

RESEARCH ETHICS ARTIFICIAL INTELLIGENCE / MACHINE LEARNING APPLICATION SUBMISSION CHECKLIST

This checklist highlights key items in evaluating the ethical acceptability and privacy-related information of a research application that involves the use and/or development of AI/ML algorithms.

How to use this checklist:

1. Address each topic as applicable to your specific application using the space provided after each bullet point
- For categories that do not apply, please indicate with "N/A" and a brief explanation.
2. If you have multiple aims that propose multiple algorithms, address each point for each algorithm separately and use labelling (e.g. a, b, c) for clarity throughout the checklist.
3. Attach the completed form to Section 9.8 of the RISE application.

1. THE ALGORITHM



Provide information on each proposed algorithm

- Please list each algorithm in your proposal and the specific clinical function of the algorithm (i.e., Treatment response, prognostic, diagnostics, improved efficiency, or other).

Response:

- Please state if this application uses supervised, unsupervised, or semi-supervised algorithms and a broad description of the ML algorithm.

Response:

- Are you developing a new algorithm or are you proposing the evaluation of a previously developed algorithm?

Response:

2. PURPOSE AND CONTEXT OF THE ALGORITHM

Describe the healthcare context and the rationale for developing the algorithm.

- Describe the healthcare setting where the model is intended to be used or is needed.

Response:

- Please describe how the algorithm would be expected to fit or complement the current clinical workflow, including when the algorithm would be applied.

Response:

- What is the algorithm compared to (the existing standard of care)? If a prior model exists, please justify developing a new model.

Response:

3. THE TARGET POPULATION

Describe the target population and the algorithm's intended purpose in the care pathway context, including its intended users (e.g., healthcare professionals, patients, and the public). Please provide this information even if this application will not be doing this directly

- Describe who is the target population for the developed or evaluated algorithm (e.g., people of a certain age, in a specific country, or with a specific disease).

Response:

- Describe any known inequalities between sociodemographic groups in the healthcare setting where the model is intended to be used.

Response:

- Describe who the intended users of the model are and if the model is for healthcare professionals, patients, the public or others. Have they been involved in the study design?

Response:

4. DATA SOURCES & STUDY DESIGN FOR ALGORITHM TRAINING / VALIDATION

- Describe the study design (prospective, retrospective) and comment on the suitability of the design for this purpose.

Response:

- Describe the sources of data separately for the development and evaluation datasets (e.g., randomized trial, cohort, routine care or registry data), the rationale for using these data, and the representativeness of the data. If synthetic data is used, please provide a rationale and describe how the data was generated.

Response:

- Describe the inclusion and exclusion criteria of participants and the inherent biases in your data sources (i.e. Selection bias, systematic bias, such as demographics, rural vs urban population, etc.) in comparison with the intended clinical use of the algorithm (e.g., underrepresented demographics in the training or evaluation datasets)

Response:



Note: It is important to acknowledge these biases in the interpretation of study results

- Define the inputs to your model and clearly define the output or outcome and the time horizon, including how and when the outcome is assessed, the rationale for choosing this outcome, and whether the outcome assessment method is consistent across sociodemographic groups.

Response:

- How reliable the labelling of the outcome in the training and/or validation data was or is expected to be?

Response:

- Explain how the study size was arrived at (separately for development and evaluation) and justify that the study size was sufficient to answer the research question.

Response:

- The degree of missing data in the training and/or validation dataset and how this was/will be handled.

Response:

5. MODEL PERFORMANCE

- Please describe the internal and external validation steps, including discrimination, calibration, and utility. Justify the performance assessment values that were or will be used in this algorithm.

Response:

- Describe the generalizability of the model and any approaches to investigate model fairness and their rationale.

Response:

- Please explicitly state the potential risk of lack of explainability if a black box model will be used.

Response:

- For algorithms that have a planned deployment, please comment on how the algorithm will be monitored, updated, and reviewed after deployment and how difficult this will be to perform.

Response:

6. ETHICAL, LEGAL OR SOCIAL CONCERNS OF THE ALGORITHM

Please indicate the following:

- Have participants consented explicitly to their data being used for this application, or is a waiver of consent being requested?

Response:

- For algorithms that have a planned deployment, who takes **responsibility for post-implementation monitoring** of the safety and efficacy of the algorithm?

Response:

- **Potential incidental findings** resulting from the use of the algorithm and the plan to address these (i.e. unanticipated findings, such as participant re-identification, unexpected information on socioeconomic status, race/ethnicity, etc.).

Response:

- How the various groups involved would be impacted if the **algorithm fails** (i.e., the patient is sent for unnecessary screening/testing).

Response:

7. PRIVACY AND SECURITY CONCERNS

Please indicate the following:

- Is this project being conducted with **partners outside the BC Cancer/PHSA**? If yes, please describe the data or products shared with these partners and indicate whether the TDO has been notified about the potential need for an agreement.

Response:

- Please clearly state **who will have stewardship of the data** used in the algorithm and if any ownership claims of the data are anticipated.

Response:

- Who will **own the intellectual property** of the algorithm.

Response:

- Will this research be **conducted internally** within a pre-vetted BC Cancer or PHSA environment? Please describe how the data will be accessed and describe the environment and platform. Note: New platforms, environments, or activities not vetted for security and privacy compliance and/or aim to connect with Health Authority systems will require a Privacy Impact Assessment (PIA) and security review.

Response:

- How will **privacy be preserved while using the algorithm**? Note that Google Collaboratory is not appropriate to use for sensitive data.

Response:

- Does **each study site have the resources/equipment** needed to implement this algorithm for the target population? If not, describe any potential concerns regarding including different institutions in the implementation and use of this algorithm.

Response: