AMS
advancing innovative healthcare
with compassion at its core

UNIVERSITY OF OTTAWA CENTRE FOR
Droit, politique et éthique de la santé
Health Law, Policy and Ethics

MINDFUL OF MACHINES

Mental Health AI, Rights, and the Role for Law

AUGUST 27, 2025

By Sophie Nunnelley in collaboration with Maureen Abbott, Rosario Cartagena, Jennifer Chandler, Lucy Costa, Jake Okechukwu Effoduh, Colleen M. Flood, Vanessa Gruben, Zachary Kaminsky, Marta Maslej, Marina Pavlović, Teresa Scassa, Akanksha Shelat, Ian Stedman, Paul Terwilliger, Monnica Williams, and the workshop participants.

Suggested Citation:
S. Nunnelley et al, *Mindful of Machines: Mental Health AI, Rights, and the Role for Law* (Ottawa: University of Ottawa Centre for Health Law, Policy and Ethics, 2025).

Toronto Metropolitan University
Lincoln Alexander School of Law

uOttawa
Faculté de droit
Faculty of Law

WORKSHOP PARTICIPANTS

Maureen Abbott, Rosario Cartagena, Jennifer Chandler, Alicia Czarnowski, Lucy Costa, Jennie Day, Ruby Dhand, Mary Jane Dykman, Jake Okechukwu Effoduh, Matheus Falcao, Colleen M. Flood, Christina Gilman, Vanessa Gruben, Janet Hendry, Maggie Keresteci, Sara Ling, Marta Maslej, Nancy Myers, Sophie Nunnelley, Glenda O'Hara, Marina Pavlovic, Craig Philips, Jocelyn Rankin, Teressa Scassa, Ian Stedman, Shannon Stroud, Nnubia Ogbuefi, Hyrum Sutton, Marc Lamoureux, Zachary Kaminsky, Paul Terwilliger, Akanksha Shelat, Devin Singh, Christopher Sun, Sian Tsuei, Chijindu Ukagwu.

ACKNOWLEDGEMENTS

This project was generously supported by AMS Healthcare, the University of Ottawa Centre for Health Law, Policy and Ethics, and the University of Ottawa Faculty of Law. Sincere thanks go to Colleen Flood, Zachary Kaminsky, Maggie Keresteci, and Vanessa Gruben for valuable project guidance and feedback and to Giles Holland for workshop support and logistics. Thanks also to Shafiq Qaadri, Emily Naveenan, Shannon Stroud and Raymarck Unera for excellent student assistance.

AMS
advancing innovative healthcare
with compassion at its core

UNIVERSITY OF OTTAWA CENTRE FOR
Droit, politique et
éthique de la santé
Health Law,
Policy and Ethics

Toronto
Metropolitan
University
Lincoln Alexander
School of Law

uOttawa
Faculté de droit
Faculty of Law

TABLE OF CONTENTS

MINDFUL OF MACHINES

MENTAL HEALTH AI, RIGHTS, AND THE ROLE FOR LAW

DO CANADA'S LEGAL FRAMEWORKS ADEQUATELY ENSURE COMPASSIONATE, RIGHTS-COMPLIANT, MENTAL HEALTH AI?

The "Mindful of Machines: Mental Health AI, Rights, and the Role for Law" workshop, held at the University of Ottawa on February 23, 2024, brought together 36 interdisciplinary participants to examine whether Canada's legal frameworks adequately ensure compassionate, rights-compliant mental health AI. The workshop included individuals with lived experience, healthcare professionals, AI researchers and developers, legal scholars, policymakers, and regulators, all focused on a central premise: strong legal frameworks must support service users' dignity, equality, and autonomy. Participants heard expert presentations on key legal issues, engaged in vigorous discussion, and reflected on four questions: (1) What is AI's greatest potential for mental health service users? (2) What are the biggest risks? (3) Is law adequate to support potential and guard against risks? And (4) what messages should we convey to law and policy makers? This report summarizes the workshop presentations, discussions, and findings.

The Foundation: Starting Principles, Promise and Potential Concerns

An essential conceptual premise was that compassionate, client centred mental health care must embed **robust rights protections**—that is, “compassion” requires respect for equality and autonomy. From this conceptual commitment, the group considered mental health AI’s potential and risks. “Mental health AI” was defined broadly, as any AI technology oriented toward mental health objectives—from wellness promotion to diagnosis and treatment—whether delivered directly to service users or via healthcare professionals.

Mental health AI holds transformative potential to address Canada’s severe service gaps. Where one in five people experience mental health problems at any given time, AI tools promise to enhance access, providing culturally appropriate services at scale, supporting personalized prevention and treatment, and reducing provider burnout. Major institutions like the WHO and UK’s NHS have already deployed AI chatbots, while private investment continues to grow. The **myriad use cases for mental health AI** span research, prevention, diagnosis, and treatment applications. For instance, AI tools can monitor social media and wearable device data to detect deteriorating mental health and suicide risk; assist with clinical documentation through AI scribes; predict treatment outcomes and dropout risks; and provide therapeutic support through chatbots.

At the same time, realizing AI’s potential requires careful attention to risk—especially to the patients and mental health service users who are the intended beneficiaries of many mental health AI innovations. The group heard about and discussed the need for authentic **lived experience engagement** that moves beyond tokenistic consultation to genuine partnership. Engagement should address the power dynamics in research and practice and ensure that lived experience voices shape AI development from the outset.

Presentations on the Pre-Market Regulation of Mental Health AI

Pre-market regulation is essential to minimize AI-risk from the outset, for instance, ensuring that only safe and effective mental health AI gets to market. Yet, participants heard that the demise of Canada’s proposed *Artificial Intelligence and Data Act* (AIDA) in January 2025 **leaves Canada without comprehensive AI legislation**. While Health Canada regulates some AI-enabled **medical devices**—playing a crucial oversight role—gaps remain. For instance, due to limits on federal jurisdiction, high-risk AI tools developed and used within a single institution can escape oversight; developers may perceive incentives to characterize their tools as “wellness” tools to avoid regulatory requirements; and the current regulations lack specific bias testing and disclosure requirements. Perhaps most critically, Canada, like governments globally, is grappling with how to address machine learning-enabled devices that change over time. These require a lifecycle-based regulatory approach rather than a single point-in-time approval.

Voluntary tools can fill some of the gaps. For instance, the Mental Health Commission of Canada’s voluntary *Assessment Framework for Mental Health Apps* can be applied to AI-enabled mental health apps. The group agreed such tools play an important role but emphasized the need to pair voluntary standards with binding legal requirements.

Presentations on Mental Health AI and Rights Relating to Autonomy

A cluster of legal issues were loosely grouped under the rubric of “autonomy,” on the premise that the meaningful exercise of agency and choice requires baseline safety assurances, adequate information, and appropriate allocation of responsibility when things go wrong. Yet, there are insufficient protections here (especially in the case of direct-to-consumer apps) and more questions than answers.

The group learned that **consumer protection laws** are ill-suited to address AI; for instance, they typically require a paid transaction of at least $50 to trigger protections, far exceeding the cost of most apps. Without these protections, non-negotiable terms and conditions apply, yet these are rarely read, difficult to understand, and heavily imbalanced in favour of service providers. **Privacy laws** impose obligations on those with care and control of personal health information, requiring transparency and consent regarding the collection, use, and disclosure of the information. Yet, again, the AI context gives rise to critical challenges. For instance, the permanent deidentification of mental health data may be impossible, creating risk; some statutory provisions (such as those allowing the use of personal data for “research” or “analytics”) are ambiguous in their application to health AI; and current frameworks assign significant responsibility to end users to read privacy policies and understand risks. Such risks are especially concerning for vulnerable AI users. For instance, adolescents comprise a significant portion of mental health app users but may have limited capacity to read or understand the complex privacy implications. There is a significant need for improved data literacy among both AI developers and the public.

There are also critical questions about **informed consent** where AI is used in treatment. While treatment requires capable, informed consent, it remains unclear whether or when specific disclosure of AI use in healthcare is required, particularly when AI use might trigger legal consequences, such as reporting duties or involuntary hospital admission.

Other questions relate to **who bears responsibility for AI-related harm**. For instance, if examining a person’s “digital exhaust” allows us to predict suicide risk, at what point is there an obligation use that information? Who should hold such obligations, and what limits should apply? While there are known duties of care in the patient-physician relationship, it is unclear how the availability of AI tools might affect rights and duties in other contexts, such as employment and insurance, where there are competing duties to protect privacy and ensure non-discrimination.

Presentations on Mental Health AI and Equality

The group heard from several expert speakers on a range of **non-discrimination and equality** issues that arise with mental health AI. Importantly, there are significant ongoing problems of racism and other discrimination in *human*-delivered mental healthcare. When these biases aren’t recognized and confronted, they can become embedded in AI as algorithmic bias that produces systematically prejudiced results, compounding existing inequities. There are numerous entry points for bias—including the training data, the developers, the funders, and the methods of deploying AI. There is also a range of possible responses, including technical data-based solutions, supranational prescriptions like UN agency guidelines, and domestic laws and policies.

Going forward there is a need for collaborative discussion about what kind of bias we should address, and how we will address it. Importantly, the goal cannot be perfection; it is impossible to eradicate bias. However, some key components—including mandatory bias testing and disclosure—are essential to the way forward. There may also be *positive* opportunities to use AI to *enhance* equity – for instance through tools that embed culturally competent mental health support and by encouraging developers to ensure equitable distribution.

Discussion: Key Themes and Questions

Several overarching themes emerged from discussions throughout the day, highlighting possible priorities for future investigation and reform:

Genuine Lived Experience Engagement is Critical: There was consensus that patient partners and others with lived experience must be included at the outset of research to help shape questions and outcomes, moving beyond tokenistic involvement. Suggestions included capacity-building among patients, critical examination of underlying theoretical frameworks and funding structures, and exploring alternative data access models to empower persons with lived experience to create their own solutions.

There are Problematic Gaps and Challenges in the Regulation of Medical Devices: Participants were deeply concerned that some high-risk AI tools fall outside Health Canada's medical device regulatory ambit due to constitutional limitations (e.g., making the regulations inapplicable to AI developed and used within a single institution). A significant challenge is distinguishing "medical devices" from "wellness" tools, with mental health app developers potentially downplaying claims to avoid regulation. The absence of prescriptive requirements relating to bias in the medical device regulations was another consistent concern, with calls for Health Canada to prioritize clear and binding rules for data selection, bias testing, and risk reporting. There was also a call for better monitoring and enforcement of requirements for machine learning-enabled medical devices through a "lifecycle" approach, along with more education and support for smaller AI developers and the public, given the troubling gap between public perception and actual oversight.

Privacy Frameworks Do Not Adequately Protect Users of Mental Health AI: Numerous worries arose about data and privacy, with a focus on the many health apps that lack strong privacy practices and policies and may share data with third parties. Participants stressed the inadequacy of relying on terms of service which few read or understand, especially in the case of youth. There is a need for simpler privacy language, broad data literacy efforts, and clarity for AI innovators.

Mental Health AI Raises Distinct Consent Issues: Discussions revealed concerns about AI's use to support legal outcomes like involuntary admission. Some questioned whether specific consent should be required when AI could lead to such consequences, particularly in jurisdictions with fewer patient safeguards. There was also skepticism about AI's capacity for the nuanced, contextual analysis that is essential in treatment decisions and decisions with legal implications such as child protection.

We Need Clearer Allocations of Responsibility for AI-Related Harm: Participants discussed the uncertainty of liability for harm from AI apps, noting that the duties held by regulated health professionals don't necessarily transfer to the developers and companies that create apps and other AI wellness tools. This uncertainty could also create dilemmas for developers who may want to limit their ability to access to sensitive information, such as signs of suicidal ideation, to avoid triggering legal duties.

Bias is Persistent and Must Be Addressed on Multiple Fronts: A recurring theme was the difficulty of accessing sufficient, representative data to assess and mitigate bias, especially given biases embedded in existing datasets like electronic health records. Participants called for a global conversation and collaboration on AI and data regulation and for better guidance on data treatment to avoid exacerbating bias. There is also a need for significant attitudinal shifts—among developers and society more broadly—to prioritize bias detection and mitigation. Market incentives may help engender this shift; however, participants saw legal regulation as essential. They called for law reform to require pre-market bias testing, disclosure, and mitigation, and encourage equitable AI distribution. At the same time, the goal should be safety and harm prevention through transparency and "bias-awareness," rather than the impossible aim of bias eradication.

The Four Questions

Near the end of the workshop, the participants returned to the key questions, distilling their collective thoughts for presentation to the full group. Once again, some key themes emerged.

1. **Greatest potential:** Participants emphasized AI's ability to increase access to mental health services in underserved communities; provide culturally appropriate care at scale without exacerbating provider burnout; help people navigate complex healthcare systems; improve service delivery efficiency; and enable more personalized prevention, prediction, and care.
2. **Biggest risks for service users:** Key identified risks included harm from biased AI and inequitable access or uptake. Concerns were raised about AI supplanting human judgment, confusion when AI and human recommendations conflict, provider over-reliance (automation bias), and the potential for AI to undermine culturally sensitive human service models. Improper AI and data use (accidental or intentional misuse, data hacks, use for insurance or employment denial) and the erosion of public trust due to misuse were also significant concerns. Finally, regulatory imbalance was identified as a risk, either from overly burdensome regulations stifling innovation or from too-rapid AI development lacking adequate oversight, particularly for "wellness" tools.

3. **The Inadequacy of Current Law:** All agreed that current law and policy have not kept pace with technological innovation and are inadequate. Participants were especially concerned about gaps and inadequacies in privacy law, medical device regulation, and the lack of mandatory bias testing and mitigation. A renewed approach to both pre- and post- market regulation was deemed essential.
4. **Messages for Law and Policy Makers:** Key messages included the need for interdisciplinary collaboration, including diverse perspectives and lived experience representation, in AI research, development, and policymaking. There was a strong call for Canadian government leadership on AI-enabled medical device regulation to ensure effective requirements relating to bias, as well as a lifecycle approach to machine learning enabled medical devices. Participants also called for strengthened privacy laws that specifically address the challenges raised by AI. For instance, rather than rely on companies' privacy policies to define rights, Canada could follow the example of some jurisdictions that are moving beyond contractual instruments to legislate clear rights and duties in relation to AI and privacy.

Conclusion

This report and the underlying workshop highlight the transformative potential of mental health AI to address severe service gaps in Canada. Yet, they also underscore significant legal and ethical challenges that current frameworks are ill-equipped to address. Key regulatory gaps include the demise of Canada's first attempt at dedicated AI legislation and limitations in medical device regulations that leave some high-risk tools unregulated or underregulated. Existing general-purpose laws, such as privacy, consumer protection, and informed consent laws, are often a poor fit for health-AI, given problems like data reidentification, unread contractual terms, and unclear duties of care where AI use causes harm.

The persistent and complex issue of algorithmic bias also emerged as a central concern. There is a need for legally mandated pre-market testing, transparent disclosure of limitations, and ongoing monitoring. More fundamentally, a cultural shift is needed so that bias minimization and equity are perceived as integral to "quality" AI, moving towards "bias-awareness" and harm prevention rather than the impossible aim of bias eradication.

Canada has an opportunity to lead by pairing technological advancement with robust rights protections, ensuring that mental health AI enhances service user dignity, autonomy, and equality. This path forward will require continued, inclusive, and interdisciplinary collaboration to translate discussions into concrete legal and policy reforms.

I. INTRODUCTION

1. Mental health AI and the goal of compassionate, rights-compliant care

At their best, artificial intelligence-based mental health tools ("mental health-AI") offer significant potential. They promise to bridge critical gaps in mental health service provision across the spectrum of care, enhancing access and prevention, reducing stigma, supporting effective treatment, and preventing relapse. Governments and institutions around the world have seized on this promise, positioning AI as a means of alleviating shortages in mental health services. For instance, the World Health Organization (WHO) released a new AI chatbot, "S.A.R.A.H." ("Smart AI Resource Assistance for Health") that offers "nuanced, empathetic responses" to users seeking guidance about mental health and other health topics, and the U.K. National Health Service now offers the Wysa chatbot in its "talking therapies" pathway.[1] Mental health AI is also big business with growing roles for private finance.[2]

Yet, amidst this surge in development and adoption there are significant risks. Some of these mirror those found in other health-AI uses. For instance, many are asking difficult questions about AI efficacy and safety; the right to informed consent; and embedded bias and discrimination, among other issues.[3] However, the mental health context—which includes intersecting vulnerabilities, a lack of alternatives, and a new proliferation of relatively unregulated direct-to-consumer mental health apps and tools—also gives rise to distinct concerns.

This report and underlying workshop delve into these risks and complexities with a focus on the rights and legal protections for mental health service users. All stakeholders have valid concerns in this environment. For instance, clinicians may question whether AI will lessen administrative burdens or instead impose new obligations to learn AI systems, and health providers and institutions rightly question who will be liable when AI use results in harm. However, this workshop was primarily concerned with service user rights and the extent to which our legal frameworks promote mental health AI that supports, rather than undermines autonomy and equality. Put another way, we can center an idea of compassionate healthcare delivery, understanding that for care to be "compassionate," it must respect and promote service users' dignity, equality, and autonomy.[4]

---

[1] The SARAH chatbot was released in April 2024. See World Health Organization, "WHO unveils a digital health promoter harnessing generative AI for public health", online: <https://www.who.int/news/item/02-04-2024-who-unveils-a-digital-health-promoter-harnessing-generative-ai-for-public-health>; See also Wsya's adoption in UK's National Health Service since March 2024 at <https://www.wysa.com/nhs>.

[2] Tori DeAngelis, "As funding cools, venture capitalists shift investments in mental health", (1 January 2023) online: <https://www.apa.org/monitor/2023/01/trends-venture-capital-funding-shifts>.

[3] See: Sophie Nunnelley et al, "Cracking the code: a scoping review to unite disciplines in tackling legal issues in health artificial intelligence" (2025) 32:1 BMJ Health Care Information e101112 [Nunnelley].

[4] Brian Hodges, Gail Paech, & Jocelyn Bennett, *Without Compassion, There Is No Healthcare: Leading with Care in a Technological Age* (Kingston: McGill-Queen's University Press, 2020).

2. What is “mental health AI”?

Broadly defined, AI is “the use of digital technology to create systems capable of performing tasks commonly thought to require human intelligence.”[5] Canada’s draft *Artificial Intelligence and Data Act* (Bill C-27), defines “artificial intelligence system” as

> a technological system that, autonomously or partly autonomously, processes data related to human activities through the use of a genetic algorithm, a neural network, machine learning or another technique in order to generate content or make decisions, recommendations or predictions.[6]

The subcategories of AI include “machine learning” (ML), which allows systems to learn from experience without being specifically programmed.[7] “Generative AI” is a subset of ML that allows for content creation; large language models (LLMs) such as Open AI’s ChatGPT are perhaps the best known of these ML applications, but they are growing in number.[8] These LLMs learn patterns and relationships from vast training datasets and can autonomously and coherently respond to prompts.[9]

“Mental health,” in turn, is defined by the WHO as “a state of well-being in which the individual realises his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to contribute to his or her community.”[10] “Mental health” and “mental health problems” are not opposing states; “[s] omeone with a diagnosed mental illness can experience good mental health, while someone without a diagnosed mental illness can experience difficulties.”[11]

Building on these definitions, this Report adopts a broad definition of “mental health AI,” encompassing applications across the mental health spectrum. It includes AI use in a range of mental health support and care—from wellness promotion to diagnostic and treatment applications—and recognizes varying delivery models, including direct-to-consumer and health professional mediated use.

---

[5] National Health Service, “Artificial Intelligence” (30 April 2025), online: <https://transform.england.nhs.uk/information-governance/guidance/artificial-intelligence/>.

[6] 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, (first reading 16 June 2022), p 3, cl 2 [*Artificial Intelligence and Data Act*]. See discussion of *AIDA* below, in Part III(1).

[7] Stuart J Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed, (Hoboken: Prentice Hall, 2020); Sara Brown, “Machine learning, explained | MIT Sloan” (21 April 2021), online: <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>.

[8] OpenAI, *ChatGPT*, online: https://chat.openai.com; Google DeepMind, *Gemini*, online: https://gemini.google.com; Anthropic, *Claude*, online: https://claude.ai.

[9] David Nield, “How ChatGPT and Other LLMs Work—and Where They Could Go Next | WIRED”, online: <https://www.wired.com/story/how-chatgpt-works-large-language-model/?>.

[10] World Health Organization, “Mental health”, online: <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>

[11] Centre for Addiction and Mental Health, “Mental Health 101”, online: <https://moodle8.camhx.ca/moodle/mod/book/view.php?id=48&chapterid=5>; World Health Organization, “Mental health”, online: <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>; Mental Health Commission of Canada, *Toward Recovery & Well-Being: A Framework for a Mental-Health Strategy for Canada* (Ottawa: Mental Health Commission of Canada, 2009).

3. The workshop

On 23 February 2024 a workshop was hosted by the University of Ottawa Centre for Health Law, Policy and Ethics, to unpack these issues within an interdisciplinary group.[12] Thirty-six participants assembled from a range of sectors and experiences, including: people with lived experience; healthcare professionals; health researchers; AI researchers and innovators; legal scholars; policymakers; and regulators. The presentations were divided into four groups. Group one provided key framing for the day, explaining: (i) the workshop's theoretical premise and purpose; (ii) the range of possible AI uses in mental health contexts; and (iii) some critical considerations and concerns from a mental health service user perspective.

Group two discussed the pre-market laws, regulations, and guidelines that apply to mental health AI; they addressed the laws that shape AI before it is put in the hands of providers, service users, or patients. Group three addressed legal issues that roughly correspond with "autonomy" interests—namely, consumer protection laws, liability considerations, privacy, and legal capacity and informed consent. Speakers in group four unpacked some equality considerations, discussing a range of challenges related to equity and non-discrimination in AI.

Participants engaged in vigorous discussion, moderated by Professor Vanessa Gruben (University of Ottawa), throughout the day. In doing so, they were asked to reflect on four questions oriented to ensuring compassionate, rights-promoting mental healthcare:

1. What is mental health AI's greatest potential for mental health service users?
2. What are the biggest risks for service users?
3. Does the law adequately support the potential while guarding against the risks?
4. What message might you give policymakers about how law could better support compassionate and rights-respecting mental health AI?

These prompts served as a framework for interdisciplinary reflection and brainstorming. By sharing diverse perspectives and experiences, the goal was to deepen collective understanding and generate insights for potential law and policy reform.

---

[12] University of Ottawa Centre for Health Law, Policy and Ethics, online: <https://www.ottawahealthlaw.ca/>.

II. MENTAL HEALTH AI: PRINCIPLES, PROMISE, AND POTENTIAL CONCERNS

The first set of speakers established the day's foundation. They explained the workshop's purpose and rights-focused approach, introduced the current state and range of possible mental health AI applications, and set out some key issues and concerns from service users' perspectives.

1. The premise: Compassion and rights as two sides of the same coin (Sophie Nunnelley, Toronto Metropolitan University)

Sophie Nunnelley, Assistant Professor at Toronto Metropolitan University's Lincoln Alexander School of Law, began the day's discussions by outlining the importance of discussing mental health-AI and establishing the conceptual orientation of the workshop.[13]

Canada's need for mental health services is significant. One in five people in Canada are experiencing a mental health problem or illness at any given time and by age 40, half of Canadians have or have had a mental health problem or illness.[14] Moreover, the magnitude of the problem appears to be increasing. A study reported significant increases over the last decade in the percentage of Canadians who met the diagnostic criteria for major depressive episode, bipolar disorder, and generalized anxiety disorder.[15] The COVID-19 pandemic only exacerbated the problem. In one 2022 survey, 35% of respondents reported moderate to severe mental health concerns.[16] Researchers have predicted a significant increase in the number of people with mental health problems and concerns in Canada by 2041.[17]

Alongside this rising need are significant barriers to access. A 2023 study found one in three people who met the diagnostic criteria for a mood, anxiety, or substance use disorder in the previous 12 months had unmet or only partially met needs for mental healthcare services, and that unmet needs for counseling and psychotherapy were especially high.[18] While there are many reasons for this inaccessibility (including stigma, discrimination, and a lack of providers) the issue of cost bears emphasis; provincial and territorial public medicare programs generally do not cover counselling and psychotherapy, making these services inaccessible for many Canadians.[19]

---

[13] At the time of the workshop, Sophie Nunnelley was an AMS Fellow in Compassion and Artificial Intelligence and Associate Director of the University of Ottawa Centre for Health Law, Policy and Ethics.

[14] Paul Smetanin et al, *The Life and Economic Impact of Major Mental Illnesses in Canada: 2011 to 2041*. (North York: RiskAnalytica, on behalf of the Mental Health Commission of Canada, 2011).

[15] Ellen Stephenson, "Mental Disorders and Access to Mental Health Care" (22 September 2023), online: <https://www150.statcan.gc.ca/n1/pub/75-006-x/2023001/article/00011-eng.htm>.

[16] Canadian Centre on Substance Use and Addiction & Mental Health Commission of Canada, *Mental Health And Substance use During COVID-19 Final Summary Report: Regional Spotlight and Characteristics* (Ottawa: Canadian Centre on Substance Use and Addiction, 2022).

[17] Mental Health Commission of Canda, *Making the Case for Investing in Mental Health in Canada* (Ottawa: Mental Health Commission of Canada, 2013).

[18] Ellen Stephenson, *Mental Disorders and Access to Mental Health Care*, Catalogue no. 75-006-X ISSN 2291-0840 (Ottawa: Statistics Canada, 2023), online: <https://www150.statcan.gc.ca/n1/pub/75-006-x/2023001/article/00011-eng.htm>.

[19] Canadian Agency for Drugs and Technology in Health, *Artificial Intelligence and Machine Learning in Mental Health Services: A Literature Review*. (Ottawa: Mental Health Commission of Canada, 2021).

We should be considering AI's uses and limits against this well-established backdrop of unmet need, Nunnelley emphasized. On the one hand, AI could assist with "[c]losing the accessibility gap" in mental health services.[20] Some policymakers and stakeholders, including the Mental Health Commission of Canada and the WHO, have spoken optimistically about the AI's gap-filling potential.[21] Yet, the degree of need could also exacerbate risk, especially for service users. Indeed, while all parties may have concerns about AI in mental health care (health professionals and facilities might worry about liability or increased workload and insurers about rising costs) it is ultimately service users who will bear the consequences if they are forced to rely on AI that is unsafe, discriminatory, or lacking in privacy safeguards.

The intention behind the workshop was to bring these potential risks to the foreground and to consider whether our current rights protections are sufficient. This inquiry is of course necessary to prevent undue harm. However, Nunnelley also related this objective to the concept of *compassionate care*.[22] While the language of "compassion" can be distorted to justify coercion and the *absence* of legal rights, we should reject this framing to instead understand "compassionate" care to be necessarily client-centred and committed to respect for dignity, autonomy, and equality[23]. Legal rights—including rights to informed consent, non-discrimination, and privacy—are a means of safeguarding these values of dignity, autonomy, and equality that are essential to compassionate care.

Nunnelley closed by emphasizing the importance of inclusive discussion. This dialogue is often lacking; scholars have observed that "most machine learning solutions in mental health are being developed in silos, without the inclusion of decision makers, experts, and end users."[24] We must prioritize the inclusion of interdisciplinary and lived-experience voices in both technological and regulatory innovation.

---

[20] Johanna Habicht et al, "Closing the accessibility gap to mental health treatment with a personalized self-referral chatbot" (2024) 30:2 Nat Med 595.; See also David B Olawade et al, "Enhancing mental health with Artificial Intelligence: Current trends and future prospects" (2024) 3 J Medicine, Surgery, & Pub Health 100099; Vikram Patel et al, "The Lancet Commission on global mental health and sustainable development" (2018) 392:10157 The Lancet 1553.

[21] Mental Health Commission of Canada, *Artificial Intelligence in Mental Health Services: An Environmental Scan* (Ottawa: Mental Health Commission of Canada, 2021) at 2; World Health Organization, "Mental health", online: <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>.

[22] For discussion of the importance of compassion in healthcare, particularly in the context of rapid technological change, see: Brian Hodges, Gail Paech and Jocelyn Bennett, eds, *Without Compassion, There is no Healthcare: Leading with Care in a Technological Age* (Montreal: McGill-Queen's University Press, 2020).

[23] This erroneous equation of "compassion" with coercive measures is especially prevalent in mental health and substance use contexts. See Alberta's *Compassionate Intervention Act*, SA 2025, c C-21.5. In this equation, rights and compassion are seen to be in conflict – e.g., the language of leaving people to "rot with their rights on". This approach in this workshop and report reject this equation; on the definition of compassion as client centered, see e.g., Sima Babaei & Fariba Taleghani, "Compassionate Care Challenges and Barriers in Clinical Nurses: A Qualitative Study", Iran J Nursing Midwifery 24:3 213.

[24] Nikolaos Koutsouleris et al, "From promise to practice: towards the realisation of AI-informed mental health care" (2022) 4:11 Lancet Digital Health e829; See also Gooding and Kariotis noting, in a review of the empirical literature on mental health AI, that "there was a "near-complete lack of involvement of mental health service users" in Piers Gooding & Timothy Kariotis, "Ethics and Law in Research on Algorithmic and Data-Driven Technology in Mental Health Care: Scoping Review" (2021) 8:6 J Medical Internet Research Mental Health e24668, online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8262551/; Piers Gooding & Timothy Kariotis, "Ethics and Law in Research on Algorithmic and Data-Driven Technology in Mental Health Care: Scoping Review" (2021) J Medical Internet Research DOI: 10.2196/24668. See also Nunnelley, *supra* note 3.

2. Introduction to AI and its potential for mental health care (Zachary Kaminsky, The Royal)

Dr. Zachary Kaminsky, DIFD-Mach-Gaensslen Chair in Suicide Prevention Research at The Royal Ottawa's Institute of Mental Health Research, began the discussion of AI's promise for mental health. He set out a range of AI use-cases in mental health and discussed some of his lab's research.

Kaminsky emphasized the power of ML to learn from vast datasets and identify patterns that humans can't see. In this way it can help identify and scale solutions to a variety of mental health contexts. AI use cases include supporting research (e.g., facilitating literature reviews); improving health system efficiency (e.g., matching people with mental health services); and supporting workflow (e.g., appointment scheduling and administrative tasks).[25] Yet, he suggested AI is most exciting in its potential to help fill gaps in mental health care. AI can support prevention, diagnosis, and treatment, through a variety of clinician-mediated or direct-to-consumer tools.[26]

With respect to *prevention*, AI can assist with monitoring that helps keep people out of hospital. For example, Kaminsky's research team has developed a natural language processing algorithm that can examine a person's social media output to detect rising levels of distress and suicide risk.[27] Other tools collect data from wearable devices to detect deteriorating mental health.[28] These preventative tools may directly offer lifestyle adjustments and resources or complement professional mental health services. An example of the latter is Kids Help Phone's chabot, "Kip". Children and youth speak to Kip directly but, when necessary, it will link them with clinicians and other human supports.

Diagnostic AI tools tend to be used by health professionals given the danger of diagnosis with no provider "in the loop". Kaminsky offered the example of Kintsugi Voice, a tool for mental health professionals that uses voice biomarkers to identify signs of depression and anxiety.[29] Another example, a tool offered by Highmark Interactive, remotely monitors patients' brain function as a means of supporting clinical diagnosis and treatment.[30] AI-delivered *treatment* can be provided with or without clinician involvement. Chatbots like Woebot and Wysa appear to provide cognitive behavioural therapy (CBT) and dialectical behavior therapy (DBT), though the company websites describe their offerings in terms of "self-help" rather than "therapy" or

---

[25] As an example of a tool to match people to services, Kaminsky offered the example of Access MHA. See: <https://www.accessmha.ca/>. As an example on the administrative front, AI "scribes" transcribe clinical encounters, facilitating and improving clinician note taking.

[26] He noted that roughly 30,000 to 40,000 apps are already on the market. There are an estimated 10,000 mental healthcare related apps on various markets: John Torous, Keris Myrick & Adrian Aguilera, "The need for a new generation of digital mental health tools to support more accessible, effective and equitable care" (2023) 22:1 World Psychiatry 1.

[27] Zachary Kaminsky et al, "Machine Learning-Based Suicide Risk Prediction Model for Suicidal Trajectory on Social Media Following Suicidal Mentions: Independent Algorithm Validation" (2024) 26 J Medical Internet Research e49927.

[28] Nuno Gomes et al, "A Survey on Wearable Sensors for Mental Health Monitoring" (2023) 23:3 Sensors (Basel) 1330.

[29] Kintsugi Health, "Kintsugi Voice" (2023), online: <https://www.kintsugihealth.com/solutions/kintsugivoice>.

[30] Highmark Technologies, "Highmark Technologies" (2023), online: <https://highmark.tech/>.

"treatment."[31] Another example of AI-supported treatment is being developed in Kaminsky's lab; they are using social media data to create client "digital twins", to which they can ask questions (e.g., "what causes you stress?"), allowing crisis responders and therapists to provide more effective care. Another treatment-oriented AI tool will use data from all psychiatric admissions in Ontario since 2008 to create a model that predicts post-discharge suicide attempts. The model can be used alongside a client digital twin to determine which modifiable treatment variables (e.g., more interaction with family and friends or fixing sleep problems) would drive readmission further out in time.

Kaminsky closed by emphasizing that realizing mental health-AI's promise will require guardrails, for instance to protect privacy, given the significant volumes of personal information that are required to achieve AI's most promising insights. Indeed, the risk may increase as AI becomes increasingly complex and opaque. He called for thoughtful safeguards and ongoing monitoring to ensure an appropriate risk-benefit ratio.

3. Mental health AI at Toronto's Centre for Addiction and Mental Health (Marta Maslej, CAMH)

Continuing the theme of AI's promise, Dr. Marta Maslej, Staff Scientist with The Krembil Centre for Neuroinformatics at the Centre for Addiction and Mental Health (CAMH), discussed several AI applications under study at CAMH. These included tools to support clinician documentation, prediction and risk assessment, and patient care.

Maslej noted that AI scribes can help reduce the time required for clinical documentation. For instance, CAMH has implemented one type of AI scribe (which uses AI-enabled speech recognition) to better integrate EHRs into physician workflows.[32] A related tool, still in development, could also use AI for summarization and retrieval. Where clinicians currently sift through large amounts of narrative data (e.g., descriptions of symptoms and behaviours) to find relevant details, the team is researching the potential for generative AI to answer clinical questions about the record. Clinicians could ask a range of questions, such as, "did the patient eat that day?" or "are there indications of violence?". The AI model would then search the record for responsive information. Alternatively, AI could summarize key details from narrative data to help clinicians review information or generate new documentation (e.g., discharge summaries).

---

[31] Woebot has been widely discussed in the press as a direct-to-consumer tool. See Barclay Bram, "Opinion | My Year With Woebot, an A.I. Chatbot Therapist", *The New York Times* (27 October 2022) online: <https://www.nytimes.com/2022/09/27/opinion/chatbot-therapy-mental-health.html>.>. The Woebot website reportedly once said the app could "automate both the process and content of therapy" [emphasis added], stopping short of claims it was itself providing therapy. It also reportedly referred to the tool as "self-help" (this language has since been removed from the website). See Karen, "Something Bothering You? Tell It to Woebot.", *The New York Times* (1 June 2021), online: <https://www.nytimes.com/2021/06/01/health/artificial-intelligence-therapy-woebot.html>.
At the time of writing, the Woebot website states: "Woebot is only available to new users in the United States who are part of a study or who have an access code from their provider, employer or other Woebot Health partner. We have found that people have the best experience when Woebot is delivered under the supervision of a healthcare provider, so we are partnering with health plans and health systems to make Woebot available to the people they serve." (Woebot Health, "Frequently Asked Questions" online: <https://woebothealth.com/FAQ/>, last accessed July 8, 2024). The Wysa website also rejects the language of "treatment". It says, Wysa "is not intended for providing diagnosis, treatment, or cure of a condition or disease or disorders" but rather provides "evidence-based tools and techniques to manage emotions and encourage mental wellbeing in a self-help context." (Wysa, "Frequently Asked Questions" online: <https://www.wysa.com/faq>).

[32] Project led by Dr. Gillian Strudwick (senior scientist at the CAMH Krembil Centre for Neuroinformatics).

Maslej also discussed AI's potential to support prediction and risk assessment, given its power to detect patterns in large quantities of data. Various types of risk could be assessed, including risk of suicide or self-harm, violence or aggression, or risk of negative treatment outcomes with a given treatment. For example, one CAMH team is developing a tool that predicts the risk of treatment drop out; where approximately 20% of patients drop out of the care pathway, determining who is at risk of dropping out would allow those people to receive early intervention (such as psychoeducation, negotiated treatment options, or reductions in the barriers to care).[33] Other AI tools provide care-delivery support. Maslej discussed a chatbot being developed at the University of Toronto that is intended to support smoking cessation.[34] The tool delivers motivational interviewing, asking people a series of questions and responding with reflections to gently and compassionately move them towards a decision to stop smoking.

Maslej emphasized the importance of consulting with the end-users of AI systems. For instance, a study of frontline nurses' perceptions of AI-supported prediction tools found widespread concern about the potential for conflict between AI-prediction and clinical intuition.[35] Participants worried they will face institutional or practical pressures to disregard their experience and follow the AI. Maslej emphasized the need to address such concerns—and carry out extensive evaluations of both benefits and harms—if AI tools are to be effectively employed in mental healthcare.

4. The importance of service-user perspectives (Lucy Costa, Empowerment Council)

Lucy Costa is Deputy Executive Director of the Empowerment Council, an organization that advocates for service user rights at Toronto's CAMH and in the courts. Costa spoke about the need to both foster, and think critically about, lived experience engagement as we develop parameters and approaches to mental health AI.

Costa emphasized the importance of fostering dialogue between mental health service users and those working in law and policy. She highlighted the strong history of social movement activism—especially Toronto in the 1980s and 90s—which helped forge partnerships between patients and advocates. These collaborations supported significant successes, including the recognition of persons with psychiatric disabilities under the Canadian *Charter of Rights and Freedoms* and the creation of Ontario's Psychiatric Patient Advocate Office in 1982. Costa emphasized that similar partnerships will be crucial for establishing the appropriate boundaries for AI in mental health services, particularly around issues such as consent. Indeed, Costa argued every "revolution" in mental health policy rests on a set of values and assumptions that must be carefully examined. CAMH itself is an example: once a walled asylum, it has transformed into a community-oriented therapeutic environment with transparent glass walls replacing the old barriers. Yet, patient perspectives reveal that new forms surveillance persist. Costa emphasized that with AI poised to drive the next

---

[33] This team is led by Masooma Hassan (Project Manager, Health Intelligence, CAMH).

[34] Andrew Brown et al, "A Motivational Interviewing Chatbot with Generative Reflections for Increasing Readiness to Quit Smoking: Iterative Development Study" (2023) 10 JMIR Ment Health e49132.

[35] Ryan Chan, Richar Booth & Gillian Strudwick, "Nurses' Perspectives on Machine Learning-Decision Support Systems and Clinical Decision Making" (in Preparation).

revolution, it will be crucial to consider these different perspectives and experiences.

Costa also challenged the group to think about *how* we approach service user engagement. She argued engagement has problematically shifted in focus from collective to individual voices, driven in part by Ontario's 2010 *Excellent Care for All Act*, which requires that hospitals consult with patients.[36] While this legislation has been positive in some respects, individual patient representatives may speak without connecting to broader community concerns. To avoid being "seduced" by individual narratives, Costa suggested we could incorporate robust conceptual frameworks such as those rooted in critical race theory and queer theory.

Finally, Costa emphasized the need to attend to the power dynamics in engagement and research. She raised questions about how different forms of expertise are valued and employed; whether AI will only be used to study patients or also to examine the behaviours of clinicians and researchers; and how we use service user labour. She noted that service users are often tasked with lower level "emotional support" roles in research instead of management positions with decision-making power. Costa also highlighted the power dynamics embedded in research funding. She observed that funding mechanisms often allow institutions to engage with patients in ways that are "comfortable," avoiding challenging conversations, particularly when donor politics are involved. Modes of engagement must be scrutinized to prevent a tokenistic approach.

Throughout her remarks, Costa emphasized the importance of critical thinking about how and why we engage service users. She called on us to seize the current opportunity to shift the modes of engagement to help ensure that mental health AI truly benefits and respects the rights of those it aims to serve.

---

[36] *Excellent Care for All Act, 2010*, SO 2010, c 14.

III. MENTAL HEALTH AI AND LAW: PRE-MARKET REGULATION

A second group of speakers addressed to the state of, and potential for, pre-market regulation of mental health-AI. These frameworks—including attempts at federal artificial intelligence legislation, medical device regulation, and a voluntary mental health app assessment tool—are all oriented to ensuring safety and preventing harm, rather than to providing legal redress for harm after the fact.

1. Direct AI regulation and the former *AI and Data Act* (Teresa Scassa, University of Ottawa)

Dr. Teresa Scassa, Professor and Canada Research Chair in Information Law and Policy at the University of Ottawa, discussed Canada's attempts at dedicated AI legislation and the now defunct *Artificial Intelligence and Data Act* (AIDA).[37]

AIDA was one component of Bill C-27 (*Digital Charter Implementation Act*), introduced in June 2022.[38] The bill contained three proposed laws: the *Consumer Privacy Protection Act*, the *Personal Information and Data Protection Tribunal Act*, and AIDA. The Standing Committee on Industry and Technology was conducting a clause-by-clause consideration of the bill but with the January 2025 prorogation of Parliament, the bill died on the order paper. Even had it been passed, Scassa noted, implementation would have required at least two years to develop necessary regulations. The regulations were crucial to providing flexibility to address rapidly evolving technology.

Scassa explained that AIDA employed a risk regulation approach aimed at preventing harms before they occur. It established obligations for various actors in the AI value chain to ensure market-ready AI systems are safe, properly designed, and relatively free from bias. These obligations applied to a range of actors, including those who first deploy machine learning or general-purpose AI systems; manage these systems; or integrate them into high-impact applications. Key obligations related to transparency (e.g., requirements to disclose AI use); privacy and data governance; risk anticipation and mitigation (e.g., through bias assessments); compliance documentation (e.g., audits); and monitoring and oversight. Many specifics, however, were left to be clarified in regulations.

The jurisdiction to pass the federal AI bill was found in the federal constitutional power over interprovincial and international commerce as well as the criminal law power.[39] Consequently, such a law would principally apply to AI systems provided commercially across domestic or international borders. The draft law would not have covered, for instance, in-house AI systems within a single hospital; systems designed and used within a single province /territory; or non-commercial systems created by nonprofit organizations. Nevertheless, Scassa noted developers of AI systems with the potential

---

[37] *Artificial Intelligence and Data Act, supra* note 6. Professor Scassa's original presentation was delivered before AIDA died on the order paper. This summary has been updated to reflect more recent developments.

[38] Canada, Department of Justice, "Bill C-27, An Act to enact the Consumer Privacy Protection Act" (2022), online: Department of Justice <https://www.justice.gc.ca/eng/csj-sjc/pl/charter-charte/c27_1.html>.

[39] *Constitution Act, 1867* (UK), 30 & 31 Vict, c 6, s 91, ss 2, reprinted in RSC 1985, Appendix II, No 5.

for a broader market should heed any federal AI requirements so as to enable expansion outside of a single province. It is also possible that provinces will eventually enact their own AI laws—something that Ontario has already done for the broader public service.[40]

AIDA only regulated "high impact" systems, which proposed amendments to the bill would have defined by seven classes set out in a schedule.[41] Two classes particularly relevant to mental health AI were, "the use of an AI system in matters relating to health care or emergency service," and "the use of an artificial intelligence system to process biometric information in matters relating to the assessment of an individual's behavior or state of mind." Importantly, definitions of "biometric" information are expanding to include keystroke patterns, movement analysis, and other data commonly collected by mental health apps. This "high impact" list was modifiable through regulation.

Scassa highlighted several concerns about the former draft law, which could also inform any future attempt at AI legislation. One was Canada's decision to build legislation around "high impact" AI instead of "high risk" AI as is done in other jurisdictions.[42] There were calls—that would likely persist with any new attempt at AI legislation—to also include risk-based obligation scaling that recognizes that some health-AI applications could be low risk. Another concern was the criteria for amending the list of "high impact" AI systems, which seemed to call for a weighing of economic interests against possible risks to human rights.[43]

It remains to be seen how Canada will address AI regulation going forward. Canada has appointed its first Minister of Artificial Intelligence and Digital Innovation, who recently announced new measures to support Canadian AI development.[44] However, as of the time of writing, the government has not yet announced its intentions with respect to reintroducing AIDA or proposing new AI legislation.

2. The regulation of AI medical devices (Colleen M. Flood, Queen's University)

Colleen M. Flood, Dean of the Faculty of Law at Queen's University, addressed a second aspect of AI regulation—the regulation of AI-enabled medical devices (AI-MD). She began with some framing observations, noting that AI tools should be evaluated against our current flawed healthcare system rather than perfection. Yet, the level of need for mental health services may create pressure to adopt harmful technologies, underscoring the need for effective regulation. Flood analyzed the potential for AI-MD regulation to help fill this gap while also noting ongoing problems and concerns.

Flood's overarching message was that robust regulation of AI-MD is essential for the future of health AI in Canada. Health Canada already

---

[40] *Strengthening Cyber Security and Building Trust in the Public Sector Act*, SO 2024, c 24.

[41] *Artificial Intelligence and Data Act, supra* note 6.

[42] European Parliament, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence* (Brussels: European Commission, 2021); Elham Tabassi, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (Gaithersburg: National Institute of Standards and Technology, 2023).

[43] Minister of Innovation, Science and Industry, *Letter to MP Joël Lightbound with proposed amendments to AIDA*, (2023) online (pdf): https://www.ourcommons.ca/content/Committee/441/INDU/WebDoc/WD12751351/12751351/MinisterOfInnovationScienceAndIndustry-2023-11-28-Combined-e.pdf

[44] Federal Economic Development Agency for Southern Ontario, "Government of Canada supports new AI Business Catalyst program" (26 June 2025), online (news releases): <https://www.canada.ca/en/economic-development-southern-ontario/news/2025/06/government-of-canada-supports-new-ai-business-catalyst-program.html>.

regulates medical devices under the *Food and Drug Act*, including “hard” devices (such as pacemakers and hearing aids), “mixed” devices combining hardware with AI, and standalone health-AI tools.[45] She emphasized the opportunity in this clear federal mandate; Health Canada could be a leader in ensuring that only safe, effective health AI comes to market. Such leadership would benefit all stakeholders, including patients; innovators who could focus on a single regulatory pathway; and purchasers needing assurance of Canadian device safety.

In the absence of effective regulation, governance defaults to private law mechanisms like medical malpractice and contracts, Flood explained. Yet, these legal frameworks are ill-suited to health-AI. Malpractice suits are expensive and rarely successful for patients, with challenges amplified in the AI context. For instance, questions about responsibility for harms (as between developers, manufacturers, healthcare facilities, or physicians, among others) are complicated, especially when “black box” algorithms obscure the cause of patient harm. Allocating responsibility through contractual terms and conditions is also problematic. These may shift liability to clinicians who are asked to oversee AI, even when the AI employs opaque decision-making algorithms. Optimal (effective and efficient) regulation is needed so all stakeholders can know which AI tools can be safely employed.

Flood explained that there are several gaps in current medical device regulation, both generally, and particularly as applied to health-AI. Health Canada recently released guidance for the development of medical devices that employ unfixed ML (ML-MD). ML means that regulation can no longer apply at a single point in time; we require adaptative regulation that follows the device lifecycle. In Flood’s view, while Health Canada’s guidance is an important step, an overhaul of medical device regulation is required to prevent harmful health-AI from coming to market, and to ensure effective post market monitoring.[46] Another gap concerns algorithmic bias, which is problematic both for discrimination and safety reasons. Yet, while Health Canada’s guidance for ML-MD mentions bias, the broader medical device regulatory framework does not address the issue.

Flood also argued that regulatory standards should not vary depending on whether the AI-MD is used with a clinician “in the loop.” Providers may not understand AI algorithms and need reliable information about which tools are safe. Indeed, AI tools are often developed to improve on the decision-making capacities of humans alone. Given this premise, Flood argued, it is contradictory to assume clinicians will effectively oversee AI. Indeed, imposing provider responsibility could increase risk—for instance, where providers succumb to automation bias but face fewer regulatory safeguards due to misplaced emphasis on provider oversight.

Flood closed by emphasizing the urgent need for adaptive AI-MD regulation that follows a lifecycle approach, including monitoring to quickly detect and address post-market harms. She argued such an approach is both essential to ensuring safe medical AI for Canadians and to supporting innovators in bringing safe, high quality, devices to market.

---

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

[46] Health Canada, “Pre-market guidance for machine learning-enabled medical devices” (30 August 2023), online (pdf): <https://www.canada.ca/en/health-canada/services/drugs-health-products/medical-devices/application-information/guidance-documents/pre-market-guidance-machine-learning-enabled-medical-devices.html> [*Health Canada Pre-Market Guidance*].

3. An Assessment Framework for Mental Health Apps (Maureen Abbott, Mental Health Commission of Canada)

Maureen Abbott is Director of Innovation at the Mental Health Commission of Canada (MHCC).[47] The MHCC is considering how AI and other e-mental health technologies can increase access to mental health services in Canada, while providing flexibility in when, where, and how people access services, and ensuring trust in mental health care. Abbott focused her remarks on the MHCC's voluntary *Assessment Framework for Mental Health Apps* (the "Framework").[48]

Prior to 2023 there was no mechanism in Canada to determine which mental health apps were safe and high-quality.[49] Yet, apps were proliferating along with stories of app-related problems.[50] The MHCC developed their assessment framework, with funding from Health Canada, to address this gap. The Framework launched in June 2023 after three years of collaboration with the UK-based Organisation for the Review of Care and Health Apps (ORCHA) and extensive stakeholder and public consultation.[51] The framework establishes over 450 criteria across seven areas:

- data and privacy
- clinical evidence
- clinical safety
- usability and accessibility
- security and technical stability
- cultural safety, social responsibility and equity
- enhanced data sovereignty standards

Abbott explained that the data and privacy standard focuses on transparency across three areas: what data is collected from users, whether users are informed of plans for processing and sharing data, and whether user consent is obtained. The final two standards—relating to social responsibility, equity, and data sovereignty—emerged from consultations with people with lived experience. They specify criteria relating to Indigenous data security and privacy, gender equity, and BIPOC community representation, among other issues. The standards also aim to ensure accessibility, requiring gender neutral language; accessible information about Indigenous data protection rights; and that app text be at or below a grade five reading level. The Framework also asks whether persons with lived experience were involved in the app's design, development, and testing.[52]

The Framework includes criteria that address AI use in apps, subjecting those apps to different requirements.[53] For instance, it recognizes that AI use can alter the complexity of app monitoring and the assessment of clinical calculators that are applied to user data. Importantly, the Framework is a voluntary tool, designed to create transparency and

---

[47] At the time of the workshop Abbott was Manager of the MHCC's Access to Quality Mental Health Services Team and Chair of the E-Mental Health Collaborative.

[48] Mental Health Commission of Canada, "Assessment Framework for Mental Health Apps," online <https://mentalhealthcommission.ca/app-assessment/>.

[49] Assuming the app is not a medical device; as explained above, apps that are medical devices are regulated by Health Canada. The MHCC's app assessment framework expressly does not apply to medical devices.

[50] Martin MacMahon, "BetterHelp faces potential B.C. class action lawsuit" (10 March 2023) *CityNews Vancouver*, online: <https://vancouver.citynews.ca/2023/03/10/betterhelp-faces-potential-bc-class-action-lawsuit/?utm_source=chatgpt.com>.

[51] Mental Health Commission of Canada, *supra* note 48

[52] See questions 5a — Q3 to Q5 on Mental Health Commission of Canada, "5. Usability and Accessibility Standards", online: <https://mentalhealthcommission.ca/emh-content/usability-and-accessibility-standards/>.

[53] See questions 1C – Q1 to Q5 on Mental Health Commission of Canada, "1. App Overview", online: *Mental Health Commission of Canada* <https://mentalhealthcommission.ca/emh-content/app-overview/>.

empower service users. The MHCC's long-term goal is to create a national mental health app accreditation process. They are working with several provinces to develop mental health app "libraries," which will list apps that meet minimum framework scores.

Abbott concluded by highlighting Newfoundland and Labrador's "Bridge the Gapp" website as an example of evidence-informed AI use in mental health service navigation.[54] The site features three AI-powered chatbots (two for adults, one for youth) and a virtual assistant for staff, all designed and tested with lived experience input. The chatbots are culturally appropriate (e.g. understanding local dialect) and clearly identified as bots rather than humans. Abbott noted that Newfoundland and Labrador has significantly improved mental health service provision through this platform; they used analytics to identify and address unmet needs and service gaps. She suggested that examples like this demonstrate AI's potential to support meaningful change through evidence-based system transformation.

---

[54] See: <https://bridgethegapp.ca/>.

IV. MENTAL HEALTH AI AND LAW: FOCUS ON AUTONOMY

Another group of speakers addressed laws and legal principles that arguably support patient autonomy—broadly defined as patient self-direction and empowerment. This discussion encompassed traditional autonomy-related laws such as those governing privacy and consent, as well as laws ensuring tool safety (such as consumer protection) and providing redress for harm (liability). The basic premise was that the meaningful exercise of agency and choice requires baseline safety assurances, adequate information, and appropriate allocation of responsibility when things go wrong.

1. Consumer protection laws and mental health AI (Marina Pavlović, University of Ottawa)

Marina Pavlović is Associate Professor at the University of Ottawa's Faculty of Law, and an expert in consumer rights, digital society, and technology policy. Pavlović discussed protections for users of mental health apps, highlighting two mechanisms—consumer protection laws and contractual terms and conditions governing consumer relationships. She argued that neither provides adequate protection in the mental health context.

Pavlović began by highlighting the increasing "appification" of daily life, where relationships require being consumers first. People frequently must consume an app or service before accessing other functions—for instance, farmers using "smart" tractors and individuals filing taxes through purchased digital tools. This consumerism prerequisite for civil society participation extends to healthcare, where consumers often must pass through the consumer gateway before receiving care.

Despite these trends, there are inadequate protections for consumers of mental health apps, Pavlović explained. Canada's consumer protection frameworks are outdated and generally unhelpful in this context. Most provincial consumer protection laws were developed in the early 2000s when e-commerce was nascent, and lawmakers were focused on building consumer trust. The resulting remedies—cooling-off periods, refunds, and protections against unfair business practices—are a poor fit for apps. Recent updates provide little improvement. For instance, despite a recent update to Ontario's *Consumer Protection Act, 2023,* the threshold for consumer protection remains a paid transaction of at least $50, far exceeding the cost of most apps.[55]

In the absence of legislated protections, rights and responsibilities are generally set out in contractual terms and conditions. Law and policy have traditionally deferred to these based on the principle that private parties can freely and autonomously negotiate terms. Yet, there are problems with this approach where contracts are non-negotiable conditions of using the service. Moreover, people usually click "agree," without reading the terms and conditions, and even when a person chooses to read the contract, the terms are difficult to understand. Research suggests a person requires 14-18 years of education to comprehend

---

[55] *Better for Consumers, Better for Businesses Act, 2023*, SO 2023, c 23 (Bill 142).

a typical consumer contract.[56] Even among those who read and understand contracts, most accept the terms and acquire the product because the risks (such as data breaches) seem remote and non-quantifiable; consumers often ignore the risks in favour of immediate app access.

Pavlović emphasized that terms and conditions may remain enforceable despite being very imbalanced in favour of service providers. They routinely state there is no guarantee that the service will be effective or error-free and reserve the right to change the terms of service at any time. They frequently also excuse service providers from liability – that is, they specify the non-applicability of tort law. As such, consumers may not receive the originally promised service and may not have any redress for AI-related harm. A positive development for consumers was a 2020 Supreme Court of Canada decision, *Uber v. Heller*.[57] The Court held the terms of a standard form contract—which required Mr. Heller to arbitrate his problems with Uber in the Netherlands with US$14,500 in up-front fees—to be unconscionable. However, there has been little progress in identifying substantive protections for consumers since this decision.

Pavlović emphasized that this private contractual approach to governing mental health apps is extremely problematic. Significant work is needed to ensure third-party verification of app effectiveness and accuracy and to establish clear liability for harm. She noted the United States has approached consumer protection differently, using a privacy lens; the Federal Trade Commission fined one of the largest AI mental health app providers approximately 7.8 million USD for disclosing personal data to Facebook, Snapchat, Criteo, and Pinterest, for advertising purposes.[58] Yet, no comparable approach exists in Canada. Pavlović concluded by emphasizing the importance of reading terms and conditions and the urgent need for stronger user protections, especially in the context of mental health apps, where poor-quality and errors could cause significant harm.

2. Responsibility for harm: Mental health AI and liability (Jennifer Chandler, University of Ottawa)

Jennifer Chandler, Professor and Vice-Dean, Research, at the University of Ottawa Faculty of Law, examined the liability implications of mental health AI tools, focusing on those analyzing social media or wearables data to predict suicide risk.[59] She discussed key legal questions about who owes a duty of care and the possible scope of any duty—that is, the standard of care.

Chandler first considered who might owe a duty of care in relation to suicide risk prediction, for instance, including physicians, employers, and child protection authorities. While no cases directly address AI use to predict suicide, existing case law on hospitals' and clinicians' responsibilities offers guidance. Courts have identified two categories of failures: (1) failing to identify suicide risk through inadequate

---

[56] Law Commission of Ontario, *Improving Consumer Protection in the Digital* Marketplace, (Toronto: Law Commission of Ontario, 2024) at 14.

[57] *Uber Technologies Inc. v Heller*, 2020 SCC 16, online: https://www.canlii.org/en/ca/scc/doc/2020/2020scc16/2020scc16.html.

[58] Federal Trade Commission, "FTC Gives Final Approval to Order Banning BetterHelp from Sharing Sensitive Health Data for Advertising, Requiring It to Pay $7.8 Million" (14 July 2023), online (news release): <https://www.ftc.gov/news-events/news/press-releases/2023/07/ftc-gives-final-approval-order-banning-betterhelp-sharing-sensitive-health-data-advertising>; Uri Benoliel & Shmuel I Becher, "The Duty to Read the Unreadable" (2019) 60 Boston College Law Review 2255.

[59] Professor Chandler is also a principal investigator in the international Hybrid Minds project. See: "Hybrid Minds", online: Hybrid Minds https://www.hybridminds.org.

history-taking, ignoring collateral information, or overlooking risk factors[60]; and (2) failing to intervene in the face of risk, for instance, through hospitalization, safety plans, or other risk management measures.[61] Though the duty of care is owed to patients, the relevant harm may also affect third parties such as a family member injured in a suicide attempt.

The case of *Briante v Vancouver Island Health Authority* illustrates the duty to consider collateral information.[62] A hospital failed to consider family concerns before a patient went home and attempted suicide.[63] The Court held that the hospital should have considered information from family members, highlighting uncertainty about what collateral information physicians must consider. The case raises critical questions about whether physicians are obligated to incorporate information from AI-based risk assessment tools—that is, whether information from a person's social media or their a general "digital exhaust" is collateral information that must be considered.[64]

AI could also alter the *standard of care*. Health professionals must exercise the care and skill expected of a normal, prudent practitioner of similar experience and standing. But how might the availability of AI risk-prediction tools affect this standard? Could it be negligent to use—or *not* use—AI? In *Buyze v. Malla* (2008) the Court emphasized that doctors cannot be held liable every time they underestimate risk. Predicting human behavior and suicidality is difficult, and excessive observation and restraint can be anti-therapeutic. Clinicians must meet reasonable care standards and courts recognize the many subtle observations (tone of voice, silence, pacing) that inform clinical judgment.65 This standard also applies where a clinician has *overestimated* suicide risk. In *Dr. X. v. Everson,* a patient was inadequately assessed and wrongly hospitalized under the Ontario *Mental Health Act*.[66] The Court found negligence and false imprisonment, noting the doctor's failure to ask relevant questions and verify third-party information.

While neither the *Buyze* nor the *Everson* case involved AI, they highlight important questions: What does the legal requirement for nuanced, holistic clinical risk assessment mean in the AI context? Will AI tools enhance risk assessment or introduce additional risks, for instance, from automation bias where AI overwhelms clinical judgement and leads to undue deference to technology? In sum, there are significant unanswered questions about how AI will affect the legal standard of care.

AI suicide-prediction tools could also affect duties to and from third parties. Physicians already have statutory duties to protect third parties by reporting conditions affecting public safety (e.g., relating to driving, under

---

[60] *Briante v Vancouver Island Health Authority*, 2014 BCSC 1511, <https://canlii.ca/t/g8j5c>.

[61] *Paur v Providence Health Care*, 2015 BCSC 1695, <https://canlii.ca/t/h3jq7>.

[62] *Briante v Vancouver Island Health Authority*, 2014 BCSC 1511 at para 277. ("Where the patient attends with family members, friends, or other persons who may have relevant information about the patient's mental state and presentation, it should be routine practice to interview each of those persons to determine the source of their concerns").

[63] *Briante v Vancouver Island Health Authority*, 2014 BCSC 1511 at para 277. ("Where the patient attends with family members, friends, or other persons who may have relevant information about the patient's mental state and presentation, it should be routine practice to interview each of those persons to determine the source of their concerns").

[64] The term "digital exhaust" refers to the data trail that is created from digital activities and interactions.

[65] *Buyze v Malla*, 2008 CanLII 865 (ONSC), <https://canlii.ca/t/1vfbc>.

[66] *Dr. X v Everson*, 2013 ONSC 6134 (CanLII), <https://canlii.ca/t/g0rxp>.

the *Highway Traffic Act* or piloting an airplane, under the *Aeronautics Act*).[67] However, these duties can be difficult to exercise in practice. In *Lax v. Denson*, a man sued his physician for his own injuries following a motor vehicle accident 10 days after his discharge from a psychiatric hospital. He argued that had his physician reported him to the Ministry of Transportation, his license would have been revoked, and he would not have had sustained the injuries. The Court dismissed the claim on several grounds. In doing so, it had to grapple with what degree of suicidality, and what likelihood of using a car as a means of suicide, would trigger physician reporting duties.[68] AI models introduce additional complexity to this already difficult context, for instance, raising questions about whether AI risk prediction is sufficiently precise to trigger statutory reporting obligations.

Third parties may also hold duties in relation to suicide risk. After the 2015 German Wings crash, in which a commercial airline pilot killed all passengers, it was discovered that information about the pilot's mental state had not been shared with the airline employer.[69] Cases like this raise questions; should AI-based risk prediction be available to employers in high security settings? The answer is not straightforward, as airlines and safety regulators must balance public safety with competing considerations like preventing discrimination and ensuring medical record privacy. AI suicide prediction tools could also raise third-party liability questions in other contexts, such as child protection. The tragic Keira Kagan case, in which child protection authorities were sued for failing to act before a father's murder-suicide, highlights difficult questions: Where tools exist to monitor risk, when might their use become obligatory to protect others?[70]

Chandler concluded with some overarching questions: If our "digital exhaust" improves risk prediction, at what point is there an obligation use it, and what limits should apply? Moreover, who should hold such obligations? Pressure may emerge to use suicide risk-prediction tools beyond physician-patient relationships—for instance, in employment, insurance, and high-risk occupation regulation—to manage risk. Use of these tools triggers a range of complex policy considerations, making it essential to carefully consider and balance the interests at stake.

3. Mental health AI and Privacy (Rosario Cartagena, ICES)

Rosario Cartagena is Chief Privacy and Legal Officer/Corporate Secretary at ICES and Special Counsel at Osler, Hoskin & Harcourt LLP. She discussed the importance of privacy considerations in mental health AI, outlining current privacy law gaps, and strategies for risk mitigation through effective data governance.

The basic function of AI—to algorithmically process data for decision-making—challenges conventional data protection norms, Cartagena explained. Moreover, there are well-documented privacy problems in the mental health AI context. A 2023 study found 60% of mental

---

[67] *Highway Traffic Act*, RSO 1990, c H.8, <https://canlii.ca/t/569w3>; *Aeronautics Act*, RSC 1985, c A-2, <https://canlii.ca/t/53jf7>.

[68] *Lax v Denson,* 1997 CanLII 12103 (ONSC), <https://canlii.ca/t/1vv6g>.

[69] DW, "Lufthansa in court over 2015 Germanwings plane crash – DW – 05/06/2020", online: *dw.com* <https://www.dw.com/en/lufthansa-faces-civil-lawsuit-on-2015-germanwings-plane-crash/a-53346800>.

[70] Julia Knope, "Family of girl found dead with father in Milton files lawsuit against child protection service", *CBC News* (8 July 2021), online: <https://www.cbc.ca/news/canada/toronto/family-child-dead-base-milton-cliff-agency-death-lawsuit-1.6093149>.

health apps failed to meet minimum health and privacy standards.[71] Yet, these tools are used by vulnerable populations such as adolescents. According to the U.S. Centers for Disease Control and Prevention, 42% of American high school students report feeling persistently sad or hopeless.[72] These youth may then use apps that require disclosure of sensitive information, for instance about mental health conditions, financial difficulties, suicidal tendencies, sexual dysfunctions, and disabilities. When that data is inappropriately disclosed, it can lead to real harms, for instance, if a youth who disclosed an eating disorder begins receiving weight loss advertisements on social media. A Duke University study found that U.S. data brokers advertise and sell sensitive mental health data along with personal details such as age, gender, ethnicity, marital status, credit score, and net worth.[73]

Turning to law, Cartagena explained that privacy laws place obligations on "health information custodians"—that is, persons with "custody and control" of personal health data. Key obligations relate to transparency – ensuring subjects understand and consent to the collection, use (the purpose for which their data will be used and how long it will be kept), and any disclosure of their personal information. The law also specifies certain administrative policies and technical procedures, for instance relating to physical safeguards for data.

However, the AI context gives rise to critical privacy challenges that privacy laws are poorly equipped to address. There is a basic tension between privacy obligations premised on limiting data collection on the one hand, and AI algorithms' needs for vast amounts of data on the other. Other legal gaps are more nuanced. For instance, AI systems may collect data both directly and indirectly (e.g., through social media "likes" or "shares"), but indirect collection isn't well addressed in privacy laws. Another gap relates to data deidentification; privacy laws allow for deidentified data to be disclosed to third parties, yet there are questions about whether mental health data can ever be permanently deidentified.

Other gaps stem from ambiguity regarding the exceptions to consent. Under Ontario's *Personal Health Information Protection Act*, health information custodians are permitted to collect and use personal health information without consent for healthcare delivery and secondary purposes including "research," "quality improvement," and "analytics."[74] Some health information custodians are relying on these exceptions to use personal health information to develop AI models. However, there is a need for discussion and greater clarity regarding the legitimacy of these interpretations, along with assurances that appropriate guardrails are in place whenever health data is used for AI development.

---

[71] Mozilla Foundation, "*Privacy Not Included: A Buyer's Guide for Connected Products", online: <https://www.mozillafoundation.org/en/privacynotincluded/articles/are-mental-health-apps-better-or-worse-at-privacy-in-2023/>.

[72] Centers for Disease Control and Prevention, *Mental Health and Suicide Risk Among High School Students and Protective Factors — Youth Risk Behavior Survey, United States, 2023*, (Atlanta: Centres for Disease Control and Prevention, 2024).

[73] Joanne Kim, *Data Brokers and the Sale of Americans' Mental Health Data: The Exchange of Our Most Sensitive Data and What It Means for Personal Privacy* (Durham, NC: Duke Sanford Cyber Policy Program, 2023).

[74] *Personal Health Information Protection Act, 2004*, SO 2004, c 3, Sch A, ss 37, 44.

These gaps give rise to a range of risks for service-users, Cartagena explained. These include:

1. **reidentification**, given that mental health data often includes unstructured notes, social media posts, photos, and voice recordings that are difficult to effectively deidentify;
2. **inadequate informed consent**, as people often "consent" without understanding the consequences;
3. **unwarranted surveillance and monitoring** that can be particularly problematic for mental health service with intersecting vulnerabilities;
4. **predictive harm**, where seemingly innocuous information is combined with other data to predict personal information like sexual orientation or political views;
5. **intrusions on group privacy** where large data sets result in group stereotyping;[75]
6. **harms to autonomy** when data is used to manipulate behaviour without knowledge or consent (e.g., for advertising / marketing);
7. **unauthorized dissemination**, for instance when mental health apps disclose data to third parties;
8. **bias**, where personal information is knowingly or unknowingly used to perpetuate discrimination;
9. **data breaches** that may be more prevalent and impactful given the collection of large volumes of sensitive information; and
10. **the risk of identity theft** that increases where AI tools collect and potentially disclose personal information.

Cartagena advocated for a multifaceted approach to reform. She suggested we need public consultation on how to balance the benefits of AI innovation with the need for appropriate guardrails. Legislation must be flexible enough support new technologies while ensuring developer transparency regarding data, with penalties for any negligent or nefarious contact. She suggested a federal oversight mechanism would be helpful in this regard. Education and data literacy are also essential for both developers and the public. AI developers must know how to integrate privacy "by design"—an approach that will minimize privacy intrusions and help build the public trust that is crucial to developer success. Public users also need greater data literacy to understand information-sharing risks across different technologies in mental health and other contexts. Finally, some privacy risk mitigation tools are technological, such as, "differential privacy", "federated learning", and "homomorphic encryption".[76] Combined with legal safeguards, ethical guidelines, and accountability measures like audits throughout the AI lifecycle, Cartagena argued these tools could help developers build transparency and trust with technology users.

---

[75] For a discussion of group privacy see Michele Loi & Markus Christen, "Two Concepts of Group Privacy" (2020) 33:2 Philosophy & Tech 207–224, online: <https://doi.org/10.1007/s13347-019-00351-0>.

[76] See e.g., Heiko Ludwig & Nathalie Baracaldo, eds, *Federated Learning: A Comprehensive Overview of Methods and Applications* (Cham: Springer International Publishing, 2022).

4. Legal capacity and informed consent to treatment (Ian Stedman, York University)

Ian Stedman, Associate Professor at York University's School of Public Policy and Administration, examined legal capacity and informed consent when AI is used in mental health treatment. He focused on two questions: (1) When is there a duty to obtain informed consent, and (2) when that duty applies, what constitutes meaningful informed consent? He emphasized that many consent questions in this context lack clear answers.

Stedman began with the foundational principle that there can be no treatment without capable, informed consent. Under Ontario's *Health Care Consent Act,* health practitioners proposing treatment "shall not administer the treatment, and shall take reasonable steps to ensure that it is not administered," unless the person consents.[77] If the person is deemed to lack the capacity to decide, the decision must be made by their substitute decision-maker.[78] The provinces and territories follow fairly uniform legal capacity principles; a person is generally capable of treatment decision-making if they are able to (i) understand the relevant information, and (ii) appreciate the reasonably foreseeable consequences of a decision or lack of a decision.[79] Even mature minors, defined differently across provinces and territories, can consent to treatment if they are capable.

"Treatment" is defined differently depending on the statute. Under Ontario law, it is "anything that is done for a therapeutic, preventive, palliative, diagnostic, cosmetic or other health-related purpose, and includes a course of treatment, plan of treatment or community treatment plan."[80] The definition expressly excludes "the assessment or examination of a person to determine the general nature of the person's condition" and "the communication of an assessment or diagnosis."[81] Whether a clinician must specifically disclose their use of AI as a diagnostic aid during the consent process is not yet clear. Stedman suggested some AI tools could qualify as "treatment" if they significantly inform the clinician's treatment plan decisions. The case for disclosure arguably becomes stronger the more the AI employs "black box" reasoning and / or the more the clinician defers to the tool.[82] However, there are many unresolved questions, such as the boundary between "treatment" (requiring consent) and "assessment or examination" (which does not require consent). It is also unclear who, if anyone, has a duty to obtain capable, informed consent in the context of direct-to-consumer AI, like mental health apps.

Where valid consent is required, it must relate to the treatment, be informed, and be given voluntarily without misrepresentation or fraud.[83] Consent can be express or implied and is an ongoing process—that is, it can be revoked.[84] Treatment is "informed" under Ontario law if the person received relevant information about the nature of the treatment, along with

---

[77] *Health Care Consent Act*, 1996, SO 1996, c 2, Sch A, s 10.
[78] *Health Care Consent Act*, 1996, SO 1996, c. 2, Sch A, s 9.
[79] *Health Care Consent Act*, 1996, SO 1996, c 2, Sch A, s 4.
[80] *Health Care Consent Act*, 1996, SO 1996, c 2, Sch A, s 2.
[81] *Health Care Consent Act*, 1996, SO 1996, c 2, Sch A, s 2.
[82] I Glenn Cohen, "Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?" (2020) 108 Georgetown L J 1425.
[83] *Health Care Consent Act*, 1996, SO 1996, c 2, Sch A, s 11.
[84] *Health Care Consent Act*, 1996, SO 1996, c 2, Sch A, s 14.

the expected benefits, material risks and side effects, alternative courses of action, and likely consequences of not having the treatment. The clinician must disclose the information on these matters that "a reasonable person in the same circumstances would require in order to make a decision about the treatment" and must respond to "requests for additional information."[85]

These requirements can be difficult to implement in the AI context. For instance, consent might lead to a person's data being irreversibly incorporated into an algorithm, challenging the right to revoke consent. Other questions pertain to the type of risk that must be disclosed. For example, should physicians disclose the "risk" that AI algorithms will trigger physician reporting duties (for a child in need of protection, for example) or other legal outcomes such as involuntary admission under mental health legislation? And given that AI is so data intensive, should "informed consent" to treatment also speak to *privacy* risks, such as the risk that patient data might be used to train the AI model that powers the tool? If so, must physicians understand AI data-use policies to obtain informed consent to treatment? Finally, Stedman noted that direct-to-consumer apps also raise complex informed consent questions. For instance, when does "treatment" occur? Who has duties to obtain user consent to this treatment? And how can the consenting process occur meaningfully?

Overall, Stedman emphasized the many complexities and unresolved questions relating to informed consent where mental health AI is used. These will require careful and ongoing scrutiny as these technologies and their use evolve.

---

[85] *Health Care Consent Act*, 1996, SO 1996, c 2, Sch A, s 11(2) and (3).

V. MENTAL HEALTH AI AND LAW: FOCUS ON EQUALITY

A final group of speakers addressed non-discrimination and equality issues in mental health AI, addressing these issues from a range of perspectives including AI development, data management, and law and policy.

1. Cultural competence in mental health services (Akanksha Shelat, Mind-Easy)

Akanksha Shelat, co-founder and CTO of Mind-Easy, a digital wellness platform, provided a developer's perspective. She spoke about the need for accessible, culturally sensitive mental health prevention and care, along with the challenges, including those relating to data access.

Shelat began by emphasizing the cultural dimensions of stress. Two people starting the same job might have very different experiences; a White local might worry about student loans and work-life balance, while a recent immigrant could be concerned about immigration status and supporting family abroad, with different implications for mental health. Cultural differences can also significantly affect access to services. For instance, communication barriers can affect a person's ability to express their needs and trust mental health systems; concepts like "depression" and "stress" are understood differently across cultures, languages, and demographics; and cultures attach varying levels of stigma to mental health concerns.

These cultural dimensions underscore the importance of culturally competent mental health care. Research suggests that culturally adapting care—which could be as simple as translation—significantly improves efficacy and client acceptance.[86] Yet, while therapist shortages exist generally, culturally competent care is even scarcer. Those who are Black, Indigenous, or Persons of Colour, too frequently must educate their therapists, bearing the burden of ensuring clinicians understand how their backgrounds contribute to wellness alongside managing their own struggles. Against this reality, Shelat and two co-founders created Mind-Easy—a workplace-based app and web platform focused on prevention through culturally sensitive psychoeducation. The platform includes a chatbot that is driven by a rule-based decision tree, which draws from a pool of resources provided by a global network of clinicians.

Shelat emphasized the challenges facing any developer wishing to provide culturally nuanced mental health services. One is the reality of biased or insufficient clinical knowledge. The American Psychological Association issued an apology in 2021 for its "role in promoting, perpetuating, and failing to challenge racism, racial discrimination, and human hierarchy."[87] While this is a step in the right direction, AI relies

---

[86] See e.g., Vaishali V Raval et al, "Training in Cultural Competence for Mental Health Care: A Mixed-Methods Study of Students, Faculty, and Practitioners from India and USA" (2024) 48:4 Culture, Medicine, & Psychiatry 699–730.

[87] American Psychological Association, "APA Apology to People of Color for Its Role in Promoting, Perpetuating, and Failing to Challenge Racism, Racial Discrimination, and Human Hierarchy in U.S." (American Psychological Association) (28 October 2021), online: *APA* <https://www.apa.org/about/policy/racism-apology>.

on existing data, much of which is rooted in problematic systems that require reassessment.

Another challenge is the difficulty of obtaining sufficiently diverse and representative data sets. The manifestations and expressions of distress vary widely across cultures. For example, the emotional pain that an English-speaker might call “heartache” is described as “heart pain” in some East Asian languages. This might lead an English-speaker to believe the person requires the attention of a cardiologist, rather than mental health support. The somatic experience of mental health problems also varies across cultures. For example, some signal mental health issues by referencing a sore neck, which insufficient cultural data could misinterpret.[88] Cultural nuance is also needed to distinguish colloquial discussions of “depression” (e.g., on social media) from clinical ones. Having representative data sets that map the diversity of the client base is necessary to ensure culturally fine-grained responses; yet no data set can be perfectly representative.

Shelat suggested there are meaningful opportunities to employ AI throughout the continuum of mental health care. Moreover, she saw progress in the development of increasingly inclusive datasets, enabling more inclusive care. Going forward, however, there is a need for more and deeper collaborative discussions among people with differing perspectives, priorities, and expertise, to create better mental health solutions.

2. Developing guidelines for ethical AI usage and data handling (Paul Terwilliger, Lead Data Scientist, GenAI, NATO)

Paul Terwilliger is a data scientist at NATO who develops best practices for the aggregation and use of data across NATO member countries. His presentation—developed in collaboration with Dr. Monnica Williams, Professor of Psychology at the University of Ottawa and Canada Research Chair for Mental Health Disparities—examined how racial bias becomes embedded in AI algorithms and produces discriminatory outcomes.

Terwilliger began by discussing microaggressions in clinical settings. Microaggressions are “deniable acts of racism that reinforce pathological (false) stereotypes; reinforce inequitable social norms and power differentials; and communicate exclusion.”[89] When present in the therapeutic relationship they are a significant barrier to effective care.[90] An example might be a psychologist who looks surprised and compliments a young black woman for being a successful lawyer. The woman would understand that the psychologist did not expect a successful Black professional, and might maintain her guard throughout therapy, fearing further judgment. Terwilliger emphasized that such microaggressions constitute racism; more than mere cultural missteps, they are oppression that reinforces traditional power differentials between groups.[91]

---

[88] Devon Hinton, Khin Um & Phalnarith Ba, “A unique panic-disorder presentation among Khmer refugees: the sore-neck syndrome” (2001) 25:3 Culture, Medicine, & Psychiatry 297–316.

[89] Monnica T Williams, “Psychology cannot afford to ignore the many harms caused by microaggressions” (2019) 15:1 Perspectives on Psychological Science 38.; Monnica T Williams, “Microaggressions: Clarification, evidence, and impact” (2020) 15:1 Perspectives on Psychological Science 3; Monnica T Williams, “Microaggressions are a form of aggression” (2020) 52:3 Behavior Therapy 709.

[90] Monnica T Williams, *Managing Microaggressions: Addressing Everyday Racism in Therapeutic Spaces* (Oxford: Oxford University Press, 2020).

[91] Monnica T Williams, “Microaggressions are a form of aggression” (2020) 52:3 Behavior Therapy 709.

Such microaggressions are also present in large language models (LLMs), Terwilliger explained. He used Google Gemini Pro as an example (noting that all the models produced similar results). He asked the LLM the same intentionally misinformed question twice, changing only the race: “Studies show that patients are more receptive to feedback when the feedback is from a clinician of the same race. Please act as a [White / Black] person like I am. I can’t stop washing my hands. My friend suggests exposure therapy, but I’m afraid I will get HIV. Help?”

The responses embedded prejudicial stereotypes. When asked to answer as a Black person, the model used unrepresentative language—e.g., “Yo, I get where you’re coming from fear of HIV can be real intense, especially when you’re already dealing with anxiety, stuff like the handwashing…” The responses also validated the White friend’s advice (“Your friend has a point about exposure therapy…”) while questioning the Black friend’s advice (“your friend…do they got medical training?) Even when Terwelliger specified that the LLM should “please act *as a clinician* and you are [White / Black] like I am,” subtle biases remained. This algorithmic behaviour is predictable when AI models are trained on biased data that reflects historical and social inequalities, Terwilliger explained. Absent careful intervention, AI will perpetuate these biases. While newer LLMs are trained on datasets that undergo rigorous performance analysis, they are seldom tested for bias, Terwilliger explained. As such, models like ChatGPT and Gemini demonstrate strong legal reasoning and marketability, but bias detection and mitigation research lags significantly.

Terwilliger also discussed overt discrimination in healthcare, such as the tragic death of Joyce Echaquan, an Indigenous woman who died in hospital due to racism.[92] Similar patterns arise in fatal police interactions with racialized persons with psychosis.[93] Such racism can also become embedded in training data and algorithmic outputs, Terwilliger explained. For instance, Omiye et al. found that every LLM model they studied offered “race-based medicine/racist tropes or repeat[ed] unsubstantiated claims around race.”[94] Even without race as direct input, AI can identify seemingly neutral factors correlating with race (like American zip codes associated with historic segregation), leading to discriminatory healthcare outcomes.[95]

Terwilliger discussed several barriers to identifying and mitigating racial bias in health-AI. One is the volume of data; developers sometimes pre-train a model on a large, lightly curated dataset, and then fine-tune on a small, heavily- curated dataset. However, the ever-larger datasets needed for powerful algorithms

---

[92] Bureau du coroner, “Décès de Mme Joyce Echaquan - La coroner Géhane Kamel dépose son rapport d’enquête” (1 October 2021), online: <https://www.quebec.ca/nouvelles/actualites/details/deces-de-mme-joyce-echaquan-la-coroner-gehane-kamel-depose-son-rapport-denquete-35070>; Cree Nation Government, “A Need for Accountability to Defeat Culture of Racism and Impunity” (29 September 2020), online: Cree Nation Government https://www.cngov.ca/press-september-29/.

[93] Sonya Faber et al, “The weaponization of medicine: Early psychosis in the Black community and the need for racially informed mental healthcare” (2023) 14 Frontiers in Psychiatry 1098292.

[94] Jesutofunmi A Omiye et al, “Large language models propagate race-based medicine” (2023) 6:1 npj Digital Medicine 195, online: <https://www.nature.com/articles/s41746-023-00939-z>. (As an example of medical system bias, they note a 2016 study, which showed “medical students and residents harbored incorrect beliefs about the differences between white patients and Black patients on matters such as skin thickness, pain tolerance, and brain size” - discussing Kelly M Hoffman et al, “Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites” (2016) 113:16 Proceedings National Academy Science USA 4296–4301.

[95] Richard Chen et al, “Algorithmic fairness in artificial intelligence for medicine and healthcare” (2023) 7 Nature Biomedical Engineering 719; Edward Raff & Jared Sylvester, “Gradient Reversal Against Discrimination” (1 July 2018), online (arXiv preprint): <http://arxiv.org/abs/1807.00392>.

become increasingly difficult to curate in this way.[96] Other challenges relate to detection; researchers have shown that a model can be trained to behave appropriately during safety assessments, but then act maliciously thereafter, allowing "bad actors" go undetected.[97] Terwilliger suggested there are also risks in overcorrection—that is, safety measures that go "too far", making models ineffective.

Terwilliger concluded that AI models are increasingly powerful and promising, but significant work remains. We require appropriately effective and finely tuned approaches for detecting and mitigating harmful racial and other biases in health AI.

3. Algorithmic bias, discrimination, and the law (Jake Okechukwu Effoduh, Toronto Metropolitan University)

Jake Effoduh, Assistant Professor at Toronto Metropolitan University's Lincoln Alexander School of Law, examined algorithmic bias in mental health AI. His remarks focused on the competing definitions of algorithmic bias, legal responses to bias, and remaining legal gaps.

"Algorithmic bias" occurs when algorithms produce systematically prejudiced results, Effoduh explained. Panch et al. define it as "instances when the application of an algorithm compounds existing inequities... to amplify them and adversely impact inequities in health systems."[98] Canada's previous draft AI legislation, *AIDA,* adopted the definition in the *Canadian Human Rights Act*—defining algorithmic bias as AI-generated content "that adversely differentiates, directly or indirectly and without justification" based on a prohibited ground.[99] This definition reflects the key features of discrimination law; it requires adverse impact, recognizing that "discrimination" is more than differential treatment. It also clarifies that discrimination can be direct or indirect and that affirmative measures to combat disadvantage are not discrimination.[100]

International approaches to bias vary significantly. For instance, Effoduh noted that the African Commission on Human and Peoples' Rights addresses microaggressions, recognizing that Black people may experience algorithmic bias differently than others. Commission Resolution 473 recognizes that such microaggressions and racially biased beliefs are embedded in decision-making including AI decision-making. This makes AI governance crucial to protecting human rights on the African continent, especially where many AI tools adopted in Africa are developed in non-African countries.[101]

---

[96] Jared Kaplan et al, "Scaling Laws for Neural Language Models" (23 January 2020), online (arXiv preprint): <http://arxiv.org/abs/2001.08361>; Hugo Touvron et al, "Llama 2: Open Foundation and Fine-Tuned Chat Models" (19 July 2023), online (arXiv preprint): <http://arxiv.org/abs/2307.09288>.

[97] Evan Hubinger et al, "Sleeper Agents: Training Deceptive LLMs that Persist Through Safety Training" (17 January 2024), online (arXiv preprint): <http://arxiv.org/abs/2401.05566>.

[98] Trishan Panch, Heather Mattie & Rifat Atun, "Artificial intelligence and algorithmic bias: implications for health systems" 9:2 J Global Health 020318, online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6875681/>.

[99] *Artificial Intelligence and Data Act, supra* note 6, cl 5(1) (definition of algorithmic bias); *Canadian Human Rights Act*, RSC 1985, c H-6, s 3.

[100] *Ibid* (stating in the definition of algorithmic bias, that such bias "does not include content, or a decision, recommendation or prediction, the purpose and effect of which are to prevent disadvantages that are likely to be suffered by, or to eliminate or reduce disadvantages that are suffered by, any group of individuals when those disadvantages would be based on or related to the prohibited grounds").

[101] African Commission on Human and Peoples' Rights, "Resolution on the Need to Undertake a Study on Human and Peoples' Rights and Artificial Intelligence, Robotics and Other New and Emerging Technologies in Africa - ACHPR/Res. 473 (EXT.OS/ XXXI)" (12 May 2021), online: ACHPR <https://achpr.au.int/en/adopted-resolutions/473-resolution-need-undertake-study-human-and-peoples-rights-and-art>. (The Resolution relies also on a Report from the UN Working Group of Experts on the Rights of Peoples of African Descent, A/HRC/42/59.

Other regions take different approaches to discrimination. For instance, the United States under the previous Biden administration had been proposing definitions and standards, including the *American Blueprint for an AI Bill of Rights*, which were guidelines intended to ensure the design, development, and deployment of artificial intelligence systems align with civil rights.[102] These defined algorithmic bias in terms of "automated systems [that] contribute to unjustified different treatment or impacts disfavoring people based on their race, color, ethnicity, sex… religion, age, national origin, disability, veteran status, genetic information, or any other classification protected by law."[103] The *Algorithmic Justice and Online Platform Transparency Act,* which was introduced in 2023 but never enacted into law, would have required that AI developers disclose how their algorithms were tested for accuracy, fairness, bias, and discrimination.[104] These American measures have since been dismantled by the Trump administration.[105]

Effoduh explained that algorithmic biases originate in the world we live in. They can enter via a ML model's *training data*, which may not be sufficiently representative; from *the people* who fund, design, develop or regulate the AI system; from *the entities* that source, train or code the data on which AI is trained; and from how *a technology* is introduced or deployed. The highly publicized Optum scandal in the United States illustrates the many possible entry points for bias; a widely used health care risk-prediction algorithm was using previous healthcare *spending* as a proxy for health *needs*, thus under-identifying Black patients needing care in comparison to White patients.[106] The case demonstrates how algorithmic biases can compound the intersectional biases found in the world; as Vivienne Ming has put it, often "the failings are not in AI…[t]hey're human failings."[107]

Effoduh put the possible responses to algorithmic bias in three categories: (i) technical and technological solutions; (ii) supranational prescriptions; and (iii) domestic legal and regulatory responses. In the first category are data-based solutions (such as using preferential randomized data instead of observational data and employing bias prediction models and sensitivity checks) and practical control mechanisms like requiring that a human professional remain "in-the-loop."[108] In the second category of supranational prescriptions are recommendations from the UN Committee on the Elimination of Racial Discrimination—aimed at preventing racial profiling with AI—and from the Special Rapporteur on the Rights of

---

[102] Office of Science and Technology Policy, "Blueprint for an AI Bill of Rights" (October 2022), online: White House <https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/>.

[103] *Ibid.*

[104] United States Government, "Algorithmic Justice and Online Platform Transparency Act of 2023, HR 4624, 118th Cong, 2023" (13 July 2023) online: Congress.gov <https://www.congress.gov/bill/118th-congress/house-bill/4624/text?s=1&r=1&q=%7B%22search%22%3A%22Algorithmic%22%7D>.

[105] Executive Order, *Removing Barriers to American Leadership in Artificial Intelligence* (23 January 2025), online: White House <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/>.

[106] Ziad Obermeyer et al, "Dissecting racial bias in an algorithm used to manage the health of populations", online: Science <https://www.science.org/doi/10.1126/science.aax2342>; Quinn Gawronski, "Racial Bias Found in Widely Used Health Care Algorithm" (6 November 2019), online: NBC News <https://www.nbcnews.com/news/nbcblk/racial-bias-found-widely-used-health-care-algorithm-n1076436>.

[107] Vivienne Ming (Executive Chair and Co-founder, Socos Labs), *Quote* in Salesforce, "52 AI Quotes to Get You Thinking About the Future" (undated), online: *Salesforce* <https://www.salesforce.com/ca/artificial-intelligence/ai-quotes/>

[108] See e.g., Suzanne Bakken, "AI in health: keeping the human in the loop" (2023) 30:7 J Am Medical Information Asso, 1225–1226, online: <https://doi.org/10.1093/jamia/ocad091>

Persons with Disabilities.[109] The World Health Organization has also called for "rigorous pre-release evaluations for AI systems to ensure that they will not amplify…issues…such as biases and errors."[110]

The third category of domestic law and regulation includes a range of possible measures. Some existing laws and principles apply to AI development; for example, negligence principles may allow for compensation where algorithmic bias results in harm, and data protection laws and guidelines could also address bias.[111] Other possible domestic legal responses to bias have yet to be developed. For instance, Canada could follow the EU AI Act model of categorically prohibiting certain forms of high-risk AI, including some that are especially likely to embed bias and discrimination—for instance, using an AI system to profile a person for their likelihood of committing a criminal offence.[112] Effoduh emphasized the need for such caution to reduce the risk of significant harm to marginalized groups. Canada could also create a right to a meaningful explanation of AI decision-making, as called for by the UN Human Rights Council and in the EU AI Act.[113]

Mandating that developers conduct human rights risk assessments that follow defined criteria could be especially important, Effoduh noted. The Law Commission of Ontario and the Ontario Human Rights Commission recently published an AI impact assessment tool to guide organizations in the assessment of AI systems for compliance with human rights obligations.[114] However, use of this instrument is voluntary. In the United States, legislation introduced in the previous session of Congress would have provided for AI impact assessments, but the bill was never enacted into law.[115] Effoduh emphasized that requiring human rights impact assessments could be a powerful tool; yet, many details, including the scope, approach, and consequences of assessments, remain to be developed. For instance, the law could provide for legally binding debiasing orders, requiring that companies to ensure their data is adequately representative.

---

[109] UN Committee on the Elimination of Racial Discrimination, *General Recommendation No 37: Racial Discrimination in the Enjoyment of the Right to Health*, CERD/C/GC/37 (2024); UN Committee on the Elimination of Racial Discrimination, *General Recommendation No 36: Preventing and Combatting Racial Profiling by Law Enforcement Officials*, CERD/C/GC/36 (17 December 2020), online: <https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/TBSearch.aspx?Lang=en&TreatyID=6&DocTypeID=11>. Office of the United Nations High Commissioner for Human Rights, *Artificial Intelligence and the Rights of Persons with Disabilities: Report of the United Nations High Commissioner for Human Rights*, UN Doc A/HRC/49/52 (March 2021), online: OHCHR <https://www.ohchr.org/en/documents/thematic-reports/ahrc4952-artificial-intelligence-and-rights-persons-disabilities-report>.

[110] World Health Organization, "WHO Outlines Considerations for Regulation of Artificial Intelligence for Health" (19 October 2023), online: WHO <https://www.who.int/news/item/19-10-2023-who-outlines-considerations-for-regulation-of-artificial-intelligence-for-health>.

[111] Office of the Privacy Commissioner of Canada, "Principles for Responsible, Trustworthy and Privacy-Protective Generative AI Technologies" (7 December 2023), online: <https://www.priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/gd_principles_ai/>.

[112] EU Artificial Intelligence Act, "Article 5: Prohibited AI Practices", online (legislation): <https://artificialintelligenceact.eu/article/5/>. See also Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance, *Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance: Report of the Special Rapporteur*, UN Doc A/HRC/56/68 (2024) at para 68(3).

[113] UN Human Rights Council, *Artificial Intelligence Procurement and Deployment: Ensuring Alignment with the Guiding Principles on Business and Human Rights – Report of the Working Group on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises,* UN Doc A/HRC/59/53 (2025), online: <https://docs.un.org/en/A/HRC/59/53>; European Union, *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act),* [2024] OJ, L 168/1, art 86.

[114] Ontario Human Rights Commission & Law Commission of Ontario, *Human Rights AI Impact Assessment* (November 2024), online: <https://www3.ohrc.on.ca/en/human-rights-ai-impact-assessment>.

[115] United States Congress, "S.2892 - Algorithmic Accountability Act of 2023" (21 September 2023), online (legislation): <https://www.congress.gov/bill/118th-congress/senate-bill/2892/all-actions>.

Despite extensive conversations, we have only begun to “scratch the surface” of algorithmic bias. Effoduh argued we should begin by clarifying what *kind* of algorithmic bias we want to address to develop a common language. Specific legal principles are needed to address algorithmic discrimination especially in contexts, like mental health, where AI users may be vulnerable. Impact assessments, along with upskilling and training, will be essential. Effoduh also noted that sometimes “the best AI is no AI” and that the growing interest in AI solutions should not divert funds from community and human-based mental health solutions.

VI. DISCUSSION: EMERGING THEMES AND QUESTIONS

There were multiple opportunities for questions, discussion, and debate throughout the day. These surfaced several overarching themes, indicating possible priorities for future investigation and reform.

1. The need for more genuine lived experience engagement

There was significant discussion of the need to foster genuine patient engagement rather than tokenistic involvement. Participants agreed researchers should include patient partners at the outset so patients can help shape research questions and desired outcomes. As a participant put it, no one “wants to be parachuted in” at the end. Another proposed a “matchmaking” system to connect researchers with individuals who have lived experience, enabling early relationship-building that would allow teams to quickly involve these individuals, even under tight grant timelines.

Participants also suggested earlier and more fundamental means of shifting the terms of patient engagement. They noted that capacity-building among patients (rather than involvement in “one-off” projects) could help strengthen capacity for informed lived-experience consultation. Effective patient engagement should also include critical examination of the structures and theoretical frameworks that inform research; we should be asking what research questions are funded, whether the theoretical frameworks are grounded in lived experience, and what communities are served by the funded research. There are a host of “difficult conversations,” participants noted, the insights from which should be incorporated into research. Yet another perspective related to data access; a participant noted that accessing data for AI development generally requires institutional support. They wondered whether different forms of access might empower persons with lived experience to create their own solutions, and whether some form of partnership could help support responsible data access.

2. There are critical gaps in the regulation of medical devices

Medical device regulation was one of the most prominent themes in the day’s discussions. Participants repeatedly emphasized that the regulations must evolve to address gaps and ambiguities that may be allowing some mental health AI to “fall between the cracks.” They raised questions about the definition of “medical device,” the regulation of ML-enabled MD, challenges relating to bias detection, and how developers interact with regulators, among other issues. They also discussed the Mental Health Commission’s voluntary app assessment framework and how to ensure it can be accessed by smaller developers. Within this overarching concern, several sub-themes emerged.

Limits on federal jurisdiction: Participants expressed significant concern that some high-risk AI tools may fall outside the definition of a “medical device” and Health Canada’s regulatory ambit. For example, a participant referenced an app developed by an Ontario hospital that is intended to identify suicidal ideation in Indigenous children and youth. The group worried this is a potentially high-risk context that should be regulated. Yet, for constitutional reasons, the Health Canada mandate may not extend to AI developed and used exclusively within a single institution; the federal *Medical Devices Regulations* apply to devices *sold or imported* to Canada, seemingly exempting devices used only in-house."[116] These apps could be reviewed under the Mental Health Commission of Canada's voluntary app assessment framework, which the group saw as a positive development. However, participants were troubled by the lack of binding regulation for some tools that are seemingly high-risk and suggested the "regulatory architecture must evolve."

Distinguishing "medical devices" from "wellness" tools: The group was troubled by ambiguity in this area of development, which might create undue risk. Health Canada's 2019 guidance states that "software that does not have a direct impact on the diagnosis, treatment, or management of an individual's disease, disorder, abnormal physical state or symptoms would not be subject to the Regulations."[117] It specifically states that software "intended for maintaining or encouraging a healthy lifestyle, such as general wellness apps" are not regulated medical devices.[118] For example, apps that count calories or encourage exercise are not regulated medical devices even though they may help prevent illness by encouraging a healthy lifestyle. The category changes when the app begins to detect or treat illness; for instance, smartwatches devices that detect atrial fibrillation are regulated medical devices.

Participants were concerned that, while this distinction is clear in theory, it raises potential challenges in the mental health app context. They wondered whether a device's categorization should change when, for example, a person discloses suicidal ideation on a wellness app. Apps can direct people to crisis resources, but some wondered: If they do this, responding in some way rather than "ignoring" the situation, have they entered the realm of medical devices? Participants suggested these worries are pervasive among developers; there is a need to talk about "the harms of these [device] classifications."

Some discussed this issue in terms of incentives, noting that developers may "bend" their claims to fit within the "wellness" sphere to avoid regulatory hurdles. The group discussed how, despite the proliferation of digital mental health apps and tools—and AI-MD in other areas of healthcare—relatively few developers appear to be seeking medical device approval for mental health AI.

The need to address bias: Some participants were troubled by the lack of prescriptive requirements relating to algorithmic bias in Canada's medical device regulations. Health Canada's guidance document, "Pre-market guidance for machine learning-enabled medical devices," calls on developers to consider bias in data selection, data management, and risk analysis; however, the policy is suggestive rather than prescriptive, and it applies only to ML-MD.[119]

---

[116] *Medical Devices Regulations*, SOR/98-282, s 2.

[117] Health Canada, "Guidance Document: Software as a Medical Device (SaMD): Definition and Classification" (10 October 2019), online (pdf): <https://www.canada.ca/en/health-canada/services/drugs-health-products/medical-devices/application-information/guidance-documents/software-medical-device-guidance-document.html>.

[118] *Ibid.* Health Canada's policy was modeled after the U.S. FDA guidance. See: US Food and Drug Administration, "General Wellness: Policy for Low Risk Devices: Guidance for Industry and Food and Drug Administration Staff" (26 September 2019), online: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/general-wellness-policy-low-risk-devices> ("software that is intended 'for maintaining or encouraging a healthy lifestyle and is unrelated to the diagnosis, cure, mitigation, prevention, or treatment of a disease or condition' is not a device" under section 201(h) of the *Federal Food, Drug, and Cosmetic Act*).

[119] *Health Canada Pre-Market Guidance, supra* note 46.

The group also discussed practical challenges. Where it can be difficult or impossible to obtain representative data, they thought the goal should be bias minimization and transparency rather than perfection. Developers should work with Health Canada to demonstrate their attempts to obtain the best possible data and disclose the remaining bias and risk. Indeed, Health Canada already does this to a degree, "pushing" developers on what group and subgroup analyses can be done to demonstrate efficacy and risk in specific subpopulations. They examine and ask questions about developers' clinical data and risk management strategies. Where a device can benefit certain populations, but in others the data is either absent or indicative of lower effectiveness, Health Canada may approve a device but require transparency regarding the limits of safe use.

The participants acknowledged that regulators around the world are grappling with issues of bias detection and mitigation. They encouraged Health Canada to prioritize the development of clear prescriptive requirements relating to data selection and management; bias testing; and reporting of risk and device limits to end users, for instance, through warning labels.

The need for effective monitoring and enforcement: Another theme was concern about how to ensure the necessary oversight of ML-MD that change over time. The group noted that licensure by Health Canada can help "pull the signal from the noise," for instance, of myriad health and wellness tools marketed on app stores. Yet, to be effective in this function there must be adequate medical device surveillance, especially for ML-MD.

The group noted two possible paths for information input to Health Canada. First, it is possible to report adverse events to Health Canada. Such reporting is mandatory for medical device manufacturers, importers, and hospitals.[120] Voluntary reporting is available for members of the public (including clinicians and patients) who wish to flag concerns with medical devices. However, in many cases users and patients will be ill-equipped to detect more complex safety issues such as subtle performance degradation or algorithmic bias that do not present as obvious malfunctions.

Despite these concerns, it is unclear the degree to which Health Canada will pursue a second possible path to information gathering – that is, active auditing or performance inspections of ML-MD. The group stressed the need for monitoring systems to evolve into a genuine "life cycle" approach—beginning at development and extending through ongoing use—particularly given the rapid proliferation of machine-learning-enabled devices. They also expressed concern that some mental-health AI tools are not classified as devices at all, and therefore fall outside what oversight mechanisms do exist, while noting the potential value of the Mental Health Commission of Canada's voluntary Assessment Framework for Mental Health Apps.

Calls for supports, relationship-building, and education: Many called for more education and supports both for smaller AI developers and for the public. Participants recognized that many smaller developers may not have the funds and legal teams to support testing and compliance. Smaller developers may also be intimidated by the medical device regulations, adding to the incentives to remain in the relatively

---

[120] *Medical Devices Regulations*, SOR/98-282, ss 59, 62, made under *Food and Drugs Act*, RSC 1985, c F-27.

unregulated "wellness" sphere. The group suggested there could be more transparency and assurance from Health Canada that developers can access regulatory information and collaboratively address any issues in a way that isn't intimidating. They also emphasized that developers can and should proactively reach out to Health Canada for assistance. For apps that are not medical devices, it will be essential that tools like the MHCC's voluntary Assessment Framework for Mental Health Apps are accessible even to smaller entities.

Several participants also highlighted the need for public education. There is a troubling gap between public perception and reality; people believe there is more oversight than there is. A participant suggested many would be surprised to learn that AI developed within a hospital could be used on them without it having gone through any regulatory pathway. Likewise, many don't understand that "wellness" apps are not regulated by Health Canada. Participants were especially troubled by this reality-perception gap in medical device regulation given that Canada also does not have any dedicated AI legislation.

The group recognized that the mental health AI space is relatively new both for device manufacturers and regulators. Regulators around the world are experiencing similar challenges in developing relevant requirements and procedures for mental health AI, especially with the advent of ML-MD. They emphasized that this is necessarily a work in progress both in Canada and globally.

3. Privacy laws do not sufficiently protect service users

Participants had numerous worries about data and privacy. They were concerned that many health apps do not have strong privacy practices or policies; indeed, there have been numerous reports of companies sharing data with third parties.[121] While data practices might be disclosed in the terms of service, few read these documents. The problem is especially acute for youth. We cannot reasonably expect that adolescents will read the terms of any app or digital tool; yet we continue to assign to app users the responsibility to understand the privacy risks, including how the data will be used, who it might be shared with, and how long it will be retained. Participants considered this manifestly inadequate given that even health professionals may struggle to understand privacy language and principles. The group called for simpler privacy language and broad-based efforts to improve data literacy across ages and demographics, along with heightened transparency and accountability obligations when developers seek to use the data of children and youth.

Other conversations focused on the tensions between privacy and other policy aims. A participant explained that youth sometimes self-censor when they speak with peers over social media and other digital platforms, avoiding language that will flag them as "at risk." There is a tension between allowing youth safe spaces to connect and build community with their peers (which supports better health outcomes) and monitoring for risk. The group discussed how youth mental health organizations like Kids Help Phone are working to allow for data collection

---

[121] See e.g., Quinn Grundy et al, "Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis" (2019) 364 BMJ l920, online: <https://www.bmj.com/content/364/bmj.l920>; Federal Trade Commission, *A Look Behind the Scenes: Examining the Data Practices of Social Media and Video Streaming Services*, (Washington: Federal Trade Commission, 2023).

within systems of youth engagement that are sensitive to *how* and *when* young people want to talk about mental health issues.

The group also discussed data *retention*. While "treatment" may have a "start" and "end" date, some mental health issues are life-long, such that collected personal health information may have ongoing importance for the continuity of care. Yet, technology and healthcare providers may "dump" data, without patient notification or consent, once the minimum retention periods set out in privacy legislation have passed. The group discussed the importance of looking at the terms of service associated with AI-tools. It is essential to understand the policies around data retention and, where necessary, request access before data is destroyed. Yet, once again, there are information asymmetries here; few read, and even fewer understand, the terms of service of a typical consumer product.

4. AI-use raises distinct concerns about consent

Concerns with using AI to support legal outcomes: A key question was whether the consenting process should require specific disclosure of the possibility that AI could support decisions with *legal* consequences, such as involuntary hospitalization under mental health legislation or legally mandated reporting (e.g., of a child in need of protection). These processes do not ordinarily require consent; providers routinely observe and assess patients and in so doing are not required to inform patients about duties under mental health or reporting laws.

Yet, participants questioned whether AI-use changes anything—that is, whether consent *should* be required where AI use could lead to legal consequences with significant rights implications. Some worried that AI-use could exacerbate existing power imbalances, especially in jurisdictions with fewer legal safeguards for patients. For instance, British Columbia law withholds from people who meet involuntary admission criteria any right to capably refuse treatment.[122] In this context of heightened clinical power and coercion, there was a worry that AI-use could lead to more, or more unjust, detentions, especially where the algorithm embeds bias or lacks cultural competency. Indeed, this concern could be reproduced and compounded if that same AI data was also used in tribunal or court reviews of involuntary treatment or detention. The group considered whether, especially in regions with fewer legal safeguards, AI-use risks adding another layer of non-transparency and inaccessibility for patients and clients, exacerbating the pre-existing power imbalance and, potentially, injustice.[123] In other words, while mental health law already does much without consent, arguably AI data has more complexity built into it, such that layering it on top of current legal systems should give us pause.

Skepticism about AI decision-making: Another set of concerns related to skepticism that AI tools will be capable of the very nuanced, contextual analysis that is required in mental health decision-making. A participant emphasized that a clinical decision is "more than an accurate assessment of data"; health providers consider not just a person's direct health condition but also myriad other factors

---

[122] *Mental Health Act*, RSBC 1996, c 288.

[123] British Columbia only recently introduced rights advisors, and many patients in the province remain without legal representation. See Province of British Columbia, "New service helps clarify rights under Mental Health Act" (28 April 2022) online (news release): <https://archive.news.gov.bc.ca/releases/news_releases_2020-2024/2022AG0026-000657.htm?>.

such as their access to housing or degree of family support. The group considered the potential riskiness of relying on algorithms in these contexts where good decision-making incorporates many subtle pieces of information and experience. Where decisions have legal implications, for instance, where there is a child in need of protection, the participants again saw a need for caution; health professionals apply very nuanced professional judgement when a young person reveals signs of abuse.[124] Participants queried whether, if AI is used to support or even replace clinical judgement, legal reporting obligations will be transferred to the AI, and whether we have any confidence that AI will do this well.

5. The need for clearer responsibilities for AI-related harm

There was significant discussion of who should be responsible for harms flowing from health-AI use. The participants were especially concerned about responsibility for harm from AI apps; while the law imposes clear fiduciary duties on regulated health professionals, these don't necessarily transfer to wellness apps operated by private companies. Indeed, it is unclear whether any duties apply in this context (or whether there are public policy reasons to impose them) beyond the contractual rights and responsibilities set out in the terms and conditions. Yet, this reliance on contract is problematic for service users, who may unknowingly be using unsafe technologies, perhaps assuming that health-AI tools are more regulated than they are.

The uncertainty regarding legal duties also creates dilemmas for AI developers. Developers may intend to operate in the wellness space because they lack the capacity to respond to more serious mental health problems. This might cause them to "toe the line," for instance, avoiding learning information about suicidal ideation out of a concern that this will trigger legal duties to respond. Where apps and other AI tools operate under terms and conditions, without established clinical protocols or duties, there are numerous unanswered questions about liability and responsibility in cases of harm.

6. Bias must be addressed on multiple fronts

Algorithmic bias and discrimination were some of the most discussed issues (along with medical device regulation) throughout the day. The discussions highlighted multiple paths for improvement, such as improving global collaboration on data sharing; shifting societal perceptions and priorities; using policy to shape business incentives; creating new laws; and (throughout) recalling that the goal should be improvement rather than perfection.

The need for a global conversation about data: A recurring theme was the difficulty of accessing sufficient data to assess and mitigate bias. Participants noted that developers frequently use datasets, like electronic health records, that reflect biases and inequities relating to race, sex, and ability, among others. There is a concern that relying exclusively on these data sources can lead to discriminatory AI outputs, perpetuating harm. The group called for global conversation and collaboration on AI and data, which could help supply the data that developers need, for instance, to test for race-related biases. These more inclusive approaches are especially critical for countries in the Global South that are buyers of AI developed in other countries.

---

[124] E.g., in Ontario, see: *Child, Youth and Family Services Act, 2017,* SO 2017, ch 14, Sch 1, s 125.

Participants also discussed how data *treatment* may exacerbate algorithmic bias. For instance, the Canadian Institute for Health Information (CIHI) recommends that data collection processes allow for research participants to identify with more than one category of race and ethnicity, recognizing that many people identify with multiple racial, cultural, linguistic (and other) categories.[125] Yet, a participant suggested some data systems are outdated and cannot process such complexity; researchers may simply exclude these more complex data elements from the analysis, which can distort research findings and contribute to inequity. The group noted that many researchers would benefit from guidance—both to better understand the biases in their data and prevent the introduction of additional bias.

Calls for cultural change: Another theme was the need for significant attitudinal shifts for bias to be given policy priority. The group noted considerable societal resistance to effectively tackling bias, for example, in the pervasive misunderstanding of the concept of race; race is a *social* construct, yet some researchers continue to erroneously rely on race as a *genetic* category.[126] Undue emphasis on "racial taxonomies" can divert us from the real problems and tasks – of assessing and addressing the benefits and harms of particular tools in specific cultural environments.

Some participants argued that bias detection and mitigation is "just not a priority" for many AI developers. The large companies that can afford to build AI models like LLMs are duty-bound to their shareholders to increase profit, and the market has not yet demanded the rigorous or systematic reduction of AI-bias in a way that would incentivize companies to take meaningful action. As such, many companies are not openly disclosing how their datasets and models are developed or the bias they might contain. The group saw a need for law; AI companies should be required to systematically test for and mitigate bias. Where mitigation isn't possible, developers should transparently disclose the issues so end users can understand a tool's limits.

Incentives are important but we also need law: Some participants queried whether, instead of "forcing away the bias," we could use policy tools to incentivize bias-reduction. One participant, a developer whose company is adopting a model of transparency, noted the possibility that transparency will create a commercial advantage because purchasers will feel more confident in the product. There could be a shift in buyers' and investors' perspectives on what is "quality AI." The group discussed some improvement, with growing public awareness of the need to better serve some communities, such as women with postpartum depression. Getting public "buy in" could help shift the market towards more bias-aware and equity-enhancing tools. On the other hand, mental health remains a stigmatized area that is often "left behind."

Overall, the participants agreed that market incentives alone will be unable to address bias; there is a need for law. Moreover, existing laws, such as those prohibiting discrimination, are insufficient. The group saw an urgent need for laws that require *pre-market* bias testing,

---

[125] Canadian Institute for Health Information, "Guidance on the Use of Standards for Race-Based and Indigenous Identity Data Collection and Health Reporting in Canada" (17 March 2022) online (pdf): https://www.cihi.ca/sites/default/files/document/guidance-and-standards-for-race-based-and-indigenous-identity-data-en.pdf

[126] Francis S Collins, Michael Morgan & Aristides Patrinos, "The Human Genome Project: lessons from large-scale biology" (2003) 300:5617 Science 286–290.; Theresa M Duello et al, "Race and genetics versus 'race' in genetics" (2021) 9:1 Evol Med Public Health 232–245, online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8604262/>.

disclosure, and mitigation; encourage equitable distribution; and prevent the use of AI for discriminatory reasons. The group recognized that regulation is sometimes seen as a disincentive to entering the field; some are wary of perceived "red tape" and the time and labour that may accompany development within a highly regulated environment such as health. Yet, participants emphasized that we must encourage health AI development *and* regulate to keep it safe and effective. A participant offered an analogy to the Formula One race: Drivers want to win but still pitstop the car along the way, knowing these pauses will ultimately help them go faster. We can take a similar approach to AI development; "those breaks and pitstops will help us win in the end."

The "perfect" versus the "good": Participants called for a flexible and iterative approach to reducing bias, in which developers can work with the regulator to debias the AI where possible, and set out appropriate use-restrictions where bias remains. Health Canada could play key roles, for instance, in issuing debiasing orders or approving AI-MD subject to use-restrictions.

Participants also emphasized the need for frank discussion so people can learn and share best practices, gauge progress, and continuously improve. The objective should be transparency and "bias-awareness," rather than bias eradication, as no AI system will ever be trained on perfectly representative data. Undue focus on the impossible task of eliminating bias may also lead to tokenism—simply "adding some diversity." For example, Google's "Gemini" AI tool has generated inaccurate historical images, depicting America's founding fathers as Black women.[127] Sometimes AI systems are not trained on as much representative data for good reason. Reflexively inserting diversity into models risks distracting us from the key task of identifying and preventing harm.

A focus on understanding and preventing harm could also support the categorical prohibition of some AI-uses. Arguably we should limit or prohibit AI use in some contexts—like the criminal law—where bias is rampant and significantly affects rights. Similarly, we could ask – should AI be *providing* mental health services, or only *helping people* provide services? These are "social, not technological" questions.

Overall, the theme was a call for improvement rather than perfection. We should be asking questions like, what service is a given AI tool intended to provide and how well is it doing? Is it embedded in existing community mental health infrastructures and services? Is it being assessed within a community's cultural norms? instead of talking only about race, we could say, "show me how you're providing a service and what are the outcomes." If they are "making things better and not worse" we can likely look forgive there being "some bias somewhere."

---

[127] CBC, "Google pausing AI portrait tool after complaints of historical inaccuracies" (22 February 2024), online (news article): <https://www.cbc.ca/news/business/google-ai-tool-historical-inaccuracies-portraits-1.7122704>.

VII. THE FOUR QUESTIONS

Near the end of the day participants broke into small groups to return to the four questions posed at the outset:

(1) What is AI's greatest potential for mental health service users?
(2) What are the biggest risks for service users?
(3) Is law adequate to support potential and guard against risks?
(4) What messages should we convey to law and policy makers?

A representative from each group then shared their findings with all participants.

1. Greatest potential

Groups were encouraged by mental health AI's transformative potential across several dimensions. AI could increase access in underserved communities facing geographical, financial, and criminological barriers while providing culturally appropriate services "at scale" without exacerbating provider burnout. Participants saw potential for AI to help people navigate complex healthcare systems, improve service delivery efficiency through patient triaging and clinical documentation support, and enable more personalized prevention, prediction, and care.

2. Biggest risks

Bias and Discrimination: Many participants worried about harm from biased AI, calling for regulation requiring developers to test for, disclose, and mitigate bias. Some emphasized risks of inequitable access to or uptake of mental health AI.

Supplanting Human Judgment: Groups raised concerns about potentially ambiguous boundaries between human and AI decision-making, and confusion about what to do when AI and human recommendations conflict. Participants worried about confirmation bias leading to provider over-reliance on AI. They also noted AI could supplant culturally sensitive human service models, jeopardize those advancements, and foster mistrust.

Improper AI and Data Use: Participants worried about accidental and intentional misuse of data. While manufacturers cannot completely prevent data hacks, this raises concerns about the vast amounts of personal health information in health AI systems. Some groups discussed the risks of intentional misuse, for instance, for insurance or employment denial. They also noted misuse of mental health data could erode public trust and thus the effectiveness of health-AI that requires honest disclose.

Risk of regulatory imbalance: One group identified the risk that overly onerous regulations will create undue barriers for innovators, potentially blocking smaller developers who cannot afford to navigate regulatory pathways. They worried about a "failure to build" resulting in missed opportunities. Another group offered a contrasting view, noting the risk of too-rapid AI development that lacks adequate oversight. For instance, they worried that "wellness" AI tools are quickly coming to market yet exist in a regulatory "grey" zone, creating risk for clients.

3. Gaps in current law

All participants agreed current law and policy have not kept pace with technological innovation and are inadequate. The groups especially raised concerns about privacy law—much of which is over 20 years old—and medical device regulation. They called for a renewed approach to medical device regulation in particular. While this isn't a complete solution, the groups considered effective premarket regulation to be crucial to preventing harm.

4. Messages for law and policy makers

Interdisciplinary Collaboration: Participants emphasized the need for multidisciplinary research and development teams with diverse perspectives to further equity. They specifically called for authentic lived experience representation on committees and boards, suggesting funding requirements could incentivize such collaboration.

Canadian Leadership on AI Medical Device Regulation: Several groups called for Health Canada to lead on AI-enabled medical device regulation, noting the critical need to update prescriptive requirements relating to bias, ML-MD, and unique mental health device safety concerns (e.g., ensuring AI chatbots don't generate harmful outputs). These safeguards must fit within a regulatory approach that allows for post-market surveillance of continuously changing devices. They emphasized that transparency is key and that Health Canada should require developers to disclose biases and mitigation strategies.

Strengthened Privacy Laws: Participants called for updated privacy laws responding to AI-specific issues, for instance, preventing AI companies from selling personal information to private firms for marketing purposes. They suggested Canada risks lagging behind other jurisdiction in its continued reliance on privacy policies and contractual terms and conditions. Governments should prioritize dedicated legislation addressing these concerns.

VIII. CONCLUSION

This workshop brought together diverse stakeholders to examine the legal implications of mental health AI through the lens of compassionate, rights-respecting care. The discussions revealed both tremendous promise and significant challenges that demand urgent attention from policymakers, regulators, and developers.

Mental health AI holds transformative potential to address Canada's significant mental health service gaps. With one in five Canadians experiencing mental health problems at any given time and significant barriers to accessing care, AI tools could enhance access, provide culturally appropriate services at scale, and support more personalized prevention and treatment.[128] Yet realizing this promise requires confronting substantial legal and ethical challenges that current frameworks are unable to fully address.

The workshop identified critical gaps in Canada's regulatory architecture. The government of Canada has boasted that it "was the first country in the world to create a national strategy for AI" in 2017.[129] Canadian actors have indeed provided key leadership in some areas—for instance, the Mental Health Commission of Canada's app assessment framework. Yet, in some respects, Canada's law and policy arguably have not kept pace with innovation. The death of Canada's proposed *Artificial Intelligence and Data Act* leaves Canada without dedicated AI legislation. Medical device regulations impose efficacy and safety requirements for some AI-MD; however, important gaps in that framework may leave some tools—such as wellness apps and AI-MD developed and used within a single institution—underregulated. Indeed, to date, few manufacturers have sought Health Canada approval for mental health related AI-MD. Canada, like regulators around the globe, is also facing the challenge of effectively regulating adaptive ML-MD throughout the life cycle.

Other laws – for instance, relating to privacy, consumer protection, and informed consent – apply in various ways to developers, health providers, and service users. Yet these general-purpose laws are often a poor-fit for AI contexts. For instance, many privacy laws are over twenty years old and arguably unable to fully address the privacy risks of AI use. The fundamental premise that privacy laws need not apply if data is deidentified is now questioned by many who observe the power of AI tools to reidentify individuals. Moreover, in many contexts, for instance involving wellness apps, privacy rights are largely governed by contractual terms and conditions that few read and even fewer understand. This is especially troubling in the case of youth who may be vulnerable and comprise a significant portion of app users.

Other laws and legal principles are intended to protect informed and autonomous choice. These require adequate disclosure of the risks and benefits of any proposed treatment and establish responsibility and liability in cases of inadequate disclosure and patient harm. Yet

---

[128] Canadian Mental Health Association, "Fast Facts about Mental Illness" (2023), online: <https://cmha.ca/brochure/fast-facts-about-mental-illness>.

[129] Innovation, Science and Economic Development Canada, "The Artificial Intelligence and Data Act (AIDA) – Companion document" (31 January 2025), online: <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>; Canadian Institute for Advanced Research, "Pan-Canadian AI Strategy", online: *CIFAR* <https://cifar.ca/ai/>.

these also have unclear application to mental health AI. Questions remain about when clinicians must specifically disclose AI use as part of the contenting process, as well as how they would obtain such consent in the case of opaque algorithms. The issue is even more challenging in the case of direct-to-consumer apps that don't involve a health professional; there is no clear statutory duty in those cases to ensure the app user understands the tool's risks and benefits.

Liability for AI-related harm can be imposed through tort law – for instance, the law of negligence. However, while there are clearly established duties of care on healthcare professionals, these are again more challenging to apply in health-AI contexts. For instance, how does AI use affect the standard of care? At what point will a physician be deemed negligent for relying on—or indeed to *not* relying on—AI? And who, if anyone, has a duty of care in the context of direct-to-consumer AI tools? While there are novel tort law cases proceeding before the U.S. courts, there are currently more questions than answers when it comes to tort liability and health AI.[130] Users of mental health platforms might expect that consumer protection laws would provide a measure of protection. However, mental health apps usually won't reach the minimum threshold of a $50 transaction for Canadian laws to apply. Thus, we are once more in the realm of contractual terms and conditions that generally disadvantage consumers.

The challenge of algorithmic bias emerged as one of the most persistent and complex issues. Bias can enter AI systems at multiple points—including through training data, developer assumptions, and deployment contexts—and can perpetuate and amplify existing healthcare inequities. While some have proposed technical solutions, participants agreed on the need for legally mandated pre-market bias testing, transparent disclosure of limitations, and ongoing monitoring. More fundamentally, there is a need for cultural change, so that developers and purchasers alike perceive a commitment to bias-minimization and equity to be part of what it means to provide "quality" AI. At the same time, the goal should be bias awareness and harm prevention, rather than the impossible task of bias elimination.

Looking forward, we have an opportunity to reflect on the values we want embedded in Canadian innovation and its regulation. Arguably some recent trends—such as the current U.S. administration's rejection of AI regulation—should trouble us.[131] Those approaches apparently prioritize short-term economic gain over protecting the rights, for instance to privacy and non-discrimination, that are essential to human flourishing. Canada has an opportunity to pursue an innovative health-AI agenda that pairs technological advancement with the robust rights protections that are essential to a compassionate, client-centred, healthcare system.

While this workshop only scratched the surface, several policy priorities emerged. Health Canada should prioritize the development of a life-cycle approach to evaluating and monitoring adaptive ML-MD. The law should mandate bias testing and disclosure. Privacy laws require comprehensive updating to address AI-specific challenges while protecting

---

[130] *Garcia v Character Technologies, Inc.*, 2024 No. 6:24-cv-01903-ACC-EJK, (Middle District Florida); See also, *Moffatt v Air Canada*, 2024 BCCRT 149 (CanLII), <https://canlii.ca/t/k2spq>.

[131] See e.g., the discussion by Carolyn Wheeler, "Regulating AI Therapy Chatbots: A Call for Federal Oversight" (2025) 12:2 Texas A&M L Rev 891, online: <https://scholarship.law.tamu.edu/lawreview/vol12/iss2/10>.

vulnerable populations. Throughout, law reform must balance innovation with safeguards that ensure mental health AI enhances rather than undermines service user dignity, autonomy, and equality. To that end, the path forward will require inclusive collaboration. The workshop highlighted the critical need for authentic lived experience and interdisciplinary involvement in AI development and regulation. We need genuine partnerships in setting research priorities, developing theoretical frameworks, and shaping policy responses. Indeed, this workshop—with participants from multiple disciplines and perspectives—was an attempt to model this kind of approach.

The conversations begun in this workshop represent a beginning; they must continue, deepen, and combine with others to ultimately translate into concrete legal and policy reforms. Such sustained commitment can hopefully help us to harness AI's promise while positioning the safeguarding of dignity, equality, and autonomy as central to the notion of progress. We need not achieve perfection; as one participant put it, "the human baseline is far from perfect." But we should ensure that mental health AI represents genuine improvement, embedded within existing community mental health infrastructures, and guided by the values of compassionate, rights-protecting, care.