PHSA AI Research Toolkit: A resource to support responsible AI research at PHSA

Purpose of the Toolkit

The PHSA AI Research Toolkit aims to support responsible AI research practices across the AI research lifecycle by highlighting available resources, infrastructure and best practices that are accessible to all researchers across PHSA. Responsible AI research entails developing AI systems that consider patients' safety, privacy, and fairness. Ultimately, the goal of this resource is to support research and development of healthcare AI systems that align with the Pan-Canadian AI4 Health Guiding Principles.

Who is the toolkit for?

This resource is intended to be used by anyone engaging in research at a PHSA program or research institute.

What is included in the toolkit?

In this toolkit you will find links to infrastructure available through PHSA that can support AI research and useful guidelines and frameworks, including Researcher Guidance on AI bias and transparency. This toolkit is intended to complement the material on the AI in Research Website, including the AI Research Lifecycle.

What is NOT included in the toolkit?

The toolkit is meant to be a living document. It does not include all possible resources available to researchers. It also does not include resources accessible through individual research or clinical programs at PHSA.

An expanded interactive version is available to anyone with a PHSA account here: AI Research Resource Bank. We encourage anyone conducting AI research at PHSA to review, suggest edits and additions to this toolkit based on their own experience.

Additional support through the PHSA AI in Research Working Group

The toolkit was developed by the AI in Research Working Group. This group is comprised of experts in AI research and partners in research support and compliance groups such as the Research Ethics Boards, Privacy, Security and Data Governance. Researchers are invited to connect with the AI in research Working group for pre-REB review and support in the development of their research projects.

For further questions related to AI research or information in this toolkit, please reach out to AIWorkingGroup@phsa.ca.

Provincial Health Services Authority

Section 1: Data Resources

RESOURCE TITLE	DESCRIPTION
PHSA Researcher Data Access, Security, and Privacy site	Resources to help clarify privacy considerations for research, and help researchers meet ethical, regulatory and institutional requirements as well as the requirements of funders. Also provides overview of data access request processes for PHSA datasets.
PHSA PANDA Data Portal	List of linked datasets available in PANDA with information on requesting data access.
First Nations Information Governance Center	Organization that focuses on First Nations data governance, maintains database of First Nations related data that can be accessed at no cost, provides training and resources for OCAP data governance principles.
STANDING together recommendations	Outlines a list of guidelines for ethical data use specific to healthcare.

Section 2: Computational Resources

RESOURCE TITLE	DESCRIPTION
UBC Advanced Research Computing **(ARC)**	High performance computational resources for UBC Researchers, with experts that can be consulted for specific domains. They also offer training and proposal support.
Digital Research Alliance of Canada **ARC**	ARC platform hosted by Digital Research Alliance of Canada.
PDHIS Compute resources	Contact the Digital Innovations team at PHSA for information on accessing PHSA instances of Azure and AWS cloud services for AI research.

Section 3: Regulatory approvals and compliance

RESOURCE TITLE	DESCRIPTION
BC Cancer REB Policies, Procedures, and Guidance	Links various BC Cancer REB policies and required checklist for submission of any research project including AI/ML at BC Cancer.
CW Research Ethics Board Artificial Intelligence/ Machine Learning Application Submission Checklist	Checklist required when submitting for REB approval for an AI/Machine learning project at BC Children's Hospital or BC Women's Hospital.
PHSA Data Governance Framework	This document defines the PHSA Data Governance Framework, establishes an approach for uses of data beyond traditional silos, and sets expectations for how

	PHSA will work with its health sector partners to improve access to data, including for research.
CIHI Standards for Race-based and Indigenous identity data collection and health reporting in Canada	Standards for collecting race-based and identifying data in healthcare, and guidelines for appropriate use of data. PHSA endorsed.
Good Machine Learning Practice for Medical Device Development: Guiding Principles	Set of 10 principles developed jointly by FDA, Health Canada, and UK MHPR specific to machine learning in healthcare.
Health Canada Pre-market guidance on Machine Learning Enabled Medical Devices	Guidance for regulatory standards for medical devices that use ML, in part or in whole, to achieve their intended medical purpose, known as machine learning-enabled medical devices (MLMD).

Section 4: Guidance documents, research frameworks and reporting standards

RESOURCE TITLE	DESCRIPTION
EQUATOR Network reporting guidelines	Best practice guidelines for publication or research. Many guidelines have been updated to specify standards for reporting research related to AI/ML interventions in healthcare, such as CONSORT-AI, DECIDE-AI and TRIPOD-AI.
A Clinical Trial Design Approach to Auditing Language Models in Healthcare Setting	Framework for model auditing developed by BC Cancer researchers for LLMs in healthcare context.
RE-AIM and PRISM implementation and sustainment frameworks	Commonly used evaluation frameworks for healthcare interventions.
Artificial intelligence self-efficacy: Scale development and validation	Validated scale for assessing self-efficacy of AI end-users.
Post-deployment evaluation framework to guide implementation of AI systems into healthcare settings	Post-deployment evaluation framework considering model performance, utility and integration/adoption in healthcare setting.

PHSA Researcher Guidance: Bias and transparency in AI research

The table below provides considerations and actions researchers should take throughout the AI research lifecycle to align with the Pan Canadian AI4Health Principles in their work[1]. Below the table is a selection of additional articles, tools and trainings that provide an overview of AI bias considerations and actions and discussion of key opportunities and challenges.

Table 2. Considerations and actions to identify and mitigate risk of bias and increase transparency during development of predictive AI tools for healthcare and decision support

PHASE 1	
Consideration	Action
Inclusivity in data: Is the data being used for model development inclusive of all population groups impacted by its use? Are demographic variables included to allow analysis of model performance among intersecting demographic characteristics such as sex/gender identity/race/ethnicity?	• Identify all target population groups that should benefit from your model. • Review your dataset for demographic distribution across target population groups to ensure all are represented as you would expect based on the population of interest. • Consider methods to boost representation of marginalized population groups in your data through simulation/synthetic data creation. • If your dataset does not include sufficient demographic variables to assess inclusivity of the population then evaluation of outcomes in Phase 2 and 3 across demographic groups is critical.
Inclusivity in data: Was the data collected in a fair and equitable way?	• Review information about methods of recruitment and data collection for the dataset being used to build your model. Is there any opportunity for systematic/ sampling bias? Ie from measurement tools, people collecting the data, or recruitment methods. • If the data was not collected for the purpose of model development, it may need significant pre-processing to be usable for this purpose. Consider where bias may be introduced in the data processing step based on assumptions being made by the analyst.

Provincial Health Services Authority

Design/developer bias: When people are involved in deciding what data is used to develop a model and how this data is used, there is an opportunity to introduce individual biases into the developed model. Researchers should consider, is the team building the model diverse and representative of the target end-users? What unconscious bias may be introduced by the team developing the model?	• Include people in the research team that have subject matter expertise and lived experience related to the clinical/operational topic you are researching. • If working with marginalized/vulnerable/medically complex populations, make sure you have representatives from these groups involved in the design of the model. • At the start of the project, have team members reflect on their positionality and potential biases that they may bring to the project.
Algorithmic bias: When data used to develop a model is limited or lacks representation of some user groups or conditions, models are at risk of supporting decisions/ conclusions that are not fair. This can lead to further increasing health disparities when the model is applied in the healthcare system. Researchers should consider, how demographic variables are represented in the model and if they are being used in appropriate ways? Are you able to determine if predictions/model outcomes are biased towards/against one population group?	• Where possible, apply Explainable AI (XAI) methods such as feature importance analysis, model visualization, natural language explanations, counterfactual explanations, SHAP (SHapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations), rule-based models, decision trees, and attention mechanisms in deep learning to understand how the developed model reached conclusions/predictions. • When evaluating model performance, include sub-group assessments to measure if predictions/outcomes are distributed among identified target population groups as expected.
Transparent data governance: Have you established transparent and structured data governance frameworks that specify ethical use, access controls, and management of data within the planned AI system from the start?	• Data governance frameworks that include information on all data sources used to train and operate the AI model should be clearly documented for future users. This should include a description of methods of data collection, pre-processing, storage and use. • A clear data governance framework ensures accountability and aligns work with regulatory requirements and organizational values of transparency.

	• Any project engaging with Indigenous communities and/or data sources should adhere to OCAP principles for data governance. • If you are developing AI for a medical device, adhere to Health Canada regulations for good engineering practices.
PHASE 2	
Output bias in context: Once the model is deployed within the clinical context intended, performance can be impacted by differences in population from the training dataset.	• Data collected in this phase must include all demographics and user characteristics needed to determine if outcomes are fair and equitably distributed among the population affected by the model use. • Evaluate model output for bias by assessing if model outputs are distributed among intersecting population groups as expected. • Consider if model use in context is introducing unexpected biased outcomes. • Model recalibration and updating may be required to ensure performance is maintained within the new population.
Human in the loop: Using AI tools outside of their intended context can lead to biased and inaccurate results. Have you planned out ways to maintain human oversight as the model is integrated into care and ensure appropriate use?	• Develop education resources that clarify how end users should be interacting with the model and its output. • Document terms of use and ensure end-users agree to abide by them before use. • Develop audit protocols to monitor use once deployed.
PHASE 3	
Model Drift: When a model is applied to a new population, differences in underlying characteristics may lead to reduced accuracy and performance—especially if the model relies on those differing factors to make predictions. Have you considered how the deployment context differs from the original dataset used for model training/development?	• Compare characteristics of the sample being used to evaluate the model to the original dataset characteristics used to train the model to determine if significant differences exist that may impact model performance. • Embed a process of model updating and recalibration within the model deployment plans that can help maintain performance of the model over time.

Post-deployment monitoring: Have you planned for long-term monitoring of model outcomes and performance once the research has been completed?	• This monitoring process should allow for ongoing review of model outcome bias, model accuracy, and appropriate use of model. • Evaluation and post-deployment monitoring should include evaluation across diverse and intersecting population groups.

Resources for Further Learning and Development in AI Equity and Bias

The following resources are a curated list of articles, tools and trainings that provide an overview of AI bias considerations and actions and discussion of key opportunities and challenges.

Peer Reviewed Papers

Aiken, C., Flann, S., Longstaff, H., Manusha, S., Pavlovich, S., Scott, J., & Wright, J. (2021). *A guidance for novel ethics of privacy issues associated with artificial intelligence in the public sector research domain*. http://www.phsa.ca/researcher/Documents/AI%20Guidance_FINAL.pdf

Celi, L.A., Cellini, J., Charpignon, M.-L., Dee, E.C., Dernoncourt, F., Eber, R., Mitchell, W.G., Moukheiber, L., Schirmer, J., Situ, J., Paguio, J., Park, J., Wawira, J.G., & Yao, S. (2022). *Sources of bias in artificial intelligence that perpetuate healthcare disparities — A global review*. PLOS Digital Health, 1(3), e0000022. https://doi.org/10.1371/journal.pdig.0000022Feng, J.,

Lieng, M.K., & Zhan, A. (2022). *Clinical artificial intelligence quality improvement: Towards continual monitoring and updating of AI algorithms in healthcare*. NPJ Digital Medicine, 5(1), 66. https://doi.org/10.1038/s41746-022-00622-0

Gichoya, J.W., Banerjee, I., McDonald, C.J., & Kohli, M.D. (2023). *AI pitfalls and what not to do: Mitigating bias in AI*. British Journal of Radiology, 96(1150), 20230023. https://doi.org/10.1259/bjr.20230023

Griffen, Z., & Owens, K. (2024). *From "human in the loop" to a participatory system of governance for AI in healthcare*. American Journal of Bioethics, 24(9), 81–83. https://doi.org/10.1080/15265161.2024.2245529

Modise, L.M., Avanaki, M.A., Ameen, S., Celi, L.A., Chen, V.X.Y., Cordes, A., Elmore, M., Fiske, A., Gallifant, J., Hayes, M., et al. (2025). *Introducing the Team Card: Enhancing governance for medical Artificial Intelligence (AI) systems in the age of complexity*. PLOS Digital Health, 3(3), e0000495. https://doi.org/10.1371/journal.pdig.0000495

Nazer, L.H., Soltwisch, M., Beasley, J., Horowitz, J.M., Campbell, J.R., & Lester, C.A. (2023). *Bias in artificial intelligence algorithms and recommendations for mitigation*. PLOS Digital Health, 2(6), e0000278. https://doi.org/10.1371/journal.pdig.0000278

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). *Dissecting racial bias in an algorithm used to manage the health of populations*. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

Xu, J., Xiao, Y., Wang, W.H., Ning, Y., Shenkman, E.A., Bian, J., & Wang, F. (2022). *Algorithmic fairness in computational medicine*. EBioMedicine, 84, 104250. https://doi.org/10.1016/j.ebiom.2022.104250

Last Updated: 13-MAR-2025

Acknowledgements: We would like to acknowledge the PHSA AI Research Working Group members (listed in alphabetic order): Aditi Bhardwaj, Akshdeep (Ash) Sandhu, Ali Taqdir, Angel Arnaout, Anna Low, Anna Meredith, Beth Payne, Benjamin Wan, Christian Schutz, Elizabeth Stacy, Farin Meralli, Haley Foladare, Hind Sbihi, Ihoghosa (Muyi) Iyamu, Jessica Gagliardi, Jessica Morrice, Jonas Bambi, Jonath Sujan, Jonathan Wong, Kacey Dalzell, Kaylie Choi, Kendall Ho, Kathryn Dewar, Lisa Tang, Matthias Görges, Navnoor Chhina, Rishav Singh, Sally Chen, Simon Roome, Sonia Brar, Tonia Nicholls, Tibor van Rooij, Timothy (Tim) Bhatnagar, Vedant Bahel, Victoria Philibert, William Wang.