AI AND DIGITAL HEALTH RESEARCH GROUP:

EXECUTIVE SUMMARY

UNIVERSITY OF TORONTO
FACULTY OF LAW

FUTURE OF LAW

Future of Law Lab
University of Toronto Faculty of Law

June 12, 2023

About the Future of Law Lab

The Future of Law Lab is a platform for students, academics, lawyers, and other professionals to participate in collaborative initiatives exploring how the law will evolve in the future. We will dive into the intersection of law, technology, innovation, and entrepreneurship, with programing dedicated to each of these streams. As a hub of interdisciplinary activity, we are dedicated to bringing together individuals from all backgrounds to examine the changing face of the legal profession.

About the AI and Digital Health Research Group

This Executive Summary is prepared by University of Toronto Faculty of Law students Micheal Antifaoff, Nancy Chen, Rao Fu, Sarah Grech, Vera Kaler, Sarrah Lal, Stephanie Ng, Isabella Papalia, Sooyeon Park, Samir Reynolds, Jaqueline Rintjema, Avital Sternin, Ammar Thaver, and Joshua Xu. The Research Group was led by Nancy Chen and Samir Reynolds, upper-year student leaders at the faculty.

The Research Group's objective is to collaborate with the Health Law, Innovation, and Policy Lab ("HLIP"), part of the Lunenfeld-Tanenbaum Research Institute of Sinai Health Systems, to develop an ongoing data governance and security framework for the hospital. The group specifically focused on the risks and challenges associated with the increasing desire for third-party for-profit providers to collect sensitive information for the purposes of creating machine learing and AI tools.

EXECUTIVE SUMMARY

Prepared by the Future of Law Lab's AI & Digital Health Research Group

Working Group Leaders

Nancy Chen and Samir Reynolds

Working Group Contributors

Micheal Antifaoff, Nancy Chen, Rao Fu, Sarah Grech, Vera Kaler, Sarrah Lal, Stephanie Ng, Isabella Papalia, Sooyeon Park, Samir Reynolds, Jaqueline Rintjema, Avital Sternin, Ammar Thaver, Joshua Xu

A special thanks to Dr. Leanne Tran, Director of the HLIP Lab, as well as other members of the Sinai Health System Team, for their guidance and constructive feedback throughout this research process.

TABLE OF CONTENTS

Contents

Data

Deidentification

Under *PHIPA*, deidentification is defined as "means to remove any information that identifies the individual or for which it is reasonably foreseeable in the circumstances that it could be utilized, either alone or with other information, to identify the individual[1]". Deidentification differs from anonymization in that it is possible to link the data back to its original producer if the appropriate connecting information is known. Effective deidentification procedures make the ability to reidentify data nearly impossible.

Currently at Sinai, there are ad-hoc ways of deidentifying data that are determined on a case-by-case basis and require large amounts of time and effort on the part of the research team (including the investigators, information technology services, and the research ethics board). For

[1] *Personal Health Information Protection Act*, 2004, S.O. 2004, c. 3, Sched. A [*PHIPA*] at s 2

example, one method of deidentifying magnetic resonance images involves a two-step internal process. First, a group of physicians works with IT to remove the identifying information and then the principal investigator with re-check all the data. The issue with this process is that it is time consuming and is not functional for large amounts of data.

The Information and Privacy Commissioner of Ontario's recent decision on the use of deidentified data (PHIPA Decision 175). This decision addresses the provincial standard expected by the process of deidentification when data is being sold to third-parties, however, these stringent standards can be incorporated into safe practices surrounding Personal Health data. Specifically, the Commissioner noted the following as being key to safe deidentification practices:

- **De-identification methods:** the use of separate servers and the use of de-identification algorithms developed by industry-recognized privacy experts.
- **Re-identification Risk Analysis:** the third-party is expected to conduct a risk analysis, investigating the probability of re-identification; the third-party must also identify mitigation controls used to prevent re-identification whether there are any motives and/or capacity for the data recipient to re-identify the data.
- **Contractual Safeguards and Confidential Agreements:** the agreement between the hospital and the third party must include privacy and security controls to ensure the data remains deidentified.
- **Audit Rights:** A provision in the data sharing agreement could provide audit rights to the hospital where an independent party could verify its practices, to ensure any third-parties is not using identifiable information unless it is authorized. Sinai may wish to allow for internal audits for their internal research as well.

Key considerations

Below is a detailed list of the key issues we identified affecting data use and deidentification and possible steps that could be taken to address those issues:

1. When proposing a research project using health data and AI it is of the utmost importance to clearly identify who has access to the data. This information affects the kinds of securities and oversights necessary to keep the data safe.
2. When deidentifying data, researchers need to use <u>indirect identifiers</u>. One of the frequent issues identified by the research ethics board (REB) is the lack of understanding of what an indirect identifier is.
    - A possible solution is having the REB pre-approve a list of identifiers that can be validly use in deidentified data and to flag to researchers that when choosing identifiers to keep in the idea the main question to keep in mind is "does the identifier further the research in any way?"
    - REB electronic system coming soon – the new system will include a help box with clarifying definitions of anonymization, pseodonymization, etc as well as suggestions on how to effectively do so. This will allow researchers to easily access Sinai's deidentification standards.

3. Currently, national research is difficult as there are different provincial guidelines on deidentification
    - New Bill C-27 will provide clear federal guidelines.

4. IT is troubled when a project requires patients to sign into a website to consent or do other research activities because the process of logging into a website collects personal information that is automatically stored on a web server.
    - This data must be stored securely in a way that prevents it from being connected to any research data.
5. Research participants and patients must be clearly informed about the deidentification process in order to provide informed consent.
6. Sinai must be transparent with the public about their information practices by describing them clearly and explicitly in their privacy notice.
    - The privacy notice must include information on the purposes of the information, including the purpose of a sale of the de-identified information to a third party.
    - This privacy notice should be easily accessible to the public.

Transfer

Cross-border transfers raise legitimate concerns about where personal data is going as well as what happens to it while in transit and after it arrives at some foreign destination. Unlike the *GDPR* in the EU, *PIPEDA* contains rules prohibiting or restricting cross-border data transfers. Similarly, *PHIPA* does not restrict cross-border transfer of personal information for outsourcing purposes, but it does prohibit the disclosure of personal information to persons outside of Ontario without the affected individuals' consent. The following sections will discuss various factors to be considered when an organization transfers data to a third party within Canada or internationally.

Key considerations

Data transfer is defined by the *GDPR* as an intentional sharing of personal data with a third party, where neither the sender nor the recipient is a data subject. In Canada, under the *PIPEDA*, organizations are held accountable for protecting personal information that is transferred to another jurisdiction, both within Canada and internationally. An important factor to be taken into consideration with data transfer is concerning adequate protection of personal information.

When seeking to release data to third parties within Canada, the following factors are important to consider:

- What type of information will be transferred? What technological routes will be used to transfer the data? How will these routes be private and secured?
- Where in the world is the information being stored? Are there existing policies in place regarding *PIPEDA* and other province-specific laws?
- Will the data be stored in the cloud or on-premise?
    - If in the cloud, what jurisdiction will the cloud be located in? Will it be different from the jurisdiction of the organization that uses it?
- Has the chosen storage provider been subject to an investigation by the Office of the Privacy Commissioner of Canada?
- What types of security and compliance certifications does the storage provider hold?
- What is the plan for the data at the end of the contract term with the provider?

- Who will have access to the data?
- What safeguards are in place to ensure the personal information will be appropriately stored?
- Does the third party have a procedure to follow in the event that privacy breaches do occur? If so, what is it?

Further, the hospital should retain the right to audit and inspect how the personal information is being stored. This may involve a regular audit (i.e. every year or every five years) to ensure the third party is appropriately handling and storing the personal information and also to consistently ensure the safeguards set in place are keeping up with rapidly evolving technology. This may involve either adding an AI component to the research and ethics board or creating a separate AI ethics board.

Guiding principles from the *GDPR*

When seeking to release data to third parties outside of Canada, the following factors may be important to consider: (taken from Art. 45 of the *GDPR*)

- Rule of law & respect for human rights and fundamental freedoms
- Legislation regarding:
  - Public safety
  - Defence,
  - National security
  - Criminal law
  - Access of public authorities to personal data
  - Effective and enforceable data subject rights
  - Effective administration and judicial redress for the data subjects whose personal data are being transferred.
- The implementation of such legislation, data protection rules, professional rules and security measures, effective and enforceable data subject rights and effective administrative and judicial redress for the data subjects whose personal data are being transferred

Additional factors to consider if the party or parties are outside of Canada, based on further research:

- Is there a data management strategy for structuring the limitations on which users can access the data, and what parts of it?
- What standard operating procedure (SOP) software does the party use, and has it historically be successful?
- When and how will the data be used or deleted after the allocated use period has ended?
  - What measures will be taken to ensure sufficient deletion from all servers/software?

Consent

Academic Research

When looking at academic research, there are already relatively robust rules in place in *PHIPA* around approving projects with health (at section 44) and the Research Ethics Board (REB) already has processes in place around approving most kinds of projects. The biggest challenge encountered with AI projects specifically is that these projects usually use retrospective data and researchers are requesting a wavier of consent. The tri-council has criteria for a wavier of consent, but the threshold of impracticality is high and not always clear to the principal investigator. While AI requires large amounts of diverse data to ensure it is representative of the true population, the sheer amount of data required is not always enough to make obtaining consent impractical under the tri-council's definition. It is worth noting that this may be partially cleared up by new electronic systems the REB is implementing soon.

Related to this is the idea of online learning AI models. These models are able to train on and continuously learn from new data. As a result, their output changes over time. These kinds of models raise additional concerns because they cannot be tested once and certified; instead, as new data is given and as the model learns, it needs to be continuously re-tested to ensure that the model's output is still appropriate.

Corporate Partners

During our interviews with stakeholders, a number of broader considerations arose with respect to consent and corporate partners.

One major consideration which continually arose in our discussions was a lack of shared understanding of standards when companies apply for data. Providing companies with more guidance about standards, as well as examples of what would and would not meet these standards, would streamline the application process and save time for both Mount Sinai and its corporate partners.

For example, the more anonymous the data, the more likely projects will be approved because waiving consent and concerns about privacy are less of an issue. However, terms such as anonymized, pseudonymized, and de-identified, which refer to three different standards, are often unclear to companies applying for data. This makes it difficult for the REB to evaluate the privacy implications of providing data. Well-understood standards, coupled with reasons for why data meets the standard, would streamline the REB's work and make it easier for them to understand the privacy implications of each application for waiving consent.

Like with academic research, issues around prospective versus retrospective waivers of consent also affect work with partners. The difference is that partners often have less knowledge about what kind of data are looking for. Companies are sometimes unclear during the application process if they are asking for already collected data, and hence a retroactive waiver of consent, or ongoing data, and hence a prospective waiver of consent. Because the REB evaluates these kinds of waivers differently, providing this information up front could also streamline the process. Also, the same issues of online versus offline learning mentioned above affect industry partners more acutely than academics. Stakeholders at Mount Sinai expressed concern that these kinds of models pose additional ethical challenges for the REB to consider and as a result it is also important for researchers (both academic and industry partner) to identify whether they want to perform online learning or offline learning with the data they are receiving.

Impending Changes

Overview of Bill C-27

Bill C-27 or the *Digital Charter Implementation Act* is currently making its way through the House. and is expected to become law sometime in the summer of this year, 2023. This bill has many effects and is partly to ensure Canada remains in the highest rating under the *GDPR.* The relevance to this project is that this bill would significantly change the Canadian regulatory landscape around data privacy by phasing out *PIPEDA* and introducing new legislation. The *Consumer Privacy Protection Act* (*CPPA*). *CPPA* would replace *PIPEDA* and will apply wherever *PIPEDA* used to apply (for example, commercial transactions not central to hospitals' mandate such as selling patient data).

Bill C-27 as it currently exists would also create several other changes relevant to this project. This briefly looks over some of those, but must be caveated with the knowledge that the Bill is not yet law and the final version could look different than what is discussed here.

Changes Concerning Data

There are two major changes around data and data transfer in the current version of the *CPPA.*

First, *CPPA* requires organizations to implement a privacy management program which details the policies, practices and procedures the organization has put into place to fulfill its obligations under the *CPPA*. Mount Sinai would want to ensure any corporate partners have this program in place.

Second, *CPPA* clarifies data responsibility: organizations are responsible for personal information under their control and remain responsible for the data when it is transferred to a third party. It is the organization's responsibility to ensure that third parties have data protection equal to their own and that complies with *CPPA*. This means Mount Sinai would be responsible for patient data and for ensuring corporate partners have an appropriate data protection framework in place.

Changes Concerning Consent

The *CPPA* requires organizations to identify purposes and obtain consent before data is collected. This means Sinai needs to decide and record why they are collecting data (for example, to provide the data to corporate partners for research) and needs to request consent in plain language anybody could reasonably understand.

However, *CPPA* also introduces several cases where organizations do not necessarily need consent to collect data. Consent is not needed when organizations have a "legitimate interest" in collecting an individual's personal information (S. 18(3)). If the organization wishes to rely on the “legitimate interest” exception, it must record an interest assessment where adverse effects and measures to mitigate them are identified (s. 18(5)). This would mean that despite consent not being required, Sinai still needs to consider and assess the privacy issues and risks that patients could potentially face. It also means that Sinai’s internal impact assessment system must be kept to standards required by statutory provisions.

Another exception is for the use of de-identified data for the organization’s internal research, analysis and development purposes (s. 21). This would mean that while de-identified information can be freely used for Sinai’s internal research projects, consent would be required the moment third parties become involved and are using the data. Limiting the consent exception to internal uses may complicate and thus hinder research partnerships with third parties.

Finally, consent is not needed to share de-identified information if the data is being shared for “socially beneficial” purposes (S. 39(1)), with the legislation explicitly identifying purposes related to health as socially beneficial (S. 39(2)). These exceptions to consent are still being refined by Parliament, but selling patient data could be seen as both a legitimate interest of hospitals and socially beneficial depending on how these are defined in the future and who is buying the data. Because this area is so novel, however, these definitions will almost certainly require deeper dives and may need to be tested in court to provide certainty around what is permissible.

Changes Concerning AI

Automated Decision Systems (ADS) are defined in the *CPPA* as any technology that assists or replaces the judgment of human decision-makers through the use of a rules-based system, regression analysis, predictive analytics, machine learning, deep learning, a neural network or other techniques (s. 2(1)). This is a fairly broad definition, as it includes systems that simply assist the judgment of human decision-makers, which suggests that systems that are not fully automated but nonetheless contains some level of automation fall under the category of ADS. This definition (assuming the one that passes is unchanged from the current version) will likely need to be tested in court to determine what qualifies as an ADS.

When an organization uses an ADS, various requirements need to be met. One such requirement is illustrated in s. 63(3) of the Bill: the organization must, on request by the individual, provide them with an explanation of the prediction, recommendation or decision rendered by the ADS. That explanation must include the type of personal information used to make the decision, the source of the information and the reasons or principal factors that led to such outcome. At first, this transparency requirement raises issues because creating an accurate prediction model requires a large set of data, often involving numerous patients. One might wonder how such detailed explanation can be given to the patient without intruding on the privacy rights of other patients involved in the same study. There are also technical questions around this requirement in that many AI algorithms (especially complicated machine learning ones) are not able to explain their decision-making process and although this is a rapidly developing field within AI research, general explainability has not yet been achieved.

One potential solution is provided under s. 70(1), which states that an organization must not give an individual access to personal information under s. 63 if doing so likely reveals personal information about another individual. Therefore, if the organization can reasonably prove an intrusion on other individuals' privacy rights, it can validly decline the individual's request. However, this is subject to three conditions: (1) the information about the requester can be severed from the information about the rest, (2) the other involved parties consent to the access, or (3) the requester's life, health or security is threatened. In these circumstances, the organization cannot decline the individual's request and must provide him or her with the information.

Quebec's Bill 64

Similar to Bill C-27, Quebec's proposed Bill 64 (which applies to public bodies in Quebec) provides that the individual must be informed of the decision and the reasons for the decision when an ADS is involved (s. 65.2). However, it must be noted that Bill 64, unlike Bill C-27, is concerned only with decisions based exclusively on an automated processing of personal information. It does not include systems that merely assist the judgment of human decision-makers. Since Bill C-27 has not been passed yet, there is still opportunity for an amendment to take place such that it resembles its Quebec counterpart. This would be less burdensome on organizations as the transparency requirement would apply only in the event of fully automated processing of personal information.

This is not directly applicable to Sinai, but the developments around this bill could indirectly affect both interpretations of the *CPPA* and potential new regulation in Ontario.

Future of Law Lab

University of Toronto Faculty of Law

Web: futureoflaw.utoronto.ca
Twitter: @Futureoflawlab