| U W  and H S D logo | **SUPPLEMENT Artificial Intelligence** |
| --- | --- |
|  | |
|  | |

PURPOSE and INSTRUCTIONS

This supplement is required for UW reviewed human subjects research involving the use of [**Artificial Intelligence Systems (AI)**](https://depts.washington.edu/comply/docs/308_G1.pdf) **when**:

1. it is led by **School of Medicine Principal Investigators** (PIs); **AND**,
2. involves either the targeted enrollment of **UW Medicine patients, OR** use of[**UW Medicine Data.**](https://depts.washington.edu/comply/docs/308_G1.pdf)*Note: UW Medicine Data includes**many types of data stored in any UW Medicine system or application and is NOT limited to clinical records.*

It should be used with the [**Guidance for School of Medicine Research using AI**](https://www.washington.edu/research/hsd/guidance/ai) to describe a plan for identifying, assessing, and mitigating risks associated with the use or development of Artificial Intelligence (AI) systems in human subjects research.

The guidance and related supplement are designed to align with the [UW Medicine Policy on Use of Artificial Intelligence in the Healthcare Setting](https://depts.washington.edu/comply/docs/comp_308.pdf) and use the definitions of AI, UW Medicine Data, and Secure UW Medicine Environment from the [glossary of this policy](https://depts.washington.edu/comply/docs/308_G1.pdf). The documents address both research involving the development of AI systems and the use of AI as a tool to facilitate the administration of a research study (e.g., recruitment, safety monitoring, data analysis).

**Exception**: The requirement to submit the supplement does **not** apply to the use of AI as a tool for the administration of research when: 1) the AI tool(s) have been approved for use in clinical care and UW medicine business operations as described in UW Medicine’s policy **and** 2) are being used for their approved purpose.

When the research will **not** involve data collection through interaction with research participants (e.g. it involves only use of secondary data), the supplement may be submitted in conjunction with the *shorter* [IRB Protocol, No Contact form](https://www.washington.edu/research/forms-and-templates/zipline-application-irb-protocol-no-contact-with-subjects/). Otherwise, the [standard IRB Protocol form](https://www.washington.edu/research/forms-and-templates/zipline-application-irb-protocol/) should be used.

Upload the completed Supplement to your Zipline application on the **Study-Related Documents** SmartForm for multi-site studies or the **Local Site Documents** SmartForm for single-site studies. It is not required for studies that will be reviewed by an external IRB instead of the UW IRB.

|  |
| --- |
| Click or tap here to enter text. |

**Study Title:**

AI DESCRIPTION

1. **Name of technology, model and version number.** [E.g., Developer – Curemetrix, Model – cmTriage, Version – v3.1]

|  |
| --- |
| Click or tap here to enter text. |

1. **Provide a detailed description of the AI model and its role in the study**.
   * Include the model’s intended purpose, functionality, limitations, and whether it is commercially available.
   * Explain whether the model is static or adaptive.
     + *Static AI has programmed objectives and relies on pre-defined rules and knowledge to generate outputs. It typically has transparent and understandable decision-making processes. Examples: Google Translate, email spam filters, credit score calculators.*
     + *Adaptive AI has objectives that have the potential to evolve over time. It learns from new data and is capable of autonomous or semi-autonomous behavior. Adaptive AI often has decision-making processes that are opaque. Examples: Google Maps, Duolingo, ChatGPT.*
   * Clearly distinguish between standard of care and investigational procedures.

|  |
| --- |
| Click or tap here to enter text. |

1. **Stage of development/use for research administration**. Review the information [about IRB Review and Stage of Development or use of AI](https://www.washington.edu/research/hsd/guidance/ai/#4) in the guidance on School of Medicine Research Involving AI and indicate the development stage of the AI system or if it will be used for the administration of research. Note that UW IRB will review only one stage of development at a time. A study modification or a new application may be submitted for subsequent stages.

**STAGE 1 – Discovery.** The conceptualization and early development of AI algorithms in research. This stage involves gathering and early analysis of training data to explore potential use cases.

**STAGE 2 - Translation.** The advancement of AI systems in research from ‘conceptual development’ to ‘validation,’ emphasizing performance testing and identifying risks.

**STAGE 3 - Deployment.** The use of a tested and validated AI system within a research context to confirm clinical efficacy, safety, and risks.

**Administration of Research.** Use of artificial intelligence technologies to facilitate various aspects of the research process (e.g. recruitment, data analysis, and patient monitoring).

DATA SOURCES and POPULATIONS – complete this section for all stages /uses

1. **Data source(s).**

* Describe the data sets that will be used in the AI model and their provenance, and/or participant populations that will be recruited.
* Include a description of the collection method (e.g., Application Programming Interface (API); scraping (automated programs to collect data, faces, voice, etc., from a website).

|  |
| --- |
| Click or tap here to enter text. |

1. **Minimum necessary.** Check the box below to confirm that you are not using more data than is necessary to meet the objectives of the research.

**CONFIRMED**

1. **Population(s).**

* Describe the relationship between the software function’s calculations (based on selected inputs) and the target problem to be solved.
* For AI that is in development (stages 2 and 3 in particular), provide information about the accuracy, representativeness, and fit of the input data for the purpose and targeted deployment population.
* For AI that is in development, explain whether the data sets the model is tested on will differ from the population(s) the algorithm is designed to impact and whether the outputs will be generalizable outside the input populations and target populations.

|  |
| --- |
| Click or tap here to enter text. |

PRIVACY and SECURITY – complete this section for all stages/uses

1. **UW Medicine Security Review.** Does this research involve the use of AI outside of a [**Secure UW Medicine Environment**](https://depts.washington.edu/comply/docs/308_G1.pdf)?

**No →** UW Medicine security review **is not required**.

**Yes** **→** A UW Medicine security review of the research **is required**. Upload documentation that the security review has been completed and a copy of the results of the security review to your Zipline application.

**CONFIRMED**

***NOTE –*** *Your IRB application cannot be approved without completion of the security review. For information about how to obtain a security review, refer to the* [*UW Medicine Information Security Risk Management Program*](https://depts.washington.edu/uwmedsec/restricted/services/risk-management/) *website. For questions about this process, contact* [*uwmed-security@uw.edu*](mailto:uwmed-security@uw.edu)*.*

1. **Inference of sensitive information, re-identification, and unintentional disclosure.** AI can infer sensitive personal attributes from seemingly innocuous data,link de-identified data back to an individual participant and even re-identify medical images (e.g. by reconstructing faces from MRI or CT scans). Additionally, there is the risk of AI models unintentionally leaking private or sensitive information in generated outputs. Describe the likelihood of these risks occurring in your research and describe any safeguards that will be in place, such as encryption, data masking and tokenization, noise addition, training models on decentralized sources, controlling access to sensitive data.

|  |
| --- |
| Click or tap here to enter text. |

1. **Future use of data.** Describe what will happen to the study data once the current project is complete. For example:
   * Will the model continue to use the data for future training?
   * Will the model be shared and if so, with whom?
   * How will data privacy be maintained post-commercialization?
   * If a commercially available tool is used, does the tool retain or use data beyond the current study?

Be as specific as possible about reasonably foreseeable purposes for which participant data may be used in the future, how it will be shared, with whom it will be shared, how long it will be stored, to what degree it is possible to destroy or scrub the data from the system and when this will happen.

|  |
| --- |
| Click or tap here to enter text. |

STAGE 1: Discovery – Do not complete for stages 2 and 3, or Research Administration

1. **Health impact.** Check the box below to confirm that the research will not impact patient or participant treatment, healthcare or clinical decision-making.

**CONFIRMED**

1. **Segmentation of AI system.** Will the AI system model be operated offline (i.e. outside of the production environment) and remain isolated from patient care systems?

Yes**→** Continue to next question.

No **→** Describe the safeguards and measures that will be in place to protect the production environment and patient care systems.

|  |
| --- |
| Click or tap here to enter text. |

1. **Plan to mitigate bias.** Provide a plan to address issues with data quality and the potential for bias in the model or the training data. Elements of this plan may include:

* Proactively identifying potential sources of bias in the data and model.
* Developing a strategy to acquire or generate diverse data to improve representation in the dataset.
* Establishing data provenance and assessing the completeness, quality, and representativeness of potential datasets.
* Identifying and addressing sensitive or proxy variables that could introduce bias into the model.

|  |
| --- |
| Click or tap here to enter text. |

STAGE 2: Translation – Do not complete for Stages 1 and 3, or Research Administration

1. **Health impact.** Check the box below to confirm that the research will not impact patient or participant treatment, healthcare or clinical decision-making.

**CONFIRMED**

1. **Segmentation of AI system.** Will the AI model be operated offline (i.e. outside of the production environment) and be isolated from patient care systems?

Yes**→** Continue to next question.

No **→** Describe the safeguards and measures that will be in place to protect the production environment and patient care systems.

|  |
| --- |
| Click or tap here to enter text. |

1. **Benefits.** Provide a description of who the technology is designed to benefit and whether and how the technology may benefit the populations from which the data originates.

|  |
| --- |
| Click or tap here to enter text. |

1. **Incorporation of early findings.**
   * Describe how the system has been adjusted based on findings from discovery (Stage 1), including changes that were made to the model architecture, training data, or feature selection, and provide the rationale for the changes.
   * Discuss any ways in which the system has evolved or adjusted as a result of its own functioning, and how these changes have been identified, evaluated, and addressed by the research team to ensure continued alignment with research objectives.

|  |
| --- |
| Click or tap here to enter text. |

1. **Plan to mitigate bias.** Provide a plan to address issues with data quality and the potential for bias in the model or the training/ assessment data. Elements of this plan may include:

* Going beyond standard accuracy metrics and evaluating performance across different subgroups.
* Implementing techniques like re-sampling, re-weighting, and fairness-aware algorithms during model training and validation.
* Challenging the model with adversarial examples and edge cases to expose biases and vulnerabilities that might otherwise remain undetected.
* Addressing missing values, outliers, noise, and duplicates in the dataset to improve data quality and prevent skewed predictions.

|  |
| --- |
| Click or tap here to enter text. |

1. **Preparation for deployment.** Describe the steps that will be taken to anticipate the harms and implications of incorrect predictions in preparation for deployment (Stage 3). These steps may include:

* Systematically identifying all potential negative impacts of incorrect AI predictions in the specific deployment context.
* Selecting appropriate metrics to track the model's performance relevant to the clinical task, such as diagnostic accuracy, precision, recall, sensitivity, and specificity for diagnostic tools, or predictive accuracy for prognostics.
* Implementing comprehensive testing methodologies to rigorously assess the model's performance.
* Involving subject matter experts throughout the testing and validation process to ensure the model's outputs are contextually appropriate and reliable.
* Designing validation processes to reflect the actual environment and challenges the AI model will encounter during deployment.

|  |
| --- |
| Click or tap here to enter text. |

STAGE 3: Deployment – Do not complete for Stages 1 and 2 or Research Administration

1. **Benefits.**
   * Provide a description of who the technology is designed to benefit and whether and how the technology may benefit the populations from which the data originates.
   * Describe any potential for the AI to provide a direct benefit to the participants in a clinical setting and any potential for the AI to provide a broad benefit for the improvement or scalability of healthcare.

|  |
| --- |
| Click or tap here to enter text. |

1. **Risk of harm.** Describe the potential for the technology to cause harm in the research context, including any negative impact on clinical care and decision-making (e.g. diagnosing, alleviating, treating, or preventing a disease or condition) and the plan to mitigate these risks.

|  |
| --- |
| Click or tap here to enter text. |

1. **Transparency, explainability and clinical oversight.**
   * Explain how the study will ensure transparency in decision-making processes, facilitate the understanding of the AI outputs for the users (e.g. researchers, clinicians, participants), and what clinical oversight procedures that will be in place. This may include:

* Implementing systems for human evaluators with appropriate expertise to review AI recommendations or decisions, particularly in high-risk areas.
* Defining a clear process for humans to override or appeal AI decisions and for investigating and resolving concerns raised by users or stakeholders.
* Providing training to users and stakeholders on how to interact responsibly with the AI system, understand its limitations, and identify potential biases.
* Integrating explainability tools or features into the AI system to help users interpret outputs (e.g. confidence scores, visualizations of decision pathways, feature importance rankings, or natural language explanations.)
  + Establishing structured feedback channels to report inaccuracies, biases, or unexpected outputs, with mechanisms to incorporate this feedback into model retraining or refinement cycles. Identify the personnel who will monitor AI performance and intervene if necessary.

|  |
| --- |
| Click or tap here to enter text. |

1. **Plan to mitigate bias.** Provide a plan to address issues with data quality and the potential for bias in the model or the training/ assessment data. Elements of this plan may include:

* Defining key performance indicators and fairness metrics for tracking the model's performance across relevant patient subpopulations, such as different demographic groups.
* Continuously auditing and testing the deployed model for bias, employing fairness metrics and stress-testing it for discriminatory patterns.
* Integrating techniques like fairness constraints, reweighting, or adversarial debiasing into the model to reduce discriminatory outcomes.

|  |
| --- |
| Click or tap here to enter text. |

1. **Monitoring plan.** Describe your plan for continuous monitoring the performance of the AI system. Elements of this plan may include:
   * Selecting appropriate metrics to track the model's performance relevant to the clinical task, such as diagnostic accuracy, precision, recall, sensitivity, and specificity for diagnostic tools, or predictive accuracy for prognostics.
   * Defining pre-specified thresholds for each metric that will trigger alerts, further investigation, or even intervention when performance deviates from acceptable limits.
   * Implementing continuous validation checks on incoming data to ensure its quality and integrity, identifying and addressing issues like missing values or inconsistencies.
   * Configuring anomaly detection algorithms to identify unusual patterns in data or model outputs that could indicate errors, biases, or unexpected behavior.
   * Maintaining comprehensive logs of all model inputs, outputs, and any human interventions or overrides for auditing, root cause analysis, and establishing a detailed record of the AI system's actions.

|  |
| --- |
| Click or tap here to enter text. |

AI for RESEARCH ADMINISTRATION – Do not complete for development stages 1, 2 or 3

1. **Validation of the solution.** Explain how the AI model has been validated for its intended purpose in the administration of the research. This may include:
   * What performance metrics were used by the developer to validate the model and were the metrics appropriate for the intended use in research administration?
   * Has the model undergone independent testing or third-party audits, and were the results made available to the research team?
   * Has the research team conducted its own assessment of the model’s performance in the specific context of the model that will be used in this study?

|  |
| --- |
| Click or tap here to enter text. |

1. **Identification of bias.** Explain how the research team will identify, evaluate, and mitigate potential for bias in the AI-driven processes. For example:
   * Has the research team reviewed the AI vendor’s documentation (e.g. model card, data sheets or system card) from the AI vendor to understand what data sources were used to train the model and the known biases and limitations of the model?
   * Has the research team conducted an independent assessment of the model’s performance using representative and diverse data sets? Were edge cases and underrepresented groups included in the evaluation to uncover hidden biases?
   * Is there a need to filter, re-weight, or otherwise process input data to reduce bias in model outputs?
   * Will the team implement strategies to balance outcomes across different demographic groups or individuals?

|  |
| --- |
| Click or tap here to enter text. |

1. **Transparency, explainability and clinical oversight.** 
   * Explain how the study will ensure transparency in decision-making processes, facilitate the understanding of the AI outputs for the end users or stakeholders (e.g. researchers, clinicians, participants), and what oversight procedures that will be in place. This may include:

* Implementing systems for human evaluators with appropriate expertise to review AI recommendations or decisions, particularly in high-risk areas.
* Defining a clear process for humans to override or appeal AI decisions and for investigating and resolving concerns raised by users or stakeholders.
* Providing training to users and stakeholders on how to interact responsibly with the AI system, understand its limitations, and identify potential biases.
  + Identify the personnel who will monitor AI performance and intervene if necessary.

|  |
| --- |
| Click or tap here to enter text. |

1. **Risk of harm.** Describe the potential for the technology to cause harm in the research context and how these risks will be mitigated.

|  |
| --- |
| Click or tap here to enter text. |

**Keywords:** Artificial intelligence; Large Language Models, Deep Learning, Generative AI, Machine Learning.