *Forbes* <forbes.com> [perma.cc/8W8H-9QV4].

93 *Ibid*.

94 See Murdoch, Jandura & Caulfield, "Privacy Considerations", *supra* note 20 at 45.

95 See Ohm, *supra* note 55 at 1752.

96 See Price & Cohen, *supra* note 27 at 42.

97 See Price, *supra* note 14 at 925.

represent Canadian subpopulations' health care needs.[98] On the other hand, inadequate privacy protections could lead to tangible and intangible harms and a loss of public trust, as well as potentially heavy-handed, reactive regulatory responses by governments.[99] So, the question then becomes: How do we generate and enable broader access to robust health care data for health care AI innovation while adequately protecting the privacy rights of individuals and groups?

Before discussing ongoing data protection legislative reform and possible policy approaches to better balance health care AI innovation and informational privacy, the following Part examines existing data protection laws in Canada, and Part V examines the constitutional constraints that have shaped those laws and ongoing legislative reform, which have historically limited federal policy approaches.

IV. A Patchwork of Data Protection Legislation Exists Across Canada

A myriad of Canadian laws protect individuals' personal information. Federally, the *Personal Information Protection and Electronic Documents Act* (PIPEDA)[100] applies to organizations across Canada in respect of personal information that those organizations collect, use, or disclose in the course of commercial activities (private-sector data protection legislation).[101] However, PIPEDA does not apply in provinces—in respect of organizations, classes of organizations, activities, or classes of activities—where the federal government has deemed provincial legislation substantially similar to PIPEDA.[102] For example, PIPEDA does not apply to private sector organizations in British Columbia, Alberta, and Quebec (except in relation

---

98 See Da Silva et al, "Regulating Safety", *supra* note 28 at 72.

99 Michael Da Silva et al, "Legal Concerns in Health-Related Artificial Intelligence: A Scoping Review Protocol" (2022) 11:123 Syst Rev 1 at 2.

100 See *PIPEDA*, *supra* note 50.

101 *Ibid*, s 4(1)(a). To note, PIPEDA also applies to federally regulated businesses' employees' personal information.

102 That is, through a Governor in Council Order in Council; however, PIPEDA still applies to federal works, undertakings, and businesses, and any organization that collects, uses, or discloses personal health information in the course of international or interprovincial trade and commerce (see *ibid*, s 26(2)(b)).

to personal information that those organizations transfer interprovincially or internationally).[103] Further, PIPEDA does not apply to private-sector organizations in Ontario, Newfoundland and Labrador, Nova Scotia, and New Brunswick in respect of personal *health* information. Here, the Federal Government has similarly determined that those provinces' health-sector laws are substantially similar to PIPEDA.[104] While all other provinces and territories have enacted health-sector data protection legislation, the Federal Government has not declared such legislation substantially similar to PIPEDA,[105] meaning that in those provinces and territories both PIPEDA and the respective provincial health-sector legislation apply to organizations that collect, use, or disclose personal health information in the course of commercial activities.

Further, PIPEDA generally does not apply to public-sector institutions. Federally, the *Privacy Act* governs federal government institutions'

---

[103] See *Organizations in the Province of British Columbia Exemption Order*, SOR/2004-220; *Organizations in the Province of Alberta Exemption Order*, SOR/2004-219; *Organizations in the Province of Quebec Exemption Order*, SOR/2003-374. See also *Personal Information Protection Act*, SBC 2003, c 63; [*British Columbia PIPA*]; *Personal Information Protect Act*, SA 2003, c P-6.5 [Alberta *PIPA*]; *Quebec Private Sector Personal Information Act*, *supra* note 2.

[104] See *Health Information Custodians in the Province of Ontario Exemption Order*, SOR/2005-399; *Personal Health Information Custodians in Newfoundland and Labrador Exemption Order*, SI/2012-72; *Personal Health Information Custodians in Nova Scotia Exemption Order*, SOR/2016-62; *Personal Health Information Custodians in New Brunswick Exemption Order*, SOR/2011-265. See also *PHIPA*, *supra* note 3; Newfoundland and Labrador's *Personal Health Information Act*, SNL 2008, c P-7.01 [*Newfoundland and Labrador PHIA*]; Nova Scotia's *Personal Health Information Act*, SNS 2010, c 41 [*Nova Scotia PHIA*]; New Brunswick's *Personal Health Information Privacy and Access Act*, SNB 2009, c P-7.05 [*New Brunswick PHIPAA*].

[105] See Alberta's *Health Information Act*, RSA 2000, c H-5 [*Alberta HIA*]. See also British Columbia's *E-Health (Personal Health Information and Protection of Privacy) Act*, SBC 2008, c 38. For an exhaustive list, see Blake Murdoch, Allison Jandura & Timothy Caulfield, *Privacy Concerns with Commercial Artificial Intelligence for Healthcare: Report Funded by the Office of Privacy Commissioner of Canada* (Edmonton: Health Law Institute, University of Alberta, 2021), online: <ualberta.ca> [perma.cc/4K2C-RWLJ] at 34–35 [Murdoch, *OPC Report*].

collecting of, using, or disclosing of personal information.[106] Provincially, legislation such as Ontario's *Freedom of Information and Protection of Privacy Act* governs personal information under provincial government institutions' control and health administrative data (for example, OHIP billing information).[107] At the time of this writing, all provinces and territories have enacted public-sector data protection legislation.[108] Interestingly, within a province such as Ontario, both public-sector and health-sector legislation may apply to certain organizations, such as public hospitals, and while PIPEDA generally does not apply to hospitals, as public-sector institutions, it may apply where hospitals engage in commercial activities outside of their core functions.

This patchwork of laws creates a web of obligations, especially for organizations that operate in more than one province and in the health and non-health sectors. In effect, amongst this web of laws, PIPEDA establishes a floor, or minimum set of rules, to protect personal information (and personal health information) in the private sector across Canada. However, this floor does not limit the provinces in responding to new or exacerbated privacy risks. The provinces remain free to deviate from PIPEDA, insofar as they (1) enact substantially similar legislation to PIPEDA, including by exceeding PIPEDA's protections, or (2) enact additional data protection legislation, while accepting that private sector organizations in the province would have to comply with both PIPEDA and the provincial legislation.

More specifically, these data protection laws protect individuals' *control* over their personal information.[109] These laws generally require entities to obtain individuals' consent before collecting, using, or disclosing those individuals' personal information, subject to certain exceptions.[110] For example,

---

106 RSC 1985, c P-21, ss 2, 4–6 [*Privacy Act*]; To note, PIPEDA also does not apply to not-for-profit charity groups, political parties, and political associations, as well as municipalities, universities, schools, or hospitals—unless these latter four types of organizations engage in commercial activities beyond their core activities.

107 *Freedom of Information and Protection of Privacy Act,* RSO 1990, c F. 31, ss 1(a)–(b), 2(1) [*ON FIPPA*]

108 See Murdoch, *OPC Report*, *supra* note 105 at 35–36.

109 To note, private-sector data protection laws also generally establish requirements for informational security.

110 However, to note, under public-sector data protection legislation consent is not

exceptions to consent in the private sector often include situations where seeking consent would be impossible or impractical, such as for certain medical, legal, or security-related purposes, or where an entity no longer has a direct relationship with the individual.[111] Further, under private-sector data protection legislation, individuals can generally withdraw consent;[112] public and private-sector laws usually only permit public and private entities to use or disclose personal information in accordance with the purposes for which it was originally collected or for appropriate purposes,[113] respectively, or for other purposes with the respective individuals' consent (again, subject to certain exceptions).[114]

However, existing data protection laws predominantly rely on ombuds models, which have "notoriously weak enforcement mechanisms."[115] These laws' remedies often include compliance agreements,[116] rights for complainants to apply to a court for redress,[117] and narrow offences with low penalties.[118] For example, under PIPEDA, organizations may be found guilty of an offence if they knowingly either (1) fail to retain personal information while an individual has made a request for that information, (2) fail to report to the Privacy Commissioner any security breaches likely to result in significant harm to an individual (or retain records of such breaches), (3) punish employee-whistleblowers, or (4) obstruct the Commissioner's investigation of a complaint or conducting of an audit. Yet, on summary conviction, the organization is only liable to a fine of up to $10,000, and, upon indictment,

---

the basis for the collection of personal information; here, government institutions collect personal information in accordance with statutorily authorized purposes, including for purposes that directly relate to government programs and activities.

111 See e.g. *PIPEDA*, *supra* note 50, Schedule 1, c 4.3.

112 See e.g. *ibid*, Schedule 1, c 4.3.8.

113 Generally meaning purposes that a reasonable person would consider appropriate in the circumstances.

114 See e.g. *PIPEDA*, *supra* note 50, ss 5(3), 7. See also, *Privacy Act*, *supra* note 106, ss 7, 8.

115 See Scassa, "AI and Data Protection Law", *supra* note 56 at 130–31.

116 See e.g. *PIPEDA*, *supra* note 50, s 17.1.

117 See e.g. *ibid*, s 14.

118 See e.g. *ibid*, s 28.

the organization is only liable to a fine of up to $100,000.[119] These offences are too narrow, and the resulting penalties are insufficient to deter large, multinational organizations from contravening PIPEDA.

Health-sector data protection laws, such as Ontario's *Personal Health Information Protection Act, 2004* (PHIPA),[120] differ from private and public-sector data protection laws in several respects. Health-sector specific legislation applies to both public and private-sector health data custodians (persons or organizations), including hospitals, long-term care homes, labs, physician offices, pharmacies, insurance companies, and other similar entities in control of personal health information in the course of their powers, duties, or work.[121] Under health-sector data protection legislation, an individual's consent for a data custodian to collect, use, or disclose that individual's personal health information may be express or implied, except in certain circumstances where the consent must be express. For example, consent must be express where a data custodian discloses an individual's personal health information to (1) a person that is not a health data custodian or (2) a health data custodian for purposes other than providing or assisting in providing health care to the individual.[122] Overall, health-sector data protection legislation generally establishes a higher (more stringent) bar concerning individual consent in most circumstances, especially where data custodians seek to *use* or *disclose* personal health information, which reflects the sensitivity of personal health information.

Ontario's PHIPA contains several exceptions to individuals' consent rights, which are particularly important for the purposes of this paper. Under PHIPA, data custodians may use an individual's personal health information for research purposes without that individual's consent where the data custodian prepares a research plan and receives research ethics board approval in respect of that plan.[123] Further, data custodians may disclose an individual's personal health information to a third party researcher if that researcher has submitted a research ethics board-approved research plan to the data custodian, and the data custodian enters into a compliance agree-

---

119 *Ibid*.

120 *PHIPA*, *supra* note 3.

121 *Ibid*, s 3.

122 *Ibid*, s 18(3).

123 *Ibid*, ss 37(1)(j), 37(3).

ment with the researcher.[124] In both cases, PHIPA largely delegates a 'privacy risk assessment' to research ethics boards. For example, under PHIPA, a research ethics board must consider (1) whether such disclosure to a third party researcher is necessary to achieve the objectives of the research, (2) whether the third party has adequate safeguards in place to protect the confidentiality of the information, (3) the public interest of the research and of the privacy risks, and (4) whether obtaining the respective individuals' consent would be impractical.[125] PHIPA also requires the third party to, among other things, comply with any conditions imposed by the research ethics board and to only use the information for the purposes set out in the research ethics board-approved research plan. "Research," in respect of these provisions, is broad and appears to include developing health care AI where such development is systematic in nature and is intended to contribute to generalizable knowledge.[126]

Compared to PIPEDA, PHIPA also contains stronger, more modern enforcement mechanisms. PHIPA provides for administrative monetary penalties to encourage compliance.[127] Further, PHIPA's offences are very broad, and include any wilful collection, use, or disclosure of personal health information in contravention of the Act or subordinate regulations.[128] Any natural person who is guilty of an offence is liable to a fine of up to $200,000 and/or a term of imprisonment of up to 1 year, and any organization that is guilty of an offence is liable to a fine of up to $1,000,000.[129]

Under PHIPA, health care AI developers may seek to access health information in several ways. Personal health information is generally more feature-rich than de-identified or anonymized information, yet data protection laws generally limit AI developers' access to such information. For ac-

---

124 *Ibid*, ss 37(1)(j), 44(1).

125 *Ibid*, s 44(3).

126 *Ibid*, s 2, where "research" means "a systematic investigation designed to develop or establish principles, facts or generalizable knowledge, or any combination of them, and includes the development, testing and evaluation of research."

127 *Ibid*, s 61.1(1)(a). See also *O Reg 329/04,* s 35, as amended by *O Reg* 342/23, s 1.

128 *PHIPA*, *supra* note 3, s 72(1)(a); *O Reg 329/*04, *supra* note 127.

129 *Ibid*, s 72(2); *O Reg 329/*04, *supra* note 127.

cess to personal health information, developers (including private corporations) may (1) seek individuals' explicit consent, (2) enter into a partnership with the data custodian, for the data custodian to develop an AI tool in accordance with a research ethics board-approved research plan, or (3) enter into a third-party research agreement with the data custodian in accordance with a research ethics board-approved research plan.[130]

While the collection of personal health information for AI-related purposes raises several important questions,[131] this paper focuses on data custodians' use and disclosure of health information that they have already collected. AI developers often seek out existing data sets rather than proactively or retroactively seeking express consent from individuals because seeking individual consent adds administrative burden through considerable costs and time to health care AI research and development.[132] It may also lead to selection biases, undermining representative data.[133] In addition, many AI developers seek out anonymized health care data for AI research and development because seeking more onerous research ethics board approvals for AI research concerning personal health information adds considerable costs and time to such research.[134]

Further, beyond the regulation-making authority and specific prohibition in PHIPA discussed below, anonymized data generally falls outside the

---

130 To note, developers may also enter into collaborations, partnerships, or data sharing agreements with entities to which personal health information may be disclosed under PIPEDA, such as the Canadian Institute for Health Information and the Institute for Clinical Evaluative Sciences, see *PHIPA*, *supra* note 3 ss 29, 39, 44–45. See e.g. *O Reg* 329/04, *supra* note 127, ss 18.1–18.1.2.

131 Such as whether data custodians, intentionally or unintentionally, blur individuals' consent to care and those individuals' consent for data custodians to collect, use, or disclose their personal health information for secondary AI-related purposes.

132 Francis McKay, Darren Treanor & Nina Hallowell, "Inalienable Data: Ethical Imaginaries of De-identified Health Data Ownership" (2023) 4 Qualitative Research in Health 1 at 2; Mark A Rothstein, "Is Deidentification Sufficient to Protect Health Privacy in Research" (2010) 10:9 Am J Bioethic 3 at 2.

133 See Bak et al, *supra* note 34 at 3.

134 McKay, Treanor & Hallowell, *supra* note 132 at 2; Rothstein, *supra* note 132 at 2.

scope of most health-sector (and other sector) data protection legislation.[135] Data custodians can use or disclose anonymized health data without requiring individuals' consent. These data custodians, as anonymized data 'access points,' commonly include hospitals, data sharing platforms, and private electronic health record vendors. In addition, a recent decision by the Information and Privacy Commissioner of Ontario will likely cause more data custodians to anonymize personal health data under their control, proliferating the availability and accessibility of health care data.[136] In this decision, the Commissioner determined that, while the act or process of anonymizing personal health information is a "use" under PHIPA, the Act permits data custodians to use personal health information in this manner without individuals' consent. More specifically, the Commissioner determined that anonymizing personal health information (in this case, to sell that information to third parties) amounted to the data custodian disposing of the information or modifying it in order to conceal the identity of the individuals.[137]

Under PHIPA, a data custodian does not require an individual's consent to dispose of or modify that individual's personal health information to conceal their identity, where the data custodian does so in a manner consistent with Part II of PHIPA.[138] Part II requires, among other things, that the data custodian provide a general description of its practices through a publicly available written statement in a practical manner. It also requires that the data custodian take reasonable steps to ensure the protection of the personal health information being anonymized—such as through terms in a sales contract that would require the receiving party to implement sufficient privacy and security protocols.[139] While PHIPA treats de-identified (read

---

135 See *PHIPA*, *supra* note 3 ss 4(1), 4(2). See also *Quebec Private Sector Personal Information Act*, *supra* note 2; *British Columbia PIPA*, *supra* note 103, ss 1–3; *Nova Scotia PHIA*, *supra* note 104, ss 3(g), 3(r), 5(1), 5(2); *New Brunswick PHIPAA*, *supra* note 104, ss 1–3(2); *Newfoundland and Labrador PHIA*, *supra* note 104, ss 2(p), 5, 6, 8.

136 See Information and Privacy Commissioner, *PHIPA Decision 175* (Ontario: IPC, 2022), online: <decisions.ipc.on.ca> [perma.cc/5NM7-L5S4] [PHIPA IPC 175].

137 *Ibid*, at para 14. See also *PHIPA*, *supra* note 3, s 37(1)(f).

138 See *PHIPA*, *supra* note 3, s 37(1)(f).

139 See PHIPA IPC 175, *supra* note 136. Problematically, many individuals may not be aware of such notices and that the respective data custodian intends to sell (use) their personal health information for secondary purposes once ano-

as pseudonymized, not anonymized) data as personal health information, relevant federal law reform—as discussed further below—would provide new exceptions to individual consent for this type of data under certain circumstances. Such reform may also influence future provincial health-sector legislative reform. It follows that the remainder of this paper focuses on the governance of anonymized and de-identified health care data.

In contrast to PIPEDA and several provincial data protection laws, Ontario's PHIPA provides for some (currently limited) oversight of anonymized data. PHIPA prohibits persons from using or attempting to use anonymized data to identify an individual, alone or in combination with other information (re-identifying individuals from anonymized data), except where permitted under that Act or another Act.[140] For example, PHIPA permits the data custodian that anonymized personal health information to re-identify individuals from that data.[141] Under the Act, a person who wilfully contravenes this prohibition is guilty of an offence subject to the high penalties for offences described earlier.[142] Additionally, the Act permits the Lieutenant Governor in Council to make regulations governing the anonymization of personal health information and the collection, use, and disclosure of anonymized information by health custodians and any other persons.[143]

Several provincial health-sector data protection laws, such as Newfoundland and Labrador's *Personal Health Information Act* and *Alberta's Health Information Act*, do not appear to directly prohibit re-identifying individuals from or by using anonymized data. In addition, they do not ap-

---

nymized; however, individuals also often support others' using their personal health information for health research, although this level of support varies depending on whether the actor conducting the research is a public or private-sector person or organization, see e.g. Paprica, *supra* note 7. Public support for secondary uses of personal health information is also contingent on privacy protections and anonymization, see Timothy Caulfield, Blake Murdoch & Ubaka Ogbogu, "Research, Digital Health Information and Promises of Privacy: Revisiting the Issue of Consent" (2020) 3:1 Can J Bioethics 164 at 165.

140 *Supra* note 3, s 11.2(1). To note, PHIPA refers to "de-identified information" to mean what this paper considers "anonymized data"—so this paper uses "anonymized data" in respect of the legislation.

141 *Ibid*, s 11.2(2).

142 *Ibid*, s 72(1)(b.1).

143 *Ibid*, s 73(1)(o.2).

pear to provide their respective executive branches of government with regulation-making authority to govern anonymized data.[144] Arguably, the act of re-identifying an individual from anonymized data without lawful excuse and without an individual's consent would likely constitute an unlawful 'collection' of personal health information under these Acts.[145] However, both Acts' personal information collection rules apply to "data custodians." Consequently, a private party, such as a malicious actor, that re-identifies an individual from or by using anonymized data does not appear to fall under the definition of a "data custodian" and would, therefore, not be subject to either Acts' data collection rules. Further, under Newfoundland and Labrador's legislation, fines for contravening the Act do not exceed $10,000.

Health care AI's data needs, outdated data protection legislation, exacerbated AI-related privacy risks, and the proliferation of anonymized health care data will likely expose individuals and groups in Canada to privacy-related harms—especially to risks of re-identification. In response and as noted above, we should not seek to minimize the availability of anonymized data. We should instead enable broad access to anonymized health care data (and, in some cases, de-identified health care data), while matching this increased access with stronger privacy protections backed by high penalties for non-compliance. Examples of such protections include more stringent cybersecurity rules, more rigorous anonymization standards, and contextual privacy risk assessments before using or disclosing such data in certain circumstances. Before discussing these de-identified and anonymized data policy reform options in greater detail in Part VII, the following Part explores constitutional constraints that would limit (and that have limited) such reform.

---

144 *Newfoundland and Labrador PHIA*, *supra* note 104; RSA 2000, c H-5. See also *Alberta HIA*, *supra* note 105.

145 Mark Phillips, Edward S Dove & Bartha Knoppers, "Criminal Prohibition of Wrongful Re-identification: Legal Solution of Minefield for Big Data?" (2017) 14:4 Bioethical Inquiry 527 at 531.

V. Constitutional Constraints Limit Federal AI and Data Protection Legislative Reform

Formulating new and effective policy requires a strong understanding of Canadian federalism and Canada's constitutional constraints, especially when designing policy for new technologies such as health care AI (and its related privacy risks)—which Canada's division of powers do not explicitly enumerate, and which may fall under multiple legislative heads of powers. As a federal state, Canada's Constitution distributes legislative powers between Parliament and the provincial legislatures, enabling unified federal and diverse provincial policy approaches. It follows that constitutional constraints may "impact the speed at which regulation can evolve, and the comprehensiveness of that regulation."[146] New policy approaches with a weak or uncertain constitutional footing also risk early judicial challenges from the other level of government, through those governments' respective reference powers,[147] or from parties with standing. In addition, where courts find legislation of no force or effect for falling outside of Parliament's or a provincial legislature's jurisdiction, the resulting legislative gaps may expose Canadians to continued risks of harm. Federally, enacting comprehensive policy in a timely manner that still respects provincial jurisdiction, autonomy, and diversity poses significant challenges for Parliament.

Historically, the provinces have enjoyed broad legislating authority over privacy and data protection under their plenary power over property and civil rights[148] and, to a lesser extent, their power over matters of a merely local or private nature.[149] However, Parliament and the provinces share several jurisdictional overlaps in respect of data protection, privacy, health, and human rights.[150] While the Constitution's enumerated heads of pow-

---

146 Teresa Scassa, "Regulating AI in Canada: A Critical Look at the Proposed Artificial Intelligence and Data Act" (2023) 101:1 Can Bar Rev 1 at 21 [Scassa, "Regulating AI"].

147 On government reference powers, see Carissima Mathen, *Courts Without Cases: The Law and Politics of Advisory Opinions* (Oxford: Hart Publishing, 2019) ch 3.

148 See *Constitution Act*, *1867* (UK), 30 & 31 Vict, c 3, s 91, reprinted in RSC 1985, Appendix II, No 5, s 92(13) [*CA, 1867*].

149 *Ibid*, s 92(16).

150 See Florian Martin-Bariteau & Teresa Scassa, "Artificial Intelligence and the Law in Canada" in Florian Martin-Bariteau & Teresa Scassa, *Artificial Intelli-*

ers are exclusive, Parliament and the provinces may enact laws that address different aspects of a matter. For example, while Parliament prohibits persons from possessing cannabis plants and the cultivation of such plants for personal purposes except where a person possesses and cultivates no more than four plants for personal use (a criminal law with exceptions), the provinces can simultaneously, and lawfully, prohibit persons from cultivating cannabis plants in dwelling-houses (a property and civil rights location-based restriction).[151] This plenary power overlap is referred to as the double aspect doctrine, which allows Parliament and the provinces to broadly address policy issues *respecting* their spheres of jurisdiction. Problematically, the scope of Parliament's powers raises important federalism concerns, primarily because where federal and provincial laws conflict, federal law will prevail under the paramountcy doctrine, although courts are hesitant to find 'true' conflicts. Returning to the cannabis example, while a federal law may permit an activity, a provincial law could still likely prohibit or regulate aspects of that activity for different purposes (such for purposes related to civil rights). In the data protection space, this likely means that where a federal private-sector data protection law permits an organization to disclose de-identified health information without an individual's consent, provincial health-sector legislation could still likely limit or prevent such disclosure. Only if a valid federal law were to *require* or *prohibit* something contrary to a provincial rule, would the federal law likely render the provincial law inoperable under federal paramountcy, insofar as they conflict.[152]

Concerning privacy and data protection legislation, constitutional constraints have historically limited the 'floor' of protection that Parliament can establish across Canada. In enacting PIPEDA, Parliament has relied on its power over the regulation of trade and commerce[153]—hence the law's focus on activities "in the course of commercial activity." More specifically, Parliament has relied on its "general regulation of trade and commerce" branch of this power, rather than the narrower branch concerning interprovincial

---

gence and the Law in Canada (Toronto: LexisNexis Canada, 2021) at 9, online: <papers.ssrn.com> [perma.cc/6964-7F5N].

151 See *Murray-Hall v Quebec (Attorney General)*, 2023 SCC 10 [*Murray-Hall*].

152 However, in addition to situations where a person could not comply with both laws simultaneously, a true conflict also includes where a provincial law frustrates the purpose of a federal law. See *Multiple Access Ltd v McCutcheon*, [1982] 2 SCR 161 at para 163.

153 See *CA, 1867*, *supra* note 148, s 91(2).

and international trade. While the general branch of Parliament's trade and commerce power is broader in scope, legislation enacted under it must (1) be part of a general regulatory scheme, (2) fall under the continuous oversight of a regulatory agency, (3) concern trade as a whole (rather than a specific industry), (4) be of such a nature that the provinces, acting alone or in concert, would be constitutionally incapable of enacting it, and (5) be of such a nature that the failure to include one or more provinces in the scheme would jeopardize its successful operation in other provinces.[154] It follows that federal legislation concerning general trade and commerce, such as PIPEDA (and depending on the nature of the regulatory scheme), rests on a weaker constitutional footing than if Parliament were to limit legislation to interprovincial and international trade. In other words, Parliament's general trade and commerce power's more nuanced legal test and the intraprovincial effects of laws enacted under this power render such laws more susceptible to provincial challenges. For example, in 1993, Quebec referred the constitutionality of PIPEDA to the Quebec Court of Appeal as intruding on the provinces' plenary power over property and civil rights,[155] even though the Federal Government had deemed Quebec's pre-existing private-sector privacy legislation as substantially similar to PIPEDA. However, Quebec abandoned its challenge and, as of this writing, PIPEDA remains valid.

Two other federal heads of power offer Parliament potential avenues to address health care AI-related privacy issues: Parliament's power over peace, order, and good government of Canada (POGG)[156] and its power over criminal law.[157] Parliament's POGG power allows it to enact laws for the "Peace, Order and good Government of Canada, in relation to all Matters not coming within the Classes of Subjects by [the *Constitution Act, 1867*] assigned exclusively to the Legislatures of the Provinces."[158] So, any subject not falling within provincial jurisdiction would fall under Parliament's

---

154 See *General Motors of Canada Ltd v City National Leasing*, [1989] 1 SCR 641 at 662–63. See also *Reference re Securities Act*, 2011 SCC 66 at paras 83–84. See also Canada, Library of Parliament, *Parliament's General Trade and Commerce Power*, by Francis Lord, Publication No. 2018-32-E (Ottawa: LP, 2019) at 4, online: <lop.parl.ca> [perma.cc/6MQ6-ZKXG].

155 See Josh Nisker, "PIPEDA: A Constitutional Analysis" (2006) 85:2 Can Bar Rev 317 at 318.

156 See *CA, 1867, supra* note 148, s 91.

157 *Ibid*, s 91(27).

158 *Ibid*.

jurisdiction. Parliament's criminal law power, on the other hand, allows it to exclusively prohibit, or prohibit and regulate, behaviours or actions with injurious or undesirable effects on various public interests, backed by penalties.

The courts have interpreted Parliament's POGG power to include three branches, two of which are relevant to health care AI and data protection: (1) the residual/gap branch and (2) the national concern branch.[159] The residual/gap branch captures new matters that the *Constitution Act, 1867* did not assign exclusively to Parliament or the provinces. However, in most cases, courts will find that an aspect of an emerging issue, such as those related to health, fall under existing federal and/or provincial powers depending on the respective law's nature and scope and the issue at hand.[160] New technologies such as AI and its health care-related privacy risks would likely fall under existing constitutional heads of power, including the provinces' plenary power over property and civil rights. POGG's second relevant branch, the national concern branch, raises more interesting questions for health care AI, and AI and data protection more broadly. This branch generally provides Parliament with legislative power over matters (1) of a sufficient concern to Canada as a whole, (2) that have a "singleness, distinctiveness and indivisibility that clearly distinguishes it from matters of a provincial concern,"[161] and (3) that have a scale and impact on provincial jurisdiction reconcilable with Canada's division of powers.[162] Once the courts find that a matter falls within Parliament's POGG national concern branch, the mat-

---

159 The third branch concerns legislative power to address emergencies of a temporary nature, see *Re: Anti-Inflation Act*, [1976] 2 SCR 373 at 423, 427.

160 See *Reference re Genetic Non-Discrimination Act*, 2020 SCC 17 at paras 22, 24, 66 [*Ref re GNDA*]. See also *Schneider v British Columbia*, [1982] 2 SCR 112 at 141–42. This overlap of exclusive heads of power is known as the double aspect doctrine. See also *Murray-Hall*, *supra* note 151.

161 Generally, matters that have a "singleness, distinctiveness and indivisibility" refers to specific and identifiable matters qualitatively different from matters of provincial concern, where evidence establishes the provinces' inability to deal with the matter. A provincial inability generally means that the provinces would not be able to deal with the matter jointly or severally, because the failure of one or more provinces to address the issue would cause the other provinces to fail in addressing the issue, leading to grave extra-provincial consequences, see *References re Greenhouse Gas Pollution Pricing Act*, 2021 SCC 11 at para 145 [*Ref re GGPPA*].

162 *Ibid* at para 160.

ter permanently becomes one of federal jurisdiction, such as aeronautics,[163] marine pollution,[164] and minimum national standards of greenhouse gas price stringency to reduce greenhouse gas emissions.[165]

While AI and its privacy-related risks pose significant policy challenges for the provinces, in addressing such risks individually or in concert, Parliament would likely face significant constitutional hurdles in convincing the courts that health care AI's privacy-related risks are grounded in its POGG national concern power. As journalist Aidan Macnab has aptly noted, borders cannot contain AI and AI-related issues,[166] yet the provinces have historically regulated private, public, and health-sector personal information and related privacy risks. While AI significantly worsens health care privacy-related risks, the provinces could likely modernize their approaches to protecting personal information and personal health information, whether individually or in concert, to mitigate such risks. Further, if health care AI (or AI)-related privacy risks could ground Parliament's POGG national concern power, then this would permanently shift such matters under federal jurisdiction. Such a shift would significantly intrude on the provinces' jurisdiction and autonomy. We also have to consider the very real possibility that Parliament could enact inadequate or unacceptable laws, especially in a space like AI where powerful private companies have a strong interest in heavily influencing the related discourse and legislative initiatives to their advantage, possibly to the detriment of others.

More importantly for the purposes of this paper, Parliament could likely rely on its criminal law power to prohibit or prohibit and regulate (through exceptions) certain aspects of health care AI and its resulting privacy risks. Parliament's criminal law power is broad, but, as noted above, a valid criminal law must (1) contain a prohibition, (2) backed by a penalty, and (3) be directed at a criminal law purpose.[167] A valid criminal law purpose is

---

163 See *Quebec (AG) v Canadian Owners and Pilots Association*, 2010 SCC 39 at paras 28–30.

164 See *R v Crown Zellerbach Canada Ltd*, [1988] 1 SCR 401 at 436–38.

165 See *Ref re GGPPA*, *supra* note 161 at para 207.

166 See Aidan Macnab, "Anchoring AI Legislation in Criminal Law Puts It in a 'Precarious Position:' Researchers" (13 September 2023), online (commentary): <canadianlawyermag.com> [perma.cc/MJ34-375A].

167 See *Reference re Validity of Section 5 (a) Dairy Industry Act*, [1949] SCR 1 at 49–50.

one in which a law targets a threat of harm to "a public interest traditionally protected by the criminal law, such as peace, order, security, health and morality, or to another similar interest."[168] While the scope of this public-interest list is not overly contentious, the Supreme Court was split in two recent cases on a general threshold of harm upon which Parliament's criminal law power must be grounded.[169] In both cases, the Court was split on whether Parliament could act on "any reasoned apprehension of harm" or "a concrete basis and reasoned apprehension of harm" in relation to health. The latter threshold suggests a role for empirical evidence to ground Parliament's criminal law power when it responds to threats of harm to health, or health-related interests, such as psychological suffering or distress.[170] While drawing a clear line between psychological discomfort and psychological suffering or distress is a difficult task, especially for harms that may manifest at the group-level, two aspects of potential criminal legislation would likely strengthen its constitutional footing: (1), as noted above, empirical evidence that quantitatively or qualitatively circumscribes such harm, or (2) by targeting threats of harm to morality in the health-privacy space, including by prohibiting behaviours or actions with specific *mens rea* elements. For example, in Ontario, a person who unjustifiably accesses another individual's medical files contravenes provincial data protection laws, and this action grounds a private right of action—intrusion upon seclusion.[171] These matters clearly fall under provincial jurisdiction over property and civil rights. While Parliament could unlikely prohibit a person from unjustifiably accessing another individual's medical files, it could likely prohibit more specific, morally reprehensible conduct (i.e., a morally undesirable behaviour rather than, exclusively, the resulting harm). More specifically, Parliament could likely prohibit a person from *wilfully* accessing another individual's sensitive personal health information *for the purposes of* exploiting, deceiving, or causing emotional distress to that person. Both *mens rea* elements – *wilfully* accessing sensitive personal health information and doing so *for the purposes of* exploiting, deceiving, or causing distress—rep-

---

168 See *Ref re GNDA*, *supra* note 160 at para 71.

169 *Ibid* at paras 260–61. See also *Reference re Assisted Human Reproduction Act*, 2010 SCC 61 at paras 55–56.

170 Ubaka Ogbogu, "The Assisted Human Reproduction Act Reference and the Thin Line Between Health and Crime" (2013) 22:1 Const Forum Const 93 at 96. See also Eric M Adams, "Touch of Evil: Disagreements at the Heart of the Criminal Law Power" (2022) 104:4 SCLR 67 at 87–88.

171 See *Jones*, *supra* note 88.

resent a threat of harm to morality. Parliament could likely respond to such conduct before (1) the harm were to materialize and (2) researchers were to quantify that harm.[172]

However, like POGG, Parliament's criminal law power also poses federalism issues. While a single uniform law for data protection (including health care data protection) may significantly decrease administrative burden, reduce uncertainty, and increase regulatory compliance, Canada's division of powers precludes such an approach. Canada's federal framework instead enables laboratories of democracy capable of balancing competing interests within each jurisdiction. Across Canada, the provinces can implement, analyze, and refine different approaches while also enabling federal-provincial cooperation. It follows that Parliament should be cautious when establishing its federal 'floor' in the data protection space, especially where it seeks to implement prescriptive rules that may conflict with (sometimes more privacy-protective) provincial rules.

VI. Ongoing Legislative Reform and Existing Laws Fail to Properly Mitigate Privacy Risks

Ongoing law reform would continue to leave several health care privacy-related gaps unaddressed.[173] Federal law reform efforts would replace PIPEDA with a modernized *Consumer Privacy Protection Act* (CPPA), introduce the *Personal Information and Data Protection Tribunal Act* (an administrative body that would hear appeals arising from the Privacy Commissioner of Canada's decisions under the proposed CPPA), and introduce the *Artificial Intelligence and Data Act* (AIDA).[174] Like PIPEDA, the proposed CPPA appears to rely on Parliament's general trade and commerce power—by again focusing on organizations that collect, use,

---

172 See generally *Criminal Code*, RSC 1985, c C-46, s 430 [*Criminal Code*].

173 Importantly, during the editing process of this paper, Bill C-27 died on the *Order Paper* when the Governor General proclaimed Parliament prorogued on January 6, 2025, on the advice of the Prime Minister (*Proclamation Proroguing Parliament to March 24, 2025*, Proclamation, 6 January 2025, SI/2025-9, C Gaz II, 159, Extra Number 1). However, this paper's discussion and policy recommendations in relation to Bill C-27 are still instructive, especially if the next Federal Government were to introduce a similar bill during the next or another future Parliamentary session.

174 See *Bill C-27*, *supra* note 2.

or disclose personal information in the course of commercial activities.[175] While the proposed Act recognizes that data constantly flows across borders and geographical boundaries, Parliament did not expand the proposed law's scope to capture a broader range of entities under, for example, its POGG national concern power. Similarly, Parliament does not appear to have relied on its POGG national concern power for the proposed AIDA. Instead, AIDA relies on Parliament's narrower trade and commerce power over interprovincial and international trade and commerce. The proposed Act specifically targets certain AI-related activities, discussed in greater detail below, *in the course of international and interprovincial trade*. Furthermore, as discussed in greater detail below, AIDA also relies, seemingly exclusively, on Parliament's criminal law power for three of its provisions. Regardless, both the proposed CPPA and AIDA represent conservative approaches within Canada's federal framework, likely capable of withstanding judicial scrutiny, which, unfortunately, would not address several health care privacy-related gaps. For example, the CPPA, like PIPEDA, would not apply to not-for-profit groups, and AIDA would not apply to any person conducting AI and data-related activities solely within a province, including in the health care sector.

The proposed legislative changes in the CPPA are vague and may not reflect how data anonymization works in practice or reality. The CPPA, like many existing data protection laws, would not apply to anonymized data;[176] however, even though the proposed Act would *explicitly* exclude anonymized data, it still appears to apply to the process of anonymizing data *until* the data becomes anonymized (like existing data protection legislation, but which does so implicitly). The proposed Act's definition of anonymized data would impose a higher standard than that of PIPEDA and provincial data protection laws. Under the CPPA, "anonymize" means "to irreversibly and permanently modify personal information, in accordance with generally accepted best practices, to ensure that no individual can be identified from the information, whether directly or indirectly, by any means."[177] This standard removes the "reasonably foreseeable in the circumstances" language that exists in federal and provincial statutes and through federal case law. The CPPA essentially creates a very high, absolute standard, yet, based on the illusory nature of anonymization, an organization may rarely, or never,

---

175 *Ibid*, cl 2, s 6(1)(a).

176 *Ibid*, cl 2, ss 2(1), 6(5).

177 *Ibid*, cl 2, s 2(1).

be able to meet this standard. Alternatively, if an organization were to meet this standard, the resulting data would likely be useless for developing health care AI, because it would include so few features. In addition, "anonymize" also relies on "generally accepted best practices," and the proposed Act does not define what this phrase means, nor does it provide the Governor in Council with regulation-making power to further define "anonymize" or govern the use of anonymized data. These approaches under the proposed CPPA diverge from, for example, Ontario's PHIPA, which allows the Lieutenant Governor in Council to define anonymized data and govern its use.[178] And, "generally accepted best practices," alongside 'irreversible' and without 'reasonably foreseeable in the circumstances' language, could impede organizations from implementing novel, yet more stringent, anonymization techniques.[179]

The proposed CPPA would also create a new statutory class of personal information: de-identified information. Similar to the definition of de-identified data described earlier in this paper, "de-identify" under the CPPA means "to modify personal information so that an individual cannot be directly identified from it, though a risk of the individual being identified remains."[180] The CPPA also clarifies that, except in respect of certain provisions, de-identified information is still personal information,[181] and that organizations must ensure that any technical and administrative measures that they apply to de-identify personal information are "proportionate to the purpose for which the information is de-identified and the sensitivity of the personal information"[182]—yet the proposed CPPA stops short of requiring such measures to be proportionate to re-identification risks.[183] Importantly,

---

178 *Supra* note 3, s 73(1)(o.2).

179 See Barry B Sookman, "CPPA: Problems and Criticisms – Anonymization and Pseudonymization of Personal Information" (6 December 2022) at 4, online (blog): <mccarthy.ca> [perma.cc/N8H7-JNX4]. See also Canadian Anonymization Network (CANON), "Proposed Amendments to De-Identification and Anonymization Provisions in the *Digital Charter Implementation Act, 2022* (*Bill C-27*)" (24 May 2023) at 2–3, online (pdf): <deidentify.ca> [perma.cc/MY68-N42Z].

180 See *Bill C-27*, *supra* note 2 at cl 2, s 2(1).

181 *Ibid,* cl 2, s 2(3).

182 *Ibid,* cl 2, s 74.

183 Office of the Privacy Commissioner, *Submission of the Office of the Privacy*

the proposed CPPA would allow organizations to use individuals' personal information without their knowledge or consent to de-identify that information (and, presumably, to also anonymize that information).[184] Further, the CPPA would allow organizations to use individuals' personal information without their knowledge or consent for the organization's internal research, analysis, and development purposes, if the organization first de-identifies the personal information.[185]

The CPPA would also permit organizations to use and disclose de-identified information without individuals' consent for several other purposes. Like PIPEDA, the CPPA would allow organizations to disclose personal information to other organizations that are parties to a prospective business transaction, subject to certain privacy-related contractual obligations between the parties, and only in respect of information necessary for the prospective transaction.[186] However, under the CPPA, the organization must first de-identify the personal information. In addition, an organization may disclose de-identified information without individuals' knowledge or consent where they disclose such information to (1) a government institution or part of a government institution in Canada, (2) a health care institution, post-secondary educational institution, or public library in Canada, (3) organizations carrying out a socially beneficial purposes on behalf of the federal or a provincial government, or (4) any other prescribed entity (in subordinate regulations), but, in each case, only for a socially beneficial purpose.[187] The CPPA defines socially beneficial purpose broadly—its definition includes purposes "related to health."[188]

Lastly, the CPPA would prohibit organizations from using de-identified information, alone or in combination with other information, to identify an individual. Exceptions to this prohibition would include where an organization tests the effectiveness of its de-identification process, complies

---

Commissioner of Canada on Bill C-27, the Digital Charter Implementation Act, 2022 (Ottawa: OPC, 2023) at 13, online: <priv.gc.ca> [perma.cc/ES7Z-6MDR].

184 See *Bill C-27*, *supra* note 2 at cl 2, s 20.

185 *Ibid*, cl 2, s 21.

186 *Ibid*, at cl 2, s 22(1).

187 *Ibid*, cl 2, s 39(1).

188 *Ibid*, cl 2, s 39(2).

with any requirements under the Act or federal or provincial law, or any other prescribed circumstance (under subordinate regulations),[189] or where the Commissioner determines, at the request of an organization, that identifying an individual is in the interests of that individual.[190] Knowingly contravening this prohibition would amount to an offence and a fine upon indictment of up to $25,000,000 or 5% of the organization's gross global revenues (whichever is greater) or upon summary conviction of up to $20,000,000 or 4% of the organization's gross global revenues (whichever is greater).[191]

AIDA, on the other hand, would apply to (1) artificial intelligence systems and machine learning models made available in the course of international or interprovincial trade and commerce, and (2) the management of the operations of artificial intelligence systems used in the course of international or interprovincial trade and commerce.[192] The proposed Act would predominantly establish an *ex ante* risk-mitigation framework that would require persons to take prescribed measures (as established in future regulations) to prevent biased and harmful AI outputs, primarily for high-impact and general-purpose (e.g., ChatGPT) systems.[193] Interestingly, AIDA would provide the Governor in Council with the authority to make regulations "respecting the data used in the development of, or the making of changes to, an artificial intelligence system."[194] This broad regulation-making power likely provides the Federal Government with sufficient authority to establish regulations concerning the use and management of de-identified

---

189 *Ibid*, cl 2, s 75.

190 *Ibid*, cl 2, s 116.

191 *Ibid*, cl 2, s 128.

192 *Ibid*, cl 39, s 4(a). See also Canada, Minister of Innovation Science and Industry, *Letter from the Honourable François-Philippe Champagne to Mr. Joël Lightbound on Proposed Amendments to Bill C-27, the Digital Charter Implementation Act, 2022* (Ottawa: MISI, 2023) at 7, online (pdf): <ourcommons.ca> [perma.cc/73X4-E6UV] [*Bill C-27 Amendments*].

193 See Scassa, "Regulating AI", *supra* note 146 at 5, 7.

194 See *Bill C-27 Amendments*, *supra* note 192. See also *Bill C-27*, *supra* note 2, Part 3, c 39(c). To note, the current iteration of the Bill explicitly requires persons (who carry out any regulated activity and who process or make available for use anonymized data in the course of that activity) to, in accordance with the regulations, establish measures with respect to (a) the manner in which data is anonymized, and (b) the use or management of anonymized data.

and anonymized data, and the manner in which data is de-identified and anonymized.

However, again, these regulations would only apply in respect of data used to develop or make changes to AI systems to which the proposed Act would apply. Therefore, any future regulations would only apply to data that a person intends to use, or has used, to develop or make changes to an AI system made available for use in the course of international or interprovincial trade and commerce—and not data for AI systems solely intended for intraprovincial trade and commerce (or for research or other non-commercial uses, such as AI developed within a hospital for use in that hospital). In addition, a contravention of the regulations would be a violation rather than an offence, so contravening potential data processing standards would only result in an administrative monetary penalty rather than a very large penalty for an offence. For example, an offence for contravening most of the proposed Act's provisions could result in a fine of up to $10,000,000 or 3% of that entity's gross global revenues (whichever is greater) upon indictment of a non-natural person, or a fine at the discretion of the court in the case of an individual.[195] Upon summary conviction, maximum fines are lower, up to $5,000,000 or 2% of that entity's gross global revenues (whichever is greater) for a non-natural person, and up to $50,000 for an individual.[196] The proposed Act does not establish any amount or range for administrative monetary penalties, which the Act delegates entirely to future regulations.[197]

AIDA would also create two criminal prohibitions, distinct from the rest of the Act and therefore not limited to actions or behaviours in the course of international or interprovincial trade and commerce.[198] First, the proposed Act would prohibit persons from possessing or using personal information (including de-identified information), knowing or believing that the information was obtained or derived, directly or indirectly, from the commission of an offence in Canada under any act of Parliament or a provincial legislature[199], but only when those persons possess or use such

---

195 See *Bill C-27 supra* note 2, cl 39, s 30(3)(a).

196 *Ibid*, cl 39, s 30(3)(b).

197 *Ibid*, cl 39 s 29(1).

198 More specifically, only Part I of the proposed Act would be limited to certain activities in the course of international or interprovincial trade and commerce, see *Bill C-27 Amendments*, *supra* note 192 at 7.

199 See *Bill C-27*, *supra* note 2, cl 39, s 38. To note, the Act would also prohibit

information *for the purpose* of designing, developing, using, or making available for use an AI system. Second, the proposed Act would prohibit persons from—without lawful excuse and knowing that or being reckless as to whether the use of an AI system is likely to cause serious physical or psychological harm to an individual—making available that AI system for use and its use causes such harm.[200] Oddly, for a person to commit this offence, the harm must materialize. The proposed Act does not prohibit persons from making available for use AI systems that they know are likely to cause serious psychological harm. This approach differs from similar legislative proposals in other jurisdictions, such as the proposed European Union AI Act, which prohibits certain systems with unacceptable risks outright.[201] Regardless, AIDA's criminal prohibitions are backed by very strong penalties, upon indictment of up to $25,000,000 or 5% of a non-natural person's gross global revenues (whichever is higher), or up to a fine at the discretion of the court or to a term of imprisonment of up to five years less a day (or to both) for individuals.[202]

Federal law reform would likely result in several follow-on effects for the provinces and potential provincial law reform. In the context of health care AI, AIDA's criminal prohibitions would apply in all provinces and to all persons (organizations and individuals), except, oddly, federal government institutions and products, services, or activities under certain federal government entities' control.[203] In contrast, the proposed CPPA, like PI-

---

such actions based an act or omission anywhere that, if it occurred in Canada, would have constituted such an offence.

200 *Ibid*, cl 39, s 39(a) To note, this prohibition would also apply in respect of substantial damage to an individual's property, and section 38 would also prohibit a person from making available any AI system with the intent to defraud the public and to cause substantial economic loss to an individual, which causes such loss.

201 See EU, *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM/2021/206, 1 at 22–23.

202 See *Bill C-27*, *supra* note 2, cl 39, s 40(a). See also (*ibid*), s 40(b) where, upon summary conviction, a non-natural person is liable of a fine up to $20,000,000 or 4% of gross global revenues (whichever is higher), or up to $100,000 or to a term of imprisonment of up to two years less a day (or to both) for individuals.

203 See *Bill C-27*, *supra* note 2, cl 39, ss 3(1)–3(2).

PEDA, would apply to health-sector organizations in the course of commercial activities, at least in provinces with health-sector legislation that the federal government has not declared substantially similar to PIPEDA (or to the CPPA, once in force).[204] In these provinces, organizations would have to comply with both the proposed CPPA and provincial health-sector legislation. Those health-sector organizations would likely be able to rely on the proposed CPPA's flexibilities concerning exceptions to individuals' consent when they use or disclose de-identified data for socially beneficial health-related purposes, except where those provinces' laws impose more stringent rules, with which those organizations would have to comply. The proposed CPPA's provisions concerning organizations' using or disclosing of de-identified information are *permissive*, so they would not likely "conflict" with more stringent provincial rules, avoiding federal paramountcy. Relatedly, even in provinces with health-sector data protection legislation that the federal government has deemed substantially similar to PIPEDA, the CPPA would establish a new 'floor,' or minimum obligations for commercial organizations. In response, those provinces may have to further modernize their health-sector laws—mainly where the proposed CPPA would render those provincial laws comparatively less protective—to remain substantially similar to the CPPA. In addition, the provinces may also amend their health-sector legislation to introduce similar flexibilities for de-identified data, so that those provinces remain competitive in the health care AI innovation space. These follow-on effects illustrate the importance of Parliament striking an appropriate balance between protecting individuals' and groups' informational privacy, while still enabling strong AI and data-related innovation under the proposed CPPA. AIDA's reach also illustrates the importance of Parliament crafting careful and effective federal criminal laws concerning AI and data, to protect individuals and groups from privacy-related harms.

---

[204] *Ibid*, cl 2, s 122(2)(b).

VII. Governments Must Implement Stronger De-identified and Anonymized Data Governance

Federal and provincial governments should expand ongoing law reform and existing laws to explicitly regulate de-identified and anonymized data. Such law reform should include more comprehensive governance regarding the use and disclosure of such data and, at the provincial level, prohibitions on persons using anonymized health data to re-identify individuals, subject to certain exceptions. At the federal level, Parliament and the Federal Government must be careful not to impede potentially more stringent provincial anonymization and de-identification standards, and Parliament should also expand AIDA's criminal prohibitions to ensure that they are sufficiently broad to protect individuals from both AI and data-related privacy harms. The remainder of this Part elaborates on these policy recommendations, while also noting that their effectiveness will likely depend on well-funded privacy commissioners, strong cybersecurity measures, and high penalties for non-compliance.

The proposed CPPA should not exclude anonymized data entirely,[205] especially if Parliament were to water down the definition of "anonymize" and re-introduce a "reasonably foreseeable in the circumstances" standard based on calls from industry. As explained above, increasingly powerful AI-enabled re-identification tools render true data anonymization illusory. The proposed CPPA should allow the Governor in Council to make regulations governing the anonymization (and de-identification) of personal information. The Federal Government could then establish standards, including by incorporating by reference external standards or standards generated by the Privacy Commissioner, on a dynamic basis for a truly agile approach.[206] The proposed CPPA should also provide the Governor in Council with regulation-making powers respecting organizations' using and disclosing of anonymized and de-identified data. Both of these regulation-making powers would align with Ontario's PHIPA (at least in respect of anonymized

---

[205] This recommendation aligns with Parliament's Standing Committee on Access to Information, Privacy and Ethics' recommendations following the Public Health Agency of Canada's use of anonymized private-sector mobility data during the COVID-19 pandemic, see House of Commons, *Collection and Use of Mobility Data by the Government of Canada and Related Issues: Report of the Standing Committee on Access to Information, Privacy and Ethics* (May 2022) (Chair: Pat Kelly) at 39, online (pdf): <ourcommons.ca> [perma.cc/TQZ5-63B8].

[206] See *Statutory Instruments Act*, RSC 1985, c S-22, s 18.1(1)–18.1(2).

health information, and the use and disclosure of that information).[207] Further, allowing for additional regulation-making authorities to govern the use and disclosure of anonymized and de-identified data under the CPPA would also generally align with AIDA's regulation-making authority "respecting data used in the development of, or the making of changes to, an artificial intelligence system."[208] In many cases, an organization may anonymize or de-identify data, and use or disclose that data, well before they or other parties in the "data supply chain" intend to use it for AI-related purposes; thus, compared to AIDA, regulations under the CPPA concerning anonymized and de-identified data would often apply to data earlier in its "data lifecycle." In addition, because absolute anonymization is a questionable concept, the CPPA's prohibition on re-identifying individuals using de-identified information should also be extended to also include using anonymized data to re-identify individuals.

Based on these recommended regulation-making powers under the proposed CPPA, the Federal Government could then develop contextual rules governing the use and disclosure of de-identified and anonymized data. For example, before an organization discloses de-identified information, regulations should require that organization to conduct a privacy impact assessment,[209] taking into account the sensitivity of the information, the quantity of information, the nature of the intended recipient (and any of the recipients' existing or likely partnerships with third parties), and the intended recipients' cybersecurity measures—while also requiring the disclosing organization to maintain records of such assessments and make those records available to the Privacy Commissioner upon request. These requirements, being contextual, would not necessarily be particularly onerous and impede data access and AI innovation, but they would force disclos-

---

207 See *PHIPA*, *supra* note 3, ss 73(1)(h), 73(1)(o.2). The ability for the Governor in Council to establish anonymization standards would also align with Quebec's private-sector law, see *Quebec Private Sector Personal Information Act, supra* note 2, s 90(3.2), as amended by *An Act to modernize legislative provisions as regards the protection of personal information*, SQ 2021, c 25, ss 158–59.

208 *Bill C-27 Amendments*, *supra* note 193, c 36(c). To note, regulations under the proposed CPPA would apply to a broader range of *private-sector* organizations than AIDA; AIDA would be limited to organizations (and individuals) that conduct certain activities in the course of interprovincial or international trade and commerce.

209 See Cofone, *supra* note 74, ss 3(a), 3(d), 5(c).

ing organizations to consider additional privacy risk mitigation measures, maintain records of those considerations, and possibly implement some of those additional measures. In addition, where organizations use or disclose de-identified data under the proposed CPPA for socially beneficial purposes, regulations should require that a data sharing agreement be in place between the disclosing and receiving entities.[210] The CPPA should also require the disclosing organization to notify the Privacy Commissioner of such disclosure and make that agreement available to the Privacy Commissioner upon request. Further, if the Federal Government were to prescribe any other organizations (in regulations) to which an organization may disclose de-identified data for socially beneficial purposes, especially private-sector or international organizations, then the Government should require more rigorous privacy impact assessments and data sharing agreements in these contexts. To enact these contextual rules, the proposed CPPA should allow regulations to establish different rules for different classes of recipient organizations.

If any province were to establish a category of "de-identified information" in its health-sector legislation—to align with the proposed CPPA's flexibilities—then it should also provide its executive branch with powers to enact regulations similar to those described above, concerning the process of de-identifying data, and governing the use and disclosure of such data in different contexts. Such an approach would support health care AI innovation, especially because de-identified data is often more feature-rich than anonymized data. However, broader use and disclosure of de-identified health information, especially sensitive health information, should attract proportionate oversight in terms of privacy impact assessments and data sharing agreements. For highly sensitive de-identified health information, requirements to notify the respective privacy commissioner could include requirements to submit information to the commissioner or even seek the commissioner's approval.

Concerning anonymized data under the proposed CPPA and provincial health-sector legislation, governments could take lighter-touch approaches to governing the use and disclosure of such data. For example, in Ontario, regulations under PHIPA could require very high standards for anonymization before a person releases anonymized data publicly,[211] and PHIPA could further require persons to keep distribution records of highly

---

210 See Minssen et al, *supra* note 9 at 16.

211 See Ohm, *supra* note 55 at 1729, 1765.

sensitive anonymized data that they confidentially disclose to other persons to create an audit trail.[212] Audit trails can provide important information for individuals and privacy commissioners with respect to such data when those individuals or commissioners commence civil or enforcement actions for privacy breaches, especially unlawful re-identification. While privacy impact assessments and data sharing agreements would likely unduly limit organizations' use and disclosure of anonymized data, in some cases privacy risks may warrant such assessments and agreements, such as when an organization or person sends anonymized data extraterritorially to other provinces or countries with weaker privacy protections. These contextual policy approaches become particularly important if provinces maintain 'reasonably foreseeable' anonymization thresholds, rather than adopting the higher absolute standard under the proposed CPPA, because of increased AI-related re-identification risks.

However, based on the provinces' broader legislative powers over consumer protection and civil rights more generally, the provinces should go beyond federal rules to directly impose requirements on third-party recipients of anonymized and de-identified data within their province.[213] For recipient entities in other jurisdictions, such as other provinces or countries, those provinces would likely still have to rely on indirect approaches, such as requiring data-sharing agreements because of issues concerning the extraterritorial application of those provinces' laws. These extraterritorial concerns raise two related points. First, the provinces, and the federal government, through the CPPA, should consider creating statutory rights for the Crown to enforce data-sharing agreements on behalf of disclosing parties, on a discretionary basis. Second, provincial and federal legislation should require a disclosing party within their jurisdiction to include high minimum penalties for breaching data-sharing agreements—again, in a context-dependent manner based on the nature of the information and the recipient.

All provinces should follow Ontario's lead in amending their health sector legislation to provide their respective governments with regulation-

---

212 *Ibid* at 1756.

213 The Information and Privacy Commissioner of Ontario has noted that it would like to see PHIPA establish rules applicable to recipients of anonymized data, as well as purposes for which personal health information can be anonymized, see Sophie MacRae & Daniel Fabiano, "Towards a More Integrated Health System? Amendments to PHIPA and Announcements about Digital and Virtual Health Care" (13 December 2019), online: <fasken.com> [perma.cc/FMQ8-EUYC]. See also Murdoch, *OPC Report*, *supra* note 105 at 15–20.

making powers to establish anonymization standards and to govern the use and disclosure of anonymized data (and de-identified data, if applicable), and those provinces, as well as Ontario, should also implement such regulations. Further, all provinces should prohibit persons from using anonymized health data (and de-identified data, if applicable) to re-identify individuals, subject to certain exceptions. Health-sector legislation, unlike private-sector legislation, applies to all persons (organizations and individuals) in respect of personal health information. Harmonized re-identification prohibitions across Canada under health-sector legislation would provide important privacy protections for individuals, regardless of the actor and the location of the contravention. Like PHIPA, *wilfully* re-identifying an individual using anonymized data (and de-identified data, if applicable) should also constitute an offence, backed by very high penalties.

In many areas, harmonization is important in order to protect individuals' privacy, reduce regulatory burden, and increase compliance. For example, if the CPPA were to allow the Federal Government to enact anonymization standards, these standards must be consistent with—or at least account for—any standards under AIDA. In other areas, experimentation and diverging approaches can also protect individuals' privacy. For example, one or more jurisdictions may enact more stringent anonymization standards. Here, Parliament and the Federal Government must be careful not to enact anonymization standards that would conflict with or otherwise impede more stringent provincial standards. However, federal regulations could likely include language to account for these circumstances, to prevent federal paramountcy from rendering stronger provincial privacy protections inoperable. This 'conflict' could otherwise occur, for example, where the proposed CPPA and provincial legislation both apply to a private, health-sector organization (i.e., in provinces with health-sector legislation not substantially similar to the proposed CPPA, once in force).

Parliament should also expand AIDA's proposed criminal prohibitions. First, the prohibition on persons possessing or using personal information (including de-identified information[214]), that they know or believe to have been obtained as a result of an offence, should not be limited to *the purpose of designing, developing, using, or making available for use an*

---

214 De-identified information, for the purpose of AIDA, would not include anonymized data as defined in the proposed CPPA (i.e. the absolute standard), but it would capture anonymized data as defined in provincial statutes that falls below the proposed CPPA's absolute standard, i.e. where a risk of re-identification remains.

AI system. This purpose element unnecessarily limits the provision, which should instead be technologically neutral. For example, a person could possess or use de-identified health information that they know was obtained as a result of (1) a cyberattack or (2) a wilful re-identification (without lawful excuse) to develop a non-AI-based software program. Such an action is no less culpable than using that information to design an AI-based system. *Developing an AI system* is not an inherently harmful or morally repugnant action. It follows that this purpose element does not likely insulate the provision from any additional constitutional risk. The threat to morality is a person *knowing the information was obtained as a result of an offence, and using it anyways*. This purpose element instead limits the criminal prohibition to one likely *use* of such information, among many possible uses, and leaves individuals and groups exposed to unnecessary risks of harm. Second, AIDA should also prohibit a person from wilfully manipulating or using an AI-based system to reveal personal health information (including de-identified health information) on which that system was trained or to which it was otherwise exposed. Alternatively, Parliament could amend, for example, section 430 of the *Criminal Code* to clarify that such conduct amounts to computer data-related mischief. [215]

Under AIDA, or related legislation, Parliament could consider prohibiting persons from using or attempting to use anonymized data to identify an individual, either alone or with other information, except in accordance with federal or provincial law; however, this approach is likely unnecessary and has several drawbacks. A similar prohibition under the proposed CPPA would apply to commercial organizations in certain provinces, and it would likely induce the other provinces with legislation substantially similar to PIPEDA to enact similar private-sector prohibitions, so that their laws would also be substantially similar to the proposed CPPA, once in force. Further, while this type of federal prohibition could permit exceptions in accordance with federal or provincial law, any province without a similar prohibition is not likely to have existing statutory exceptions. In this case, a federal prohibition would likely adversely apply to some persons in circumstances where re-identification is morally or clinically justified—i.e., that exceed any potential federal exceptions. In addition, a province may challenge a federal prohibition of this nature as encroaching on its jurisdiction. To mitigate constitutional risks under its criminal law power, Parliament would

---

[215] However, further analysis is required to determine whether existing criminal legislation that prohibits persons from interrupting or interfering with the lawful use of computer data would capture such behaviour, see *Criminal Code*, *supra* note 172, s 430(1.1).

likely have to (1) limit this type of prohibition to sensitive information, such as personal medical information, biometric data, and information about minors, and (2) include a *wilful mens rea* component—thereby limiting the provision's scope. Preferably, as recommended above, all provinces should follow Ontario's health-sector lead in enacting provincial prohibitions that would apply to all persons (including malicious actors), backed by very high fines and/or terms of imprisonment.

Importantly, all of these policy recommendations depend on well-funded privacy commissioners, strong cybersecurity measures, and high penalties for non-compliance. While this paper has discussed several large fines where persons or organizations commit offences or contravene data protection legislation, many statutes still provide for inadequate penalties and other recourse. For example, where a court determines that a person has suffered harm as a result of a contravention or offence under PHIPA, which the defendant engaged in wilfully or recklessly, then the court may include in its award for damages a separate award of up to $10,000 for mental anguish.[216] Further, under Ontario's *Freedom of Information and Protection of Privacy Act,* a person who uses or attempts to use de-identified health information (that was de-identified under a public-sector data integration unit) to identify an individual is guilty of an offence and liable to a fine of up to $5,000.[217] These penalties and separate awards are insufficient to mitigate privacy risks and resulting harm; governments must significantly increase outdated penalties and statutory awards.

---

[216] See *PHIPA*, *supra* note 3, s 65(3).

[217] See *ON FIPPA*, *supra* note 107*,* ss 49.8, 61(1)(b.1), 61(2), as amended by *Protecting What Matters Most Act (Budget Measures)*, *2019*, SO 2019, c 7, Sched 31, ss 6, 8.

VIII. Conclusion

Existing data protection laws and ongoing data protection law reform, by largely excluding anonymized data, may ultimately incentivize persons and organizations to anonymize health care data and to subsequently use and disclose that data, which is a valuable asset for health care AI. Further, ongoing law reform, particularly the proposed CPPA, would also allow organizations to use and disclose de-identified data in several circumstances without requiring individuals' consent, including disclosing de-identified data to hospitals, government institutions, and universities for socially beneficial health-related purposes. These federal law reform efforts, and possible provincial law reform that may follow, would better support health care AI innovation and likely improve Canadian health care and individuals' health outcomes. However, existing laws and law reform efforts do not adequately mitigate health care AI-related privacy risks, especially re-identification risks.

Federal and provincial governments must further modernize data protection laws and law reform efforts to better mitigate these risks and prevent privacy-related harms, while still enabling access to critical health care data for health care AI innovation. Federal and provincial data protection legislation should explicitly regulate anonymized data. Data protection legislation should establish standards for anonymization, whether directly or indirectly, and govern the use and disclosure of such data (and de-identified data, where applicable). Governments can approach strong de-identified and anonymized data oversight in a context-dependent manner that would not unduly limit health care data sharing and health care AI innovation. Most importantly, all provincial health-sector laws should prohibit persons from re-identifying individuals using anonymized data (and de-identified data, where applicable), except under certain circumstances. In addition, while constitutional constraints limit the reach of federal consumer protection legislation, Parliament could likely expand proposed criminal prohibitions in a technologically neutral manner, such as prohibiting the use or possession of personal information that was likely obtained, directly or indirectly, as a result of an offence—whether or not for AI-related purposes. Parliament should also enact or amend criminal legislation to clearly prohibit certain cyberattacks on AI systems, such as model inversion attacks.

Governments around the world are balancing competing interests and, unfortunately, making important data protection policy choices

behind closed doors,[218] which also gives rise to risks of relying too heavily on commercial innovators' input because of their often-greater access to and influence on decision-makers. In Canada, the Federal Government should comparatively engage in a more deliberative process with a broader range of stakeholders to make these important policy choices in a more informed and democratic manner, including by engaging individuals, health care groups (including patient groups), academics, other privacy experts in the not-for-profit sector, and a range of interdisciplinary experts and scholars.[219] Governments' policy choices concerning AI, health care, and data protection over the coming years will have lasting impacts on both privacy rights and AI innovation. We must properly balance competing interests to incentivize AI innovation while ensuring proper privacy protections for individuals and groups. This balancing necessarily includes policies that generate robust health care data and enable access to that data, and they must also encourage the responsible use of that data[220]—including through the explicit regulation of de-identified and anonymized health care data.

---

218 See Bak et al, *supra* note 34 at 4.

219 *Ibid* at 5.

220 See Teresa Scassa, "Data Sharing for Public Good: Does *Bill C-27* Reflect Lessons Learned from Past Public Outcry?" (11 July 2022) at 2, 7, online (blog): <teresascassa.ca> [perma.cc/E4D5-CAPG].