

# A Practical Framework for Operationalizing Responsible and Equitable AI in Healthcare

**Prepared By:** Mattea Welch, Benjamin Grant, and Christopher Deutschman

**In Consultation With:** Clare McElcheran, Adam Badzynski, Jennifer A.H. Bell, Andrew Hope, Robert C. Grant, Tran Truong, Kelly Lane, Patti Leake, Divya Sharma, Ian Stedman, Laleh Seyyed-Kalantari, Mike Lovas, Jeremy Petch, Benjamin Haibe-Kains, and James A. Anderson

## Motivation/Background

The next healthcare era hinges on artificial intelligence (AI) advancements, leveraging vast clinical, imaging and operational data to streamline and augment existing processes. However, the data that is used to train AI solutions is biased due to the inequities that characterize our society (e.g. the social determinants of health), our health, and healthcare in particular. Therefore, there is a risk that AI will exacerbate health inequity by recapitulating and magnifying biases in healthcare data<sup>5-7</sup>, thereby undermining clinicians' ability to provide safe and compassionate care.

This document outlines best practices and recommendations to 1) encourage responsible, safe, and compassionate AI development and deployment within UHN clinics and 2) safeguard UHN patients and care-providers from biased and inequitable AI predictions.

## Document Overview

Many comprehensive frameworks have been developed to guide AI in healthcare. This document aims to present a single overarching best-practices guide based upon the synthesis of three specific frameworks: 1) HEAAL([Kim et al. 2024](#)); 2) JustEFab([Mccradden et al. 2023](#)); and 3) Normative Framework([McCraadden et al. 2023](#)). An appendix has also been developed to provide further details and recommendations for different steps in this guide. The appendix is meant to be updated as new approaches become validated and accepted standards.

This document is most relevant for individuals or teams within healthcare focused on AI solutions that will have an impact clinically; either through direct clinical deployment or by generating or interpreting data that will later be used clinically. It specifically covers questions and steps that should be considered during: 1) the conceptualization of a clinical problem that may benefit from AI; 2) developing and/or testing AI solutions in a research or commercial environment; 3) deploying AI solutions in clinic through silent or prospective modes; or 4) maintaining and auditing existing clinical AI solutions. Although this document has a deployment focus, AI projects aiming to perform research on healthcare biases, fairness and inequity would also benefit; partially from the earlier steps of the framework.

Furthermore, it is recognized that all steps of this document may not be relevant or feasible for every project. However, it is recommended that the best efforts be made to achieve as many steps as possible.

## Best-Practices Guidelines

Stage	Task
<b>Problem Identification and Study Design (Pre-Data)</b>	
Define the Problem and the Solution	Clearly state the problem you aim to solve and its clinical relevance
	Justify why your team is well-positioned for this task
	Assemble a cross-functional team for consultation and guidance
	Explore non-technical solutions alongside technical ones. If there are viable non-AI solutions to this problem, start there
Focus on Equity in the Problem Space	Conduct a literature search to understand social determinants of health, and historical injustices specifically affecting target population
	Document baseline inequities in target population
	Identify and map intersecting identities and social factors contributing to inequities in the target population
	Determine if the AI solution will involve First Nations or Indigenous data; if so, take appropriate measures to respect data sovereignty
	Determine if the AI solution will involve other underserved populations; if so, seek appropriate educational resources and advocacy groups for guidance
Outcome Measurements and Data Requirements	Assess how outputs will be integrated into clinical settings and the value of the solution
	Determine what outcome should be measured and ensure the outcome is measurable and fair across target subpopulations
	Identify the necessary data for model development and verify its availability
<b>Model Training and Development (Retrospective Data)</b>	
Accessing Retrospective Data	Obtain REB or QI approval for using retrospective data
	Request retrospective data from relevant institutional data repositories
Appropriateness of Retrospective Data	Identify current clinical benchmark solutions
	Recruit and interview patients from disadvantaged subgroups to identify where current clinical benchmarks fall short
	Examine whether your data is fit for purpose and assess data for disparities that may skew performance (e.g., differential missingness, imbalance populations, etc.)
Defining Objectives and Metrics	Choose appropriate statistical measures of performance
	Define equity objectives and fairness metrics considering both discriminatory performance and inequitable impact between subgroups
	Specify the analytical plan, including handling intersectional impacts and statistical penalties
Model Training and Testing	Assess the representativeness and quality of training and testing data
	Train the model, seeking evidence of safety, efficacy, and equity
Model Evaluation	Assess whether the model meets defined statistical measures of performance, equity objectives and fairness metrics.

	<p>If the model fails to meet any of the defined statistical measures of performance, equity objectives and fairness metrics, consult with the cross-functional team to determine if the model should be retrained or if alternate solutions should be explored.</p>
<b>Silent Deployment and Clinical Trial (Prospective Data)</b>	
Prospective Deployment Preparation	<p>Obtain REB or QI approval for prospective silent deployment of AI model</p> <p>Map retrospective data to prospective data, if necessary</p> <p>Ensure prospective data is representative of retrospective training and testing datasets. Consider spectrum effects and changes in disparities</p> <p>Develop required workflow integration components that carefully consider necessary human factor and user experience measures</p>
Prospective Model Evaluation (Silent Mode)	<p>Prospectively deploy AI model with no clinical intervention</p> <p>Compare prospective model performance with retrospective model performance to ensure that prospective model continues to meet defined equity objectives and fairness metrics</p> <p>If the model fails to meet any of the defined statistical measures of performance, equity objectives and fairness metrics, consult with the cross-functional team to determine if the model should be adapted or if mitigation strategies should be implemented</p> <p>Review results of silent deployment with cross-functional team and define auditing methods for future deployments that track known biases in the solution</p>
Prospective Clinical Trial of AI Solution	<p>Obtain REB or QI approval for conducting a prospective interventional clinical trial to test AI solution</p> <p>Develop educational materials for end-users that includes known AI solution biases and automation bias</p> <p>Undertake prospective clinical trial of AI solution with diverse representation of participants and clinical end-users</p> <p>Is your trial population representative of your silent deployment prospective dataset? Consider spectrum effects and changes in disparities</p> <p>Evaluate AI workflow integration through qualitative assessment with end-users</p> <p>Evaluate trial AI solution performance and compare with retrospective and silent prospective model performance. Additionally, assess how often the AI's recommendation is correctly executed in the clinic.</p> <p>If the AI solution fails to meet any of the defined statistical measures of performance, equity objectives and fairness metrics, consult with the cross-functional team to determine if the solution should be adapted or decommissioned</p> <p>Compare the AI solution performance metrics to clinical benchmarks and review results with cross-functional team</p> <p>Recruit end-users and patients to provide critical perspectives to inform the future clinical integration of this AI solution</p>
<b>Operationalization and Lifecycle Monitoring (Clinical Care)</b>	
Preparation and Documentation	<p>Seek institutional approval for AI solution rollout</p>

	<p>Prepare documentation on model development, statistical performance, fairness metrics, and equity objectives</p> <p>Include information on non-technical components supporting equity goals</p>
Communication and Education	<p>Modify educational materials for a broader audience including stakeholders, end-users, and patients</p> <p>Prepare communication plans to disseminate education and lifecycle monitoring reports to stakeholders, end-users, and patients</p>
AI Solution Rollout and Monitoring	<p>Execute AI solution rollout</p> <p>Recruit end-users and patients to provide critical perspectives to inform the continued clinical integration of this AI solution</p> <p>Regularly monitor AI solution using the outlined Audit plan to ensure that it continues to meet defined equity objectives and fairness metrics</p> <p>Regularly report on disparities in AI solution performance, clinical adoption, and any adverse events</p>
Updating or Decommissioning of AI Solution	<p>If inequities worsen, consult with the cross-functional team to determine if the solution should be paused, updated, or decommissioned</p> <p>If AI solution updates are feasible, assess potential upgrades to the model or its environment and pilot or silently evaluate these changes. If performance improves and equity objectives and fairness metrics are met, rollout upgraded AI solution</p> <p>If updates do not improve performance, conduct a final assessment and decommission AI solution</p> <p>Communicate any changes to the AI solution based on the pre-established communication plans</p>