Artificial Intelligence Systems: Island Health Research Ethics Guide

April 2024

Image generated with DeepAI

Island Health Research Ethics & Compliance Office
Queen Alexandra Centre for Children's Health
2400 Arbutus Road | Victoria, BC V8N 1V7 Canada
on Songhees and Esquimalt First Nations' Territory

Tel: (250) 519-6726
ResearchEthics@islandhealth.ca
https://www.islandhealth.ca/research-capacity-building/research-ethics-compliance-office

Table of Contents

1.0 What are artificial intelligence (AI) systems and automated decision systems?

An **artificial intelligence system** is a technological system that, autonomously or partly autonomously, processes data related to human activities by a genetic algorithm, a neural network, machine learning or another technique to generate content or make decisions, recommendations, or predictions (The Digital Charter Implementation Act, 2022). A decision system is **automated** when a technology that *assists* or *replaces* the judgment of human decision-makers does so through the use of a rules-based system, regression analysis, predictive analytics, machine learning, deep learning, a neural network, or any other technique.

2.0 Is your AI related work research or quality improvement (QI)?

Island Health is a recipient of Tri-Agency funding from the Government of Canada. As a condition of funding, the Agencies require that researchers and their institutions apply the ethical principles and the articles of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS: 2022) and be guided by the Application sections of its articles. As per 4.3 of the Agreement on the Administration of Agency Grants and Awards by Research Institutions, Island Health must ensure that research conducted under their auspices complies with TCPS2 as well as ensure, through the use of financial or other controls, that the Institution's research ethics board ("REB"), or an REB designated by the Institution, has approved the research project before research activities involving humans have commenced, and that REB approval is maintained as long as activities involving humans are carried out.

Article 2.5 Quality assurance and quality improvement studies, program evaluation activities, and performance reviews, or testing within normal educational requirements when used exclusively for assessment, management or improvement purposes, do not constitute research for the purposes of this Policy, and do not fall within the scope of REB review.

Article 2.5 of TCPS2 (p. 21) stipulates that quality improvement initiatives fall outside of REB review. However, the research ethics board makes the final decision on what is exempt from research ethics review and so is responsible for the demarcation of quality improvement and research within an institution (TCPS2, p. 16).

Article 2.6 of TCPS2 (p.21) stipulates that creative practice activities, processes through which artist(s) make or interpret a work or works of art, also do not require REB review. As the distinction between research and creative practice is not always clear, the final assessment of whether an activity is research is the responsibility of the REB (TCPS2 Interpretations, March 2024, p. 57).

In order to evaluate whether a project is research or quality improvement, Island Health's QI Ethics Team has developed a QI Screening Tool to classify and assess the risks of quality improvement projects. If you are embarking on a project that uses AI, please either consult the Research Ethics & Compliance office or submit your project to the QI Ethics Team using the QI Screening Tool. **Please note that the generated score and classification received from this screening tool does not constitute an official exemption from REB review. For exemptions, please contact the Research Ethics & Compliance Office.**

To publish findings pertaining to human data or biological materials, most publications require proof that the researchers or project leads underwent an ethical evaluation of some kind. Often, publishers will require a certificate of research ethics board approval and/or an exemption letter prior to agreeing to publish. **The REB does not provide retroactive approvals.** Ethical approval to conduct research with humans is required prior to

undertaking any research activities.

Labeling a project as quality improvement (QI) is only appropriate when the benefits and risks of the interventions are already known. This is because QI is typically undertaken to improve something already known to be beneficial (Lynn, 2007). For most AI in health care settings, the risks and benefits of the AI systems are not known (McCradden et al., 2020; 2021). Even if something has proven safe in one environment, it may not necessarily be safe in another. The current consensus is that if a particular system has not been first tested in a live clinical environment with patients, there may be unknown variables that could produce harm.

Grounding AI systems in the research ethics pathway achieves two crucial goals for our organization:
(1) A pathway for progress in clinical settings becomes clearer;
(2) Trust is fostered by aligning AI system use with established legal and ethical standards in clinical research.
This approach has already been implemented at *The Hospital for Sick Children* in Toronto, Canada.

3.0 Generative AI & consent

Many generative AI tools like ChatGPT are built by large language models (LLMs) and multimodal foundation models (MFMs) that process and output visual or linguistic information for users. When given written prompts, LLMs will generate human-like responses to carry on text-based conversations with users. Many are fast, free, and easy to use. In research, these tools can help researchers identify research questions and hypotheses, design studies, write code, and analyze research data.

- While generative AI tools have many uses, they are **not** to be used to process *personally identifiable research data* in Canada without consent. This applies to information that is directly identifiable, such as a person's date of birth, address, full name, personal health number (PHN), as well as any indirectly identifiable information such as spoken or written data of individuals describing themselves, their beliefs, or their opinions in their own words, and/or photographs and video.

- Qualitative textual data, for example, can be used to predict mental illness diagnoses and other personality traits (Zhang et al., 2022; Jang, J., et al. 2022). When combined with authorship identification algorithms, capturing a sample of someone's speech or writing can be easily re-identified across the web, even when explicit identifiers are not collected. Like many kinds of personally identifiable information, there can be legal, economic, political, social, and dignitary harms if disclosed.

*Recommendation: De-identify your data **before** using AI tools for analysis and use secure, encrypted platforms for data processing.* For more information on de-identification, please visit this Privacy Implementation Notice from the Government of Canada.

- Because the full scope of applications, uses, and risks of each individual generative AI tool may not be presently known, researchers cannot always guarantee the confidentiality or the welfare of their participants when they are used to process their data. Accordingly, participants must be told about any limitations of privacy and explicitly consent to having their data inputted into these AI systems before researchers do so.

- If researchers want to use a generative AI tool in their study, meeting the ethical standard of autonomy requires that participants know and consent to sharing their data with the specific generative AI tool in advance. The intention to share research data with a specific tool or to use it in data analysis should be on the informed consent form (ICF), alongside the country where the data holdings are located (e.g., Canada,

U.S.A, the EU). This way, the prospective participants can gauge which privacy laws may apply to this data before they decide whether to consent to share.

- If researchers were using ChatGPT in a research study, here are some examples of what they may include on the ICF for prospective participants:
    - "This research uses ChatGPT, owned by OpenAI, as a tool for data analysis. OpenAI is owned and operated in the United States. As such, there is a possibility that your information may be accessed without your knowledge or consent by the U.S. government, in compliance with the USA Freedom Act (formally known as the Patriot Act)."
    - "**Limit of Withdrawal:** It is impracticable and even impossible in some cases to remove any data we have collected about you once it has been de-identified and shared with OpenAI."
- As of February 2024, OpenAI is currently under investigation by the Office of the Privacy Commissioner of Canada for unlawful collection and disclosure of personal information without consent. It is also under investigation by the United States Federal Trade Commission for unfair and deceptive practices relating to risks of harm to consumers. It is the responsibility of researchers to stay informed about the legal and ethical standing of the AI tools they plan to use, including regular checks for updates on investigations or changes in terms of service that may affect participant's rights and welfare.
- ***[Recommended 02/23/24]*** If ChatGPT is to be sought as an aspect of a study's data analysis, please make **a Privacy Request** to OpenAI to opt-out of your data being used for training purposes, and upload the response to the relevant REB for review. This restricts any inputted research data from being used to train the ChatGPT algorithm. If you are using **OpenAI Playground**, inputted data may already be opted out from algorithmic training. If using a different tool, check out the developer's website to see if they also have a similar opt-out process to help protect participant data.

4.0 Machine learning-enabled medical devices (MLMDs)

Health Canada's definition of an **machine learning-enabled medical devices (MLMD)**: A medical device that uses machine learning, in part or in whole, to achieve its intended medical purpose.

If your research involves the use of a Health Canada defined MLMD, you will need to present to the REB its Class designation. MLMDs can be labeled as Class II, III and IV medical devices under the current regulations. The class designation of an MLMD is determined based on the safety and risk of the device. For more information on how to determine which class your device falls under, please consult Special Rules 13 to 16 from Health Canada.

4.1 Predetermined change control plans (PCCPs)

- The use of a PCCP allows timely and ongoing management of risks while retaining high regulatory standards to ensure device safety and effectiveness.
- A predetermined change control plan (hereafter "PCCP") is the documentation intended to characterize a device and its bounds, the intended changes to the machine learning (ML) system, the protocol for change management and the change impacts. If included, a PCCP is considered part of the device design.
- If you are submitting an ethics application involving a MLMD, you will need to submit a separate PCCP or add a section in your clinical protocol for it.
- PCCPs should be risk-based and supported by evidence, take a total product lifecycle perspective, and provide a high degree of transparency.

- All modifications listed in a PCCP must ensure that the device continues to operate within its intended use. **Changes listed in a PCCP should not include changes to the medical conditions, purposes, or uses of an MLMD**. **Such changes require a medical device license amendment application prior to implementation.**
- The content of a PCCP is threefold: (1) Change description; (2) Change protocol; (3) Impact Assessment.
- For more information on how to present the content of a PCCP, please visit: Draft guidance document: Pre-market guidance for machine learning-enabled medical devices.
- For more information on PCCPs, please consult the following document from Health Canada: Predetermined change control plans for machine learning-enabled medical devices: Guiding principles.

4.2 Clinical and behavioural research protocol requirements

- If you are submitting a research proposal that involves the use of an MLMD, you will need to provide to the REB information pertaining to the device.
- Your protocol should include sections on:
    - Design
    - Device description (Identify whether a Class 1, 2, 3, or 4 medical device)
    - 3-part predetermined change control plan (PCCP)
    - Risk management
    - Data selection and management
    - Descriptions of development, training, and tuning approaches
    - Testing/evaluation/software verification
    - Clinical validation
    - Transparency
    - Labelling
    - Terms and conditions (T&Cs)
    - Health Canada license status
    - How the pre-market requirements for medical device cybersecurity will be met
- For more information on each aspect to include in your protocol, please consult the Pre-Market Guidance for MLMDs by Health Canada.
- The REB will evaluate MLMDs in alignment with the Guiding Principles of Good Machine Learning Practice as jointly identified by the U.S. Food and Drug Administration (FDA), Health Canada, and the United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA). Broadly, these are:
    - Multi-Disciplinary expertise is leveraged throughout the total product life cycle
    - Good software engineering and security practices are implemented
    - Clinical study participants and data sets are representative of the intended patient population
    - Training data sets are independent of test sets
    - Selected reference datasets are based upon best available methods
    - Model design is tailored to available data and reflects the intended use of the device
    - Focus is placed on the performance of the human-AI team
    - Testing demonstrates device performance during clinically relevant conditions
    - Users are provided clear, essential information
    - Deployed models are monitored for performance and re-training risks are managed

Over the lifecycle of the MLMD, manufacturers and researchers are encouraged to apply a framework for designing AI and data-driven research such as Sex-and-Gender-Based Analysis (SGBA Plus) and consider the

unique anatomical, physiological, and identity characteristics of patients, participants, and system users. This includes considering intersectional identity variables, collecting, and analyzing disaggregated data on sub-populations in clinical studies, and transparently presenting training and test data, as appropriate. Please visit this website for a free course on Sex-and-Gender-Based-Analysis from the Government of Canada.

5.0 Software as a medical device (SaMD)

Mobile health apps and new software applications may meet the definition of software as a medical device (hereafter SaMD). Though we often think of medical devices as clunky pieces of hardware, like pacemakers, eyeglasses, and magnetic resonance imaging (MRI) scanners, anything that intervenes on a patient's state within a healthcare context can be construed as a medical device, such as but not limited to software or apps facilitating cognitive behaviour therapy (CBT), dieting, exercise, smoking cessation, medication compliance, and sleep monitoring.

5.1 Does your software meet Health Canada's definition?

- If a software application has a **direct impact on the diagnosis, treatment, or management of an individual's disease, disorder, abnormal physical state, or symptom(s)**, it meets Health Canada's definition of software as a medical device (SaMD).
- Collecting or monitoring health data, even if used with clinicians, is still considered to directly impact a medical patient's management. SaMDs may be capable of running on mobile phones on their own or in combination with other products including medical devices.
- Applications that are excluded as falling under Health Canada's definition of a SaMD are:
    - Software intended for administrative support of a healthcare facility,
    - Software that enables clinical communication and workflow including patient registration, scheduling visits, voice calling, video calling,
    - Software intended for maintaining or encouraging a healthy lifestyle, such as general wellness apps, and
    - Software intended to serve as electronic patient records or tools to allow a patient to access their personal health information.
- When evaluating whether software qualifies as a SaMD according to Health Canada's definition, the deployment of AI—whether on the edge or in the cloud—plays a crucial role. Edge computing, where data processing happens directly on the device, can offer benefits in privacy and real-time analysis without needing constant internet access. In contrast, cloud-based AI leverages significant computational power and extensive datasets, potentially enhancing the software's capabilities. This distinction is vital for regulatory considerations, impacting risk assessments and privacy measures, and thus influencing the classification under Health Canada's SaMD framework.
- Identifying whether software qualifies as a SaMD can become complex with AI advancements. An illustrative grey area involves AI applications that transition from wellness advice to personalized health management based on user data, such as dietary apps evolving to manage specific health conditions like diabetes. This shift from general wellness to direct disease management underlines the nuanced categorization challenges AI introduces into the SaMD framework.
- If your team thinks that the software used in your research study is **not** a SaMD, please explain this to the REB by showing how it meets all four criteria as set out by the FDA and Health Canada.

5.2 The four criteria set out by the FDA and Health Canada

1. Software that is not intended to acquire, process, or analyze a medical image or signal from an in vitro diagnostic device (IVDD) or a pattern/signal from a signal acquisition system.
2. Software intended to display, analyze, or print medical information about a patient or other medical information (such as demographic information, drug labeling, clinical guidelines, studies, or recommendations).
3. Software that is only intended to support a healthcare professional, patient, or non-healthcare professional caregiver in making decisions about prevention, diagnosis, or treatment of a disease or condition. *Software that is used to treat, diagnose, or drive clinical management **does not** fit under this criterion.
4. Software that is not intended to replace the clinical judgment of a healthcare professional to make a clinical diagnosis or treatment decision regarding an individual patient.

Tip: Both #3 and #4 exclusions make a distinction between an **immediate and near-term action** and one that occurs after **independent review and reliance on a clinician's own judgment to reach a recommendation without primarily relying on the software function**. Medical contexts such as high-risk surgery or emergency room care may not provide the adequate space needed for reflection to determine when the information provided by the software will be acted upon in clinical care. When seeking an exclusion on these grounds, be sure to provide the REB with enough information about the contexts in which the software will be deployed to align an assessment with Health Canada's.

5.3 Clinical Decision Support Software (CDS) VS Patient Decision Support Software (PDS)

- Please clarify for the REB whether a SaMD in your research can be broadly categorized under the terms **Clinical Decision Support Software** (CDS) or **Patient Decision Support Software** (PDS).
- CDS software is usually intended for health care providers (HCPs), whereas PDS software is intended for patients and caregivers who are not HCPs.
- These distinctions are primarily made based on the *intended* users of the software. Just because a graduate student may be using a software does not change its intended user – these are aspects of a software's design.
- If you or your study team is unsure of how to classify the software being used in your research, reaching out to ask the developers or manufacturers is a good first step.
- For more information, please consult Health Canada's Guidance Document: Software as a Medical Device (SaMD): Definition and Classification.

5.4 Research Ethics Board Requirements

- Researchers to show familiarity with the Health Canada regulations surrounding SaMD.
- Researchers to supply the REB with enough information for them to independently assess whether the AI system deployed in their research study falls under the definition of a SaMD.
- Researchers are ready to respond to questions from REB reviewers and delegates about any SaMDs used in the context of their research.
- If your SaMD is through a mobile app, please consult our Mobile App Checklist Document, as the REB may require this in order to conduct their review. This document includes questions on what kinds of data participants will share, the terms of service or any end user license agreements (EULAs) involved, the use of

any unique device identifiers (UDIDs), the market status of the app and any data monetization or sharing plans.

- If you have applied for a Health Canada license, please upload your application as an attachment to your research ethics application.
- When collecting personally identifiable data from research participants, all future, secondary use of research data is subject to TCPS2 Article 3.13, which states that when seeking consent for a specific research project at the same time as seeking consent for storage of data and human biological materials for future unspecified research, prospective participants must be provided with an option to consent to each separately, either through separate consent forms or separate sections on the same form.

Article 3.13: To seek broad consent for the storage and future unspecified use of data and human biological materials, researchers shall provide prospective participants, or authorized third parties, with applicable information as set out in Articles 3.2 and 12.2, as well as the following details, as appropriate to the research project:

a. the type, identifiability, and amount of data and human biological materials being collected and stored for re-use, and for what potential purpose;
b. the voluntariness of the participant's consent, including any limitations on the feasibility of withdrawal;
c. a general description of the nature and types of future research that may be conducted, including whether the research might be conducted outside of Canada (if known);
d. the risks and potential benefits of storage of data and human biological materials, and of their use in future unspecified research, including areas of uncertainty where risks cannot be estimated;
e. access to a general description of the repository and its governance;
f. a statement regarding participants' preference to being recontacted for additional future research;
g. whether the data or human biological materials could be shared with researchers who are not subject to the TCPS;
h. whether the research will (if known) or might include whole genome sequencing or similar technologies that may pose a substantial risk of re-identification of the participant or identification of material incidental findings (when appropriate);
i. whether linkage of data gathered in the research or derived from human biological materials with other data about participants – either contained in public or personal records – is anticipated (Article 5.3); and
j. separate options for consenting to participate in a specific research project and for consenting to the storage of data and human biological materials for future unspecified research.

Please ensure that sections referring to the prospective use of research data to train an AI algorithm and/or sell to a third party for similar such uses is in its own separate section on the consent form alongside a separate consent agreement/signature or is given its own separate consent form.

6.0 Requesting health data to train an external AI system

In general, all requests for data go through Decision Support at Island Health. There are a number of resources available to internal employees through the Decision Support website. External requests can start by sending an e-mail to DataRequest@islandhealth.ca.

For internal Island Health employees, there is:

- the *Report PORTAL* (data associated with improving the health system for quality improvement purposes only. This data is **not** available for research purposes).

- *The Enterprise Data Warehouse*. This contains a metadata catalogue of data offerings available.

Additional data sources that may be of interest are:

- Population Health Statistics & Publications (Island Health)
- Health Data Platform BC (Government of British Columbia)
- BC Community Health Data (BCCDC via the Provincial Health Services Authority)

7.0 Clinical Deployment Requirements

Prior to being deployed in clinical settings with real-world patients, it is important that each distinct AI system, product, or device, undergo three phases of testing at the intended site with the intended population. As the very same AI system can have different impacts such as harms at different sites with different patient populations, it is important that each system be tested prior to deviating from the standard of care. This creates a safer and more ethical clinical environment and safeguards researchers and clinicians from legal liability in the case of unintended harms. For more information on legal liability, please see 10.6 of this document.

Researchers can submit a clinical research ethics application for Stage 1 and later add Stage 2 and Stage 3 in subsequent post-approval activities (PAAs). Even if a research ethics application has been approved by another research ethics board for Phases 1 and 2, a specific Stage 2 trial will be required for REB review. By developing a phased integration framework for AI systems, emphasizing pilot testing and evaluation against established standards of care before full deployment, we can ensure that AI systems are aligned with clinical expectations and legal requirements from the outset.

> **Stage 1**: Exploratory AI systems are governed by data-access protocols. This is important to protect patient information during this exploratory phase. Data used in testing should reflect the diversity of the intended patient population to help identify and mitigate biases in the AI system. Data security experts are engaged to enhance the protocols' effectiveness and implement robust monitoring mechanisms to continuously assess the AI system's performance and safety after deployment, allowing for timely adjustments as needed.
>
> **Stage 2**: Silent period (the model is run in 'silent mode', where its predictions do not influence decision-making) and trialed against real patient encounters to establish its clinical performance, simulating its intended integration into clinical care.
>
> **Stage 3**: Clinical trial that establishes a particular AI model's causal impact on relevant outcomes. Established trial designs (e.g., standard randomized controlled trials, stepped wedge, adaptive platform, cluster, micro-randomization) can accommodate both rigorous evaluation and ethical requirements for research conduct.

If interested in undertaking a clinical research study involving the use of an AI system, please consider consulting with the Research Ethics & Compliance office before submitting your application to receive support.

8.0 Clinical Trial Requirements

The CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) extension is a new reporting guideline for clinical trials evaluating interventions with an AI component. It was developed in parallel with its companion statement for clinical trial protocols: SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence). Both CONSORT-AI and SPIRIT-AI offer checklists that researchers can

use to consider the unique aspects of AI to include in their protocols to the REB.

9.0 Data Management Plans

Increasingly within the research community, data management plans are encouraged by funding agencies and research ethics boards to exhibit the stages of data collection, storage, and dissemination during a research project.

- For studies involving an AI system, it is important that a data management plan is appended or added to the primary protocol that includes information about the following:
- Method and sources of data used to train the model.
- Description of any privacy impact assessments (PIA) and conclusions
- Confirmation that the diversity in the data sources meet the needs of the study design and procedures to ensure equitable selection.
- Description of the study team's framework of ongoing monitoring of AI-driven decisions to prevent bias, discriminatory or unjust impacts.
- Description of what features of data will be used at which stage of the model's development.
- A description of what will happen to the data when the specific project is complete.
- The extent to which participants have control over sharing the data from them and generated about them.
- Description of limitations of privacy and confidentiality.
- Data use and Terms of Use/Service (ToU/ToS) requirements of third-party sources, if applicable.
- Data linkage procedures, if applicable, including consent, methods of combining new data with existent datasets.
- Describes any reasonably foreseeable purposes in which participant data may be used in the future.
- Describes a clear plan for data deletion or anonymization after project completion.

10.0 Ethical considerations that may be considered by the Island Health REBs

10.1 Algorithmic harm

- An algorithmic bias can be thought of as a systematic deviation in algorithm output, performance, or impact, relative to some norm of standard. An algorithm can be morally, statistically, or socially bias, depending on the normative standard used.
- Not all statistically biased behaviours are ethically or morally problematic, while not all statistically fair or unbiased predictions are ethically or morally acceptable. Just because a bias has been identified in an AI system's outputs does not necessarily entail an ethical concern. (Fazelpour & Danks, 2021).
- Researchers using AI are expected to understand how AI systems may produce biases, be equipped to understand the impact of various kinds of biases and be willing to develop mitigation plans with their teams.
- During an ethical review, the REB may request information about how biases are managed with respect to the AI training data, testing conditions, and data management plans.
- Algorithmic bias is not merely a function of the mathematical values or codes embedded within its program, but is also influenced by the domain, scope, and range of its applications, the goals to which it is directed, and a myriad of other variables that change depending on the context.
- While it may not be possible for AI developers and users to anticipate every algorithmic bias, occlusion, or error, the REB expects researchers to show value weight transparency and pre-emptive accountability in their research design. We encourage researchers to involve domain experts, ethicists, and affected

communities early in the AI development process to provide comprehensive perspectives on potential biases and their implications.

- **When algorithmic bias goes unchecked, unjust harms can be produced.** Even when protected attributes like race, gender, and nationality are explicitly removed from an algorithm, other variables in the input data can be correlated with and serve as proxies for these protected attributes. As a result, merely removing certain attributes is an insufficient strategy of addressing this kind of bias. Increasingly researchers are turning to tools that make use of standardized bias assessment frameworks or advanced debiasing techniques to monitor and mitigate biases.
- Because the goal of clinical trials in particular is to arrive at statistically unbiased estimate of the average effect of a treatment under study, the REB may put more scrutiny on randomized control trials (RCTs) using AI to consider algorithmic bias in the reporting of results as well as any ongoing adaptations to trial design (for more, see TCPS2 Chapter 11).

10.2 Clinical equipoise and duty of care

- TCPS2 (2022) stipulates that clinicians have a fiduciary responsibility to their patients to not deny them sub-standard treatment to participate in research. This can come into conflict with another one of TCPS2's requirements of clinical equipoise, an ethical criterion of uncertainty within the relevant expert community about which interventions are most effective for a given condition (p. 199). Though this expert disagreement and uncertainty creates the need for research to determine the comparative therapeutic merits of different interventions, this research should not create conditions where participants are denied therapies that are demonstrably safe and effective.
- Crucially, it may not be justified to carry out a research study if clinical equipoise cannot be demonstrated at the relevant level. While equipoise can be claimed to exist due to data-driven research, some focus only on the local clinical environment in which the research proposes to take place.
- In the context of research involving AI systems, the REB expects researchers to be prepared to exhibit an examination of the extant literature in the domain of interest and to be able to provide evidence-based justifications for why clinical equipoise is posited and site-specific understandings of local standards of care when deviations are expected.

10.2 Therapeutic misconception

- Therapeutic misconception occurs when research participants do not understand that research is aimed primarily at producing knowledge and that their participation may not provide any therapeutic benefit to them. This can occur when the possible benefits are overstated, the possible risks are underestimated or omitted, or when there is a lack of disclosure about how participation can interfere with their own health care objectives.
- It is particularly important when clinician-researchers (clinicians who also conduct research) are managing the conflicts that may arise from their dual role that participants fully appreciate the difference between their clinical care and research participation. Because AI systems have been shown to be capable of making great strides in medical innovation, it is important for research participants in studies that feature the use of an AI system to understand the risks, benefits, and impacts on their welfare clearly so they can offer fully informed consent.

10.3 Privacy and information security

- As discussed in TCPS2 Chapter 5, researchers as individuals and as members of organizations owe participants an ethical duty of confidentiality to safeguard entrusted information. This includes an obligation to protect information from unauthorized access, use, disclosure, modification, loss, or theft.
- Many artificial intelligence systems and machine learning algorithms involve third parties, such as the manufacturer, developer, or technicians. It is important for researchers to understand the terms of using such devices and how the information that researchers share is managed.
- Some artificial intelligence systems, be they software or conventional medical devices, will require a privacy impact assessment (PIA) before they can be deployed in the organization. This may be the case irrespective of whether the system will be deployed in a capacity to serve clinical care, administration, or research.

If you are an Island Health employee and are thinking about using artificial intelligence in your research project, it may require a consultation with the Information Stewardship, Access & Privacy (ISAP). ISAP is responsible for ensuring that our organization follows the provincial and federal privacy legislation. To learn more about consultations and privacy impact assessments, visit the Intranet website here.

If you are external to Island Health but need information on whether an initiative, innovation, product, or process requires a privacy review, please reach out to privacy@islandhealth.ca.

10.4 Synthetic Data

Please be patient while we learn more about synthetic datasets and how researchers can use them to protect participant and patient data.

10.5 Liability and research related harm

- According to the Artificial Intelligence and Civil Liability Project Committee's 2023 Consultation Paper from the BC Law Institute (BCLI), it is anticipated that in the case of civil wrongs where the harm is unintended, alleged **negligence** will predominate in tort litigation concerning artificial intelligence. Based on tort law, compensation for harm caused by AI depends on some human or corporate entity being legally liable to compensate the person harmed. As such, researchers should be mindful of the requirements of the tort of negligence:
    1. The defendant owed the plaintiff a duty of care.
    2. The defendant breached the standard of care.
    3. The plaintiff incurred damage.
    4. The breach of the standard of care by the defendant was the cause of the damage.

- Whether a defendant owed the plaintiff a duty of care depends on whether (1) a **relationship of proximity** existed between them, and (2) whether harm to the plaintiff was **reasonably foreseeable** if the defendant failed to take reasonable care.
- The first concept, 'a relationship of proximity' is one in which the plaintiff could be "directly affected" by the defendant's conduct, such that the defendant should have the plaintiff's interests in contemplation. Most clinician-patient relationships will meet this definition. Further, to perform a duty of care, the defendant must meet the standard of care. Specifically, "the court will ask itself the question: is it more probable than not that the damage would not have occurred but for the breach of the standard of care?" (BCLI, 2023, p.25).
- When first deployed, many AI systems in healthcare will initially be breaches of the standards of care. As such, recognizing the known standards of care and the grounds for clinical equipoise that prompted the

research question in the region, site, and context of the research will be important grounds for evaluating the legal risks of researchers and participants.

- The second concept, "reasonable foreseeability of harm," is positioned between mere possibility and probability—it only applies to risks that can reasonably be contemplated and prevented from materializing through the defendant's actions or omissions. This clarifies to whom a duty of care is owed, what that duty can reasonably demand of its provider, and when the harm is too remote to be attributable.
- However, there are also many legal challenges surrounding AI which complicates how fault can be assigned in cases of AI-related injuries. Some AI systems lack transparency around their decisions; in these cases, if no source of error can be identified, it may not always be possible to assign blame to a particular defendant. The more autonomous an AI system, the more challenging it may be to prove any causal link between an identifiable act or omission by a human and an output of an AI system that causes harm.
- A common disclaimer for research participants during the consent process is that by agreeing to participate in this research, they are not giving up or waiving any legal rights if they are harmed during the research. However, their ability to have their legal rights upheld may be constrained when an AI system is involved in research in BC. It may be important for researchers to undergo a legal and/or risk assessment to weigh how much legal risk exists for themselves and their participants and to inform their participants proportionately during the consent process.

10.6 Citing the use of AI

If using data derived from generative AI tools in your research (such as graphs, summations of literature reviews, bibliographies, spreadsheets), you will want to disclose the use of AI in your research. This will encourage transparency and help guide your scientific colleagues to understand your results and how to reproduce them. To do this, specify the AI tools, models, and versions used in your research in the methods section of your manuscript.

For more guidance on how to cite AI tool use in research, please see How to cite text generated with A.I., Scholarly use of A.I. tools, University of Victoria Libraries.

10.6 First Nations Health Data, OCAP®, and **C.A.R.E principles**

Increasingly and urgently, frameworks such as the First Nations Principles of OCAP® and the C.A.R.E. Principles for Indigenous Data Governance urge non-First Nation individuals, institutions, and governments to ask themselves if they have the authority or consent to make data on First Nations peoples available for surveillance, research, and commercial purposes. While these principles and initiatives have many features in common, they are not pan-Indigenous and cannot be used as 'rules' that will apply to each and every instance of use. Each First Nation may choose whether to assert these principles or not and in what context, requiring researchers to anticipate assertions of OCAP® (ownership; control; access; possession) in the form of data repatriation, intellectual property, and copyright.

It is the onus is on the researchers to justify their use and collection of First Nations data to the REB, who as per Chapter 9 of the TCPS2 (p. 146), may require ongoing engagement and consultation with the community prior to commencing the research as well as a data management plan that includes a means of making that data available to First Nations communities should assertions of data sovereignty be made on the holdings in the future.

It is important to recognize that First Nations have called for reform and action on deficiencies of current privacy legislation, such as PIPEDA, to represent their rights and interests. As such, it is important to consider and consult

with each Nation and to avoid using Indigenous data in artificial intelligence systems or devices without community consent.

We recommend researchers take the Fundamentals of OCAP Course and to acquaint themselves with the resources offered by the First Nations Information Governance Centre (FNIGC)'s Online Library as a starting point.

Island Health sponsors the training of all Island Health employees who wish to enroll in the San'yas Anti-Racism Indigenous Cultural Safety Training Program. For more information on how to register, please visit this Island Health bulletin.

If you are unsure whether your project impacts First Nations individuals or communities, please reach out to the Research Ethics and Compliance office for a consultation.

11.0 Contacts

Research Ethics & Compliance Office	ResearchEthics@islandhealth.ca	General Inquiries
Institutional Approvals	ResearchOperations@islandhealth.ca	For studies accessing Island Health resources and sites.
Information Stewardship and Privacy Office	Privacy@islandhealth.ca	Privacy Consults
Innovation, Analytics & Information	Ai@islandhealth.ca	AI-related questions

12.0 Consultants

The following persons have contributed to this guidance.

Victoria Philibert	Victoria.Philibert@islandhealth.ca	HREB Coordinator, Research Ethics & Compliance, Island Health
Sam Obeidat	Sam.Obeidat@islandhealth.ca	Chief AI Officer at WAIU Island Health HREB Member
Tracy Wong	Tracy.Wong@islandhealth.ca	Research Quality Assurance Specialist, Research Ethics & Compliance, Island Health
E. Sarah Bennett	Elizabeth.Bennett@islandhealth.ca	Manager, Research Ethics & Compliance, Island Health
Dr. Paul Hasselback	Paul.Hasselback@islandhealth.ca	Island Health HREB Member
Dr. Louise Costello	Louise.Costello@islandhealth.ca	Island Health HREB Member
Dr. Jan Boxall	Jan.Boxall@islandhealth.ca	Island Health HREB Member
Dr. Laura Shanner	Laura.Shanner@islandhealth.ca	Island Health HREB Member
Michael Litchfield	Michael.Litchfield@islandhealth.ca	Director, Business Law Clinic, University of Victoria; Island Health HREB Member

13.0 External Resources

Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS 2 (2022).
TCPS 2022: Interpretations (Last updated: March 2024)
The Digital Charter Implementation Act, 2022
the Office of the Privacy Commissioner of Canada
the United States Federal Trade Commission
How to cite text generated with A.I., Scholarly use of A.I. tools, University of Victoria Libraries
Health Canada: Special Rules 13 to 16
Health Canada: Draft guidance document: Pre-market guidance for machine learning-enabled medical devices
Health Canada: Predetermined change control plans for machine learning-enabled medical devices: Guiding principles
Health Canada: 3-part predetermined change control plan (PCCP)
Health Canada: the Pre-Market Guidance for MLMDs
Health Canada: the Guiding Principles of Good Machine Learning Practice
Sex-and-Gender-Based-Analysis from the Government of Canada
Medical License four criteria as set out by the FDA and Health Canada
Health Canada: Software as a Medical Device (SaMD): Definition and Classification.
Government of Canada: Privacy Implementation Notice: De-identification
OpenAI's privacy policy

14.0 References

(2020) Agreement on the Administration of Agency Grants and Awards by Research Institutions. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada. Government of Canada.

Artificial Intelligence and Civil Liability Project Committee. (2023). Consultation Paper on Artificial Intelligence and Civil Liability. British Columbia Law Institute. Vancouver, BC. <https://www.bcli.org/project/artificial-intelligence-and-civil-liability-project/> Accessed 01.08.24.

Bell, G., Burgess, J., Thomas, J., and Sadiq, S. (2023). Rapid Response Information Report: Generative AI - language models (LLMs) and multimodal foundation models (MFMs). *Australian Council of Learned Academies.*

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 Session, 44th Parliament, 2021-2022.

CBC News. (2023) More Canadian privacy authorities investigating ChatGPT's use of personal information. May 25, 2023. <https://www.cbc.ca/news/canada/british-columbia/canada-privacy-investigation-chatgpt-1.6854468> Accessed 27-Feb-2024.

Campbell et al. (2022). Open Science and Data Sharing in Trauma Research: Developing a Trauma-Informed Protocol for Archiving Sensitive Qualitative Data. *American Psychological Association.* https://doi.org/10.1037/tra0001358

Fazelpour, Sina & Danks, David. (2021). Algorithmic bias: senses, sources, solutions. *Philosophy Compass*, 16(8). https://doi.org/10.1111/phc3.12760

Genin, K., & Grote, T. (2021). Randomized Controlled Trials in Medical AI: A Methodological Critique. *Philosophy of Medicine*, *2*(1). https://doi.org/10.5195/pom.2021.27

Gilbert J-P, Ng V, Niu J, Rees EE. (2020). A call for an ethical framework when using social media data for artificial intelligence applications in public health research. *Can Commun Dis Rep*, 46(6), 169–73. https://doi.org/10.14745/ccdr.v46i06a03

Health Canada. (2015). Guidance Document – Guidance on the Risk-based Classification System for Non-In Vitro Diagnostic Devices (non-IVDDs). Guidance documents – Medical devices. Government of Canada. < https://www.canada.ca/en/health-canada/services/drugs-health-products/medical-devices/application-information/guidance-documents/guidance-document-guidance-risk-based-classification-system-non-vitro-diagnostic.html> Accessed 27-Feb-24.

Jang, J., Yoon, J., Son, G., Kang, M., Choeh, J. Y. (2022). Predicting Personality and Psychological Distress Using Natural Language Processing: A Study Protocol. *Frontiers in Psychology*, 13. DOI=10.3389/fpsyg.2022.865541

Kang, Cecilia & Metz, Cade. (2023). F.T.C. Opens Investigation Into ChatGPT Maker Over Technology's Potential Harms. The New York Times. July 13, 2023. <https://www.nytimes.com/2023/07/13/technology/chatgpt-investigation-ftc-openai.html> Accessed 27-Feb-2024.

Lynn, J., et al. (2007). The ethics of using quality improvement methods in healthcare. *Ann Intern Med*, 1:146(9), 666-73. doi: 10.7326/0003-4819-146-9-200705010-00155. PMID: 17438310.

McCradden, M.D., Stephenson, E.A. & Anderson, J.A. (2020). Clinical research underlies ethical integration of healthcare artificial intelligence. *Nat Med,* 26, 1325–1326. https://doi.org/10.1038/s41591-020-1035-9

(2023) OpenAI Privacy Request Portal. January 12, 2024. Open AI. <https://privacy.openai.com/policies> Accessed 27-Feb-2024.

(2023) Privacy Implementation Notice 2023-01: De-identification. Access to Information and Privacy Notices. Treasury Board of Canada Secretariat. Government of Canada.

(2022) Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS 2 (2022). Panel on Research Ethics. Government of Canada.

Zanger-Tishler, M., Nyarko, J., Goel, S. (2023). Risk Scores, Label Bias, and Everything but the Kitchen Sink. arXiv. https://arxiv.org/abs/2305.12638

Zhang, T., Scoene, A. M., Shaoxiong, J., Ananiadou, S. (2022). Natural language processing applied to mental illness detection: a narrative review. *npj Digital Medicine*, 5(46). https://doi.org/10.1038/s41746-022-00589-7