CIFAR

MACHINE MD:
Law and Ethics of Health-Related AI Case Study 2:

Suicide Artificial Intelligence Prediction Heuristic

Workshop held: March 11, 2022

Report published: March 10, 2023

This report was drafted by Caroline Mercer and Sophie Nunnelley in collaboration with the participants of the Machine MD: Law and Ethics Case Study on the Suicide Artificial Intelligence Prediction Heuristic (SAIPH).

Acknowledgments

This event was co-hosted by CIFAR and the Canadian Institutes of Health Research (CIHR)-funded Machine MD: How Should We Regulate AI in Health Care? project, with support from the Alex Trebek Forum for Dialogue. It is part of CIFAR's AI & Society Program. CIFAR's leadership of the Pan-Canadian AI Strategy is funded by the Government of Canada, with support from Facebook and the RBC Foundation. The organizers thank CIFAR, CIHR, and the Alex Trebek Forum for Dialogue for their support.

Citation

C. Mercer, S. Nunnelley, A. Goldenberg, C. Régis, C. M. Flood, T. Scassa, Z. Kaminsky, J. Chandler, N. Martin, R. Cartagena and the workshop participants, *Machine MD: Law and Ethics of Health-Related A.I. Case Study 2: The Suicide Artificial Intelligence Prediction Heuristic* (Toronto: CIFAR, 2023).

Table of Contents

Law and Ethics Case Studies in Health-Related AI

The Machine MD project is committed to the use of case study analyses to explore the law and ethics of health-related artificial intelligence (AI). The team seeks to identify and analyze the legal issues associated with AI in healthcare by looking at real technologies, identifying any issues they raise, and analyzing how they are treated in Canadian and foreign law. The objective of these case studies is to move beyond abstract concerns into concrete realities, helping to inform law reform with a better understanding of real-world applications. The goal is to support beneficial AI technology innovation, while minimizing associated risks through appropriate legal governance.

The Machine MD team and CIFAR are partnering to host events dedicated to these case studies. Each event assembles an interdisciplinary group of experts in AI, law, ethics, policy, and medicine to discuss regulatory issues raised by a specific AI technology. These events follow earlier AI & Health Care: A Fusion of Law & Science collaborations.[1] This report summarizes the findings of the second of three online case study events in the spring of 2022. The other two events discussed the OR Black Box (March 4, 2022) and "digital twins" technology (April 1, 2022).

[1] See: *AI & Health Care: A Fusion of Law & Science – An Introduction to the Issues*, drafted by Michael Da Silva in collaboration with the participants of the AI & Society workshop for AI & Health Care: A Fusion of Law & Science (Toronto: Canadian Institute for Advanced Research, 2021), online: <https://uploads-ssl.webflow.com/5e94a26db210b579bca67e7c/60b15d4338d77688f3056d12_604140e8419b84275713ca86_CIFAR%20AI%20Report%20(Final).pdf>; *AI & Health Care: A Fusion of Law & Science – Regulation of Medical Devices with AI,* drafted by Michael Da Silva in collaboration with the participants of the second AI & Society workshop for AI & Health Care: A Fusion of Law & Science (Toronto: Canadian Institute for Advanced Research, 2021), online: <https://uploads-ssl.webflow.com/5e94a26db210b579bca67e7c/60b159da764a10e80837a75a_AI-Healthcare-A-Fusion-of-Law-Science-II.pdf>.

Case Study #2: The Suicide Artificial Intelligence Prediction Heuristic (SAIPH)

11 March 2022 (Online via Zoom)

The Suicide Artificial Intelligence Prediction Heuristic (SAIPH) is a natural language processing tool that uses Twitter data to identify suicidal ideation risk. The algorithm analyses speech patterns in public Twitter posts to quickly identify patients who are showing signs of distress. More specifically, it uses neural networks that are trained to recognize concepts including burden, stress, loneliness, hopelessness, insomnia, depression, and anxiety. Although the tool is still in development, it could be used (1) as a clinical decision-making aid in point of care settings (e.g., helping providers determine whether to do suicide screening), (2) for mental health appointment triage (e.g., assessing which patients need to be seen first based on their level of distress), and (3) promoting real-time, application-based, interventions for suicidal outcomes before they develop (e.g., warning Twitter users' social circle of the risk). The algorithm is also capable of analysing archival patterns in social media posts to provide clinicians with a sense of their patients' mood over time. Beyond its eventual clinical use, SAIPH has the potential to be used (4) at a population level, for public health purposes and to track the effectiveness of suicide prevention interventions.[2]

This event examined the potential benefits and challenges of SAIPH through a presentation by one of its developers, commentaries by legal scholars on three legal issues raised by the tool, and breakout sessions where participants sought to better understand – and help resolve – problems.

[2] See: The Royal, "Homegrown innovation in mobile health monitoring" (20 April 2020), online: <https://www.theroyal.ca/news/homegrown-innovation-mobile-health-monitoring>; Arunima Roy et al, "A machine learning approach predicts future risk to suicidal ideation from social media data" (2020) 3 NPJ Digit Med 78

SAIPH (Zachary Kaminsky, The Royal Hospital)

Researcher Zachary Kaminsky, one of the developers of SAIPH, began by describing the unmet need that SAIPH was designed to address. Suicide is a major public health problem, and rates have remained stable for decades.[3] In the hopes of identifying individuals at elevated risk of suicide, Kaminsky turned to social media, and in particular to publicly-available Twitter data. Twitter was selected because Twitter users are informed that the data they generate when using the platform is public.

According to Kaminsky, it's easy to find clues that may indicate suicidality by searching terms like "suicidal, thinking" on Twitter. However, it would take a human researcher a long time to read through enough posts to draw meaningful conclusions. Instead, SAIPH processes Twitter data quickly and efficiently, using neural networks to score psychological constructs that signal expressions of suicidality to varying degrees (e.g., stress, loneliness, anxiety). The model also generates a picture of a user's suicidal risk over time, allowing researchers and clinicians to observe seasonal trends. SAIPH has demonstrated a high level of specificity, although false positives are admittedly relatively common, as with all suicidal ideation research, given the low incidence rate. Kaminsky noted that the tool's predictive abilities are not influenced by gender, although it performs much better for younger users.

Kaminsky discussed a range of possible applications for the technology. These include use in point of care settings to assist with clinical judgment and decision making. Clinicians could evaluate imminent risk and observe trends over time. As well, Kaminsky is working with a start-up company to develop a digital app that would use the data for personalized support. Users would grant confidants in their social circle access to their data, and these contacts could reach out if they noticed or were informed of any distress signals–termed a virtual "SAIPHty-NET." Confidants would then be encouraged to reach out to those at risk well before any suicidal ideation takes hold. One benefit of this approach is that it would be tolerable to false positives – persons who are showing suicidal ideation but who are not in fact at risk of suicide. The tool could also include

[3] Matthew N Nock et al, "Prevalence, Correlates, and Treatment of Lifetime Suicidal Behavior Among Adolescents: Results From the National Comorbidity Survey Replication Adolescent Supplement" (2013) 70:3 JAMA Psychiatry 300.

a messaging function to facilitate communication between users and their peers. It could also target users digitally with resources. Such early interventions could minimize the chances of many serious mental health incidents. Further applications could include triaging patients on wait lists in order to determine which patients are at elevated risk of crisis and should be seen first, and monitoring outpatients and discharged patients in psychiatric settings to ensure that their mental health remains stable. Finally, the technology could be used at a public health or population level to identify increased mental health challenges in geographic areas.

Commentaries

The legal commentaries focused on three issues that have been discussed at previous CIFAR events and that were pre-identified by planners as raising potential issues for SAIPH.

A. Liability (Jennifer Chandler, University of Ottawa)

Jennifer Chandler began by discussing possible liability issues that may arise in the context of SAIPH. She noted that if SAIPH were integrated into the standard of care, healthcare providers could be liable if suicide risk were detected and they failed to intervene in situations where there was a duty of care relationship between the patient and healthcare provider. On the other hand, false positives could also lead to liability issues: clinicians who act overzealously upon detecting suicidality could come up against causes of action including battery and false imprisonment.

Chandler also raised the possibility that the technology could present issues for other parties, such as employers, insurers, and regulatory authorities, if information about employees' suicidality risk was made available to them. The liability risk might be particularly acute in high-risk contexts and occupations such as aviation. Liability questions might also arise in the context of child protection services if suicidality were detected, and the organisation failed to act. Questions about possible liability on the part of the person's confidantes were also raised during the question period. A participant queried whether, for example, a SAIPH user's friends and family could be held accountable for failing to respond to distress signals sent through the "SAIPHty-NET". Chandler discussed the general rule that observers are not obligated to assist individuals in peril under the common law but suggested these tensions may need to be worked out contractually and in terms of standards of reasonableness. Participants also raised the possibility that this issue would be dealt with differently under the Quebec *Charter of Human Rights and Freedoms*[4].

[4] *Charter of Human Rights and Freedoms*, CQLR c C-12

B. Informed Consent (Nyranne Martin, The Ottawa Hospital, with assistance from Jodie Al-Mqbali and Sarah Grieve)

Nyranne Martin provided an overview of informed consent to treatment and to collection, use, and disclosure of personal health information, and to how a consent structure might work for SAIPH. Martin emphasized that the consent structure is complex and depends on the scope of the technology, including whether it is being used for diagnostics or treatment, and whether it is eventually commercialized. While provincial law will always apply to consent to treatment, the scope of technology would determine whether federal or provincial law applies for consent to collection, use, and disclosure of personal health information.

Martin explained informed consent to treatment requires that a patient or their substitute decision maker understand the nature of the treatment, its benefits and risks, possible side effects, alternative treatments, and any consequences of not having the treatment. In the context of SAIPH, patients may be consenting to the use of the technology, as well as to medication and follow ups, as a single package. Consent is an ongoing process, and clinicians would need to inform patients of updates to the technology if the app were to evolve over time.

Given the data collection and artificial intelligence components, Martin added, patients would need to understand how their data was being used. Martin raised related questions for the group to consider, including what specific data will be collected from patients and their close contacts, whether data from private messages would be collected, at what point the app would move from simply collecting data to informing treatment, and how substitute-decision makers might come into play when minors use the technology.

Following Martin's presentation, the group discussed the extent to which consent is necessary when applying machine learning analyses to publicly available social media posts. In medicine generally, consent is not needed to evaluate a patient's physical condition, but the duty to obtain consent arises when the information informs the patient's treatment. The group suggested a consent structure for SAIPH might be laid out similarly, with its use as an assessment tool versus a treatment tool determining the type of consent needed. Finally, the group discussed the possibility that the adverse mental health effects of social media use ("doom scrolling") might be mentioned to patients as a risk of engaging with a technology that relies on active posting.

C. Privacy (Rosario Cartagena, ICES)

Rosario Cartagena delivered a presentation on privacy and cybersecurity laws and their possible applications to SAIPH. First, she explained that data collection generally is governed by a collection limitation principle, which limits collection to what is necessary, lawful, and fair. She noted the particular challenge of applying this principle in the AI context, where the development of machine learning algorithms depends on large amounts of data, rather than minimum collection. She suggested this is one area where law reform may be needed. Cartagena also highlighted other potential data privacy issues, including the risk of re-identification, data quality, and biased data sets. She also noted that AI developers should consider the specific purpose of their data collection, as well as who will have access to the data (noting that SAIPH plans to set up a data trust that will ensure trustee control of patient data, rather than provide for its private ownership or sale).

Cartagena expressed particular concern about data security, noting that cybersecurity threats, including cyberwarfare and cyberespionage, continue to grow. She emphasized the need for start-ups to implement effective cybersecurity infrastructure at the outset of product development. She also discussed regulatory developments that are on the horizon, including a new version of Canada's "Digital Charter", Bill C-11, which was proposed by the last government, and further reforms led by Innovation, Science and Economic Development Canada.[5]

Additional privacy and data security issues that arose during the question period included whether SAIPH data might become part of a patient's medical record and be made available to insurance companies. The group also briefly discussed the principles of Indigenous data sovereignty and how they might apply in this context.[6]

---

[5] Government of Canada, *Canada's Digital Charter: Trust in a Digital World* (Innovation, Science and Economic Development Canada) online: <https://ised-isde.canada.ca/site/innovation-better-canada/en/canadas-digital-charter-trust-digital-world>; Bill C-11, *An Act to enact the Consumer Privacy Protection Act and the Personal Information and Data Protection Tribunal Act and to make consequential and related amendments to other Acts,* 2nd Sess, 43rd Parl, 2020 (first reading completed 17 Nov 2020), online: <https://www.parl.ca/LegisInfo/en/bill/43-2/c-11>. A revised version of the bill has since been introduced, 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 (Introduction and first reading, 16 June 2022), online: <https://www.parl.ca/legisinfo/en/bill/44-1/c-27>.

[6] See: University of Toronto Libraries Research Guides, "Indigenous Data Sovereignty", online: <https://guides.library.utoronto.ca/indigenousstudies/datasovereignty>.

Breakout Sessions

The commentaries were followed by breakout sessions on (1) liability, (2) privacy, and (3) informed consent. Rapporteurs then summarized the findings during a debriefing session. The core thematic concerns that arose in each session are summarized below.

Breakout #1: Informed Consent

Attendees: Lisa Schwartz (Rapporteur), Caroline Mercer (Scribe), Jason Millar, Karyne Vaillant, Florian Martin-Bariteau, Michael Froomkin, Sophie Nunnelley, Sylvain Bédard

The breakout session on informed consent focused on two main themes: (1) the roles and responsibilities of individuals in a patient's "SAIPHty-NET," and (2) the risks and benefits of using the technology.

I. The Roles and Responsibilities of the "SAIPHty Net"

While discussions of informed consent in the context of SAIPH are rightly centred around the patient whose social media activity is being monitored, the group considered that individuals who become members of a person's "SAIPHty-NET" may also take on liability. Several questions were raised in this regard. For instance, the group queried how members of the network might be chosen, particularly where the individual at risk of suicide is already struggling, and perhaps isolated. Moreover, once selected, what kind of consent structure would apply to these individuals? Participants considered whether it would be fair to hold members of the peer group accountable for changes to a person's mental health. The group also raised questions about expectations that might fall on external organizations and third parties, such as colleges or universities, if they were given access to SAIPH data. They considered, for example, whether a faculty should be expected to follow up with a student who was flagged as being at risk of suicide.

II. Explaining the Risks and Benefits of the Technology

The breakout group discussed what a full explanation of the benefits and risks of using SAIPH technology might encompass. Some members raised the possibility that patients will modify their

behaviour after beginning to use the technology. For example, they may start spending more time on social media, which could have unintended negative consequences for their mental health. As well, they discussed the need for patients to be clearly informed of the technology's limitations, and for any unrealistically high expectations to be tempered; for instance, patients should not be told the technology will enable them to jump queues to access healthcare services, even though it may eventually be used for triage purposes. At the same time, the group discussed the possibility that patients will benefit from using the technology, for instance, through better access to resources or treatments. Members also noted the need for informed consent to the collection of data, and the challenge of collecting data for an algorithm the use of which might evolve over time. They discussed the possibility that consent would have to be re-obtained at the point of any significant shift in data use.

Finally, the group noted the likelihood that some people who might benefit from the technology will not have access to the internet or social media. Other individuals with mental health issues may have access but be sceptical about using the app.

Breakout #2: Liability

Attendees: Marc Bilodeau (Rapporteur), Nicole Davidson (Scribe), Colleen Flood, Genevieve Lavertu, Jennifer Chandler, Daniel Buchman, Zach Kaminsky, Ian Stedman, Jennifer Gibson

The breakout group on liability focused on three main themes: (1) product liability and developers' responsibility to mitigate algorithmic risks, (2) liability concerns for healthcare providers, and (3) liability risks for other service providers.

I. Product liability for AI developers

The group began by discussing possible liability risk for AI technology developers. They noted that developers must meet a standard of care when designing and marketing their tools. They are responsible for ensuring their technology is reasonably safe and used for its intended purposes. If a tool is being knowingly misused, engineers may be responsible for adding warning labels to the product or otherwise advising against its use. The breakout group discussed the possibility that liability might differ depending on whether the tool was marketed for suicide prevention or mental health assessment.

II. Liability for healthcare providers

Members of this group also discussed the standard of care that healthcare providers owe to patients. They noted that if SAIPH were integrated into medical practice settings, practitioners would need to use the tool only for its intended purpose, and to evaluate the risks and benefits associated with employing the tool in a given context. The group discussed a possible tension between relying on a patient's self-reported health information – for example, if a patient indicated their risk of suicide was low – and algorithmic reporting, where there is a conflict between the two sources. The group also raised the possibility of automation bias, where clinical decision makers defer to algorithmic tools without employing clinical judgement. For comparison, the group discussed NarxCare, a prescription drug tracking tool used in the U.S. to predict risk of substance abuse or overdose during opioid prescribing. American healthcare providers are required to follow that tool's direction and can lose their medical license for failure to do so. Group members considered that a similar tension between clinical judgement and algorithmic prediction could arise while using tools like SAIPH.[7]

III. Liability for other service providers

The group discussed the possibility that employers or organizations might use SAIPH data inappropriately. For instance, employers could be liable for discrimination based on mental health disability if candidates are not hired because of SAIPH screening. The group discussed the importance of regulation to ensure the tool is only approved in circumstances where its benefits outweigh its risks.

Breakout #3: Privacy

Attendees: Bryan Thomas (Rapporteur), Michael Da Silva (Scribe), Jodie Al-Mqbali, Cécile Bensimon, Rosario Cartagena, Christina Gilman, Gagan Gill, Sarah Grieve

This breakout group focused on SAIPH's implications for data security and privacy, raising important questions about whether and how privacy law should apply. An interesting feature of SAIPH is its use of publicly available social media data to generate highly sensitive predictions about suicidality. The group discussed whether current privacy laws are capable of regulating this

[7] See: Maia Szalavitz, "The Pain Was Unbearable. So Why Did Doctors Turn Her Away?" (11 August 2021), *Wired,* online: <https://www.wired.com/story/opioid-drug-addiction-algorithm-chronic-pain/>.

unique scenario, given that privacy legislation does not typically apply to publicly available data. However, most agreed that federal private sector privacy law will apply in some way; there was debate as to whether these laws apply to the Twitter data itself, but most agreed that the federal private sector privacy law will at least apply to the products of SAIPH analyses, where SAIPH is commercialized. However, participants stressed the importance and difficulty of regulating the production of sensitive information that is derived from analysis of publicly available data (an issue that applies beyond SAIPH).

The group also discussed what information may be collected by SAIPH and potential reidentification risks from the processing of de-identified data. For instance, while the tool does not capture location or age, it may be possible to draw inferences in both categories. At the same time, members noted this data could be useful for public health initiatives. Indeed, some worried privacy law puts undue limits on data collection and thought privacy law should permit access to public health-relevant data in particular. This led to a broader discussion of whether SAIPH's focus on Twitter data raises equity concerns as certain populations are more likely to spend their time on social media.

The group also considered the possible privacy obligations for confidantes in a user's "SAIPHty-NET." While privacy laws would not typically apply to these individuals, the group discussed the possibility of integrating privacy obligations into a consent process for using the tool. Finally, the group noted that SAIPH developers are considering creating a data trust and discussed whether this is the best way to protect privacy. Members noted the term 'data trust' admits of many specifications but thought using a data trust does appear preferable to many of the alternatives, for instance, in its ability to limit possible conflicts of interest. On the other hand, they considered that trusts may place overly onerous duties on fiduciaries.

Conclusion

This case study highlighted both the transformational potential of using AI in healthcare, and the regulatory challenges and tensions that arise. The themes raised during the presentations and breakout sessions on SAIPH included:

- Liability risks for healthcare providers using the technology, for instance, for acting on false positives
- The risk of automation bias if healthcare providers begin to rely on the technology over their own clinical judgement
- The roles and responsibilities of members of a user's "SAIPHty-NET," including the kind of informed consent structure that should apply to these members and the associated risk of liability
- Liability for third parties including employers, educational institutions, child protective services, and community organizations
- Data governance and privacy issues including risk of re-identification, data quality, and algorithmic bias
- The difficulties of applying privacy regulations when publicly available data is used to create sensitive, individual risk profiles, and the need for nuanced legislation
- The regulatory and liability distinctions that arise depending on whether the tool is used for diagnostics or treatment, and whether it is marketed as a suicide prevention or mental health tool
- The privacy and regulatory requirements triggered by commercialization
- Tensions in applying data collection principles when collecting data for AI algorithms that require large amounts of data
- The value and utility of data trusts

This list is non-exhaustive. Some concerns were unique to particular breakout sessions. However, discussions regarding (i) the roles and responsibilities of "SAIPHty-NET" members, (ii) privacy and data security, (iii) the need to consider both the benefits and risks of the technology, (iv) risks for healthcare providers, and (v) implications for third parties, arose across the breakout groups. Participants especially emphasized certain regulatory tensions, for instance, relating to the applicability of Canadian privacy law to the collection of publicly available information; the need for clear liability and consent structures for individuals who agree to participate in a user's "SAIPHty-NET"; and the variability of liability concerns for healthcare providers when the tool is used for diagnostics versus treatment. They also discussed the importance of exploring the utility and implications of using data trusts to store the personal health information that informs AI algorithms.

Some answers and clarity regarding these issues may emerge as the SAIPH technology is further developed and commercialized. However, such developments may also raise new regulatory questions. These concerns and questions highlight the need for further discussion of the unique regulatory issues that arise in the context of AI in healthcare.

Workshop Participants:[8]

Caroline Mercer	Gagan Gill	Pascal Thibeault
Michael Da Silva	Jodie Al-Mqbali	Jacqui Sullivan
Colleen Flood	Karyne Vaillant	Tess Sheldon
Zach Kaminsky	Lisa Schwartz	Anna Goldenberg
Bryan Thomas	Marc Bilodeau	Nyranne Martin
Christina Gilman	Michael Froomkin	Robert Bacigalupo
Cécile Bensimon	Nicole Davidson	Louise Bernier
Florian Martin-Bariteau	Rosario Cartagena	Devin Singh
Ian Stedman	Sarah Grieve	Camille Brosseau
Jason Millar	Sylvain Bédard	Sophie Nunnelley
Jennifer Gibson	Genevieve Lavertu	Catherine Régis

[8] The following people participated in the workshop but were not part of a breakout session: Jacqui Sullivan; Tess Sheldon; Pascal Thibeault; Anna Goldenberg; Nyranne Martin; Robert Bacigalupo; Louise Bernier; Devin Singh; Camille Brosseau; Catherine Régis.

CIFAR

cifar.ca/ai