Key Principles for Evaluating AI-Driven Interventional Trials

The evolution of artificial intelligence (AI) in the dynamic landscape of medical technologies has ushered in a new era of possibilities in healthcare. According to the OECD definition, an AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.[1]

AI-based tools are increasingly being integrated into the healthcare system, as is the prevalence of submissions for clinical trials using these tools. In light of this increase, there is also a need to present guidelines for reviewing and approving applications, while ensuring safety, efficiency and observance of ethical rules.

In this document, we outline the guiding principles for examining submissions for interventional clinical trials using AI-based tools, tailored to the unique considerations inherent in these technologies. Inspired by the SPIRIT AI list for clinical trials,[2] these principles are designed to provide comprehensive guidance to Institutional Review Boards (IRBs), also known as Helsinki committees, responsible for evaluating submissions for trials involving AI-based products. These principles are designed to improve decision-making processes by highlighting issues unique to these technologies and promoting the values of safety, privacy, accountability, fairness, disclosure, transparency and explainability in the use of AI models.

It should be noted that while this document is intended for examination of interventional clinical trials, the issues raised in it could be addressed at earlier stages in the research and

[1] https://oecd.ai/en/wonk/definition
[2] Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension | Nature Medicine

development process of the tools. In this context, it is recommended to also read the guiding principles for Good Machine Learning Practice for Medical Device Development.[3]

At the core of the approach is the recognition of the unique characteristics of AI-based products, such as: data reliance, algorithmic complexity, and the ability to continuously learn and autonomously update over time. As a result, evaluating these products requires disclosure and transparency of the sources and databases used in its training, training techniques, model limitations, data on past performance and information on existing control pathways.

The principles, presented herein in the form of questions, emphasize the importance of assessing the benefits of AI-based interventions versus assessing risks to patient safety and privacy. The principles highlight the need to examine the quality of the information used to train and test the models, as well as possible biases arising from the data bases used for their training. In addition, there is reference to the processes of information flow from the environment to the models and back, their mode of operation, and the manner in which these technologies are used, with an emphasis on the interaction between human and machine and the training required for sensible and supervised use. Most of this information will be provided by the applicant, with the purpose of the questions in the document being to clarify the information requested by the committee, and the rationale behind it.

Recognizing that not all AI applications are equal in risk or clinical benefit profiles, we support an approach that matches the complexity and novelty of the technology being evaluated, and the clinical context in which it will be applied. There is room for the Helsinki Committees to exercise discretion in implementing these principles, in consultation with experts in the field, and to examine the applications that come before them according to the perceived risk and benefit. To identify and classify risks in the tool, it is recommended

[3] Good Machine Learning Practice for Medical Device Development (gov.il)

to use the International Medical Device Regulators Forum (IMDRF) documents related to risk management[4][5].

To date, we identify three main types of studies throughout the different stages of development, training, testing and implementation of AI tools in healthcare organizations: the model development and training phase, the silent prospective study (where no intervention occurs in the clinical process), and the active prospective study, where the tool is actively used within the clinical process:

- **Model development and training phase**: At this stage, when there is no defined tool yet, the goal is to examine different models and identify the model that performs best in predicting the target variable. If there is already a model to be tested, retrospective data is required for validation and testing of the model against the data that the organization can provide. In these cases, it is important to consider the privacy and security of patients' data, in accordance with the principles set forth in the Secondary Uses of Health Information Circular [6] and in the Collaborations Based on Secondary Uses of Health Data Circular.[7] ***For example:*** *a study focusing on early detection of signs of heart attack according to ECG data. At this stage, several different models are tested using existing ECG data to identify the best model for predicting heart attacks.*
- **Silent prospective study**: At this stage, the model is tested in the real world without affecting patients. Added here is the risk of integrating the model into the operational systems of the health organization, which may affect their functioning. Integration poses additional challenges in information security and ongoing data de-

[4] "Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations (imdrf.org)
[5] IMDRFSaMD WGN81 DRAFT 2024, Medical Device Software Considerations for Device and Risk Characterization - final draft.pdf
[6] https://www.health.gov.il/hozer/MK01_2018.pdf
[7] https://www.health.gov.il/hozer/MK02_2018.pdf

identification, which requires attention in accordance with the principles set forth in the Secondary Uses Circular.

***For example**: a study for early detection of signs of a heart attack that is activated in silent mode in hospitals, analyzing ECG data in real time but not affecting treatment decisions. At this stage, there will be a comparison between the records of the relevant department regarding heart attacks and the predictions of the system in order to test its performance.*

- **Active prospective study:** At this stage, the model enters a real-world interventional trial and therefore may influence the medical decision-making process and, accordingly, the treatment. Ethical and safety issues related to the unique characteristics of AI technologies as proposed in the table below should be examined.

  ***For example**, a study for early detection of signs of a heart attack that is in active mode and alerts medical staff about patients at high risk of heart attack in real time. At this stage, the effects on treatment decisions, the accuracy in preventing heart attacks, and other ethical and safety challenges of the AI tool are examined.*

These principles emphasize the importance of creating clinical evidence in supporting clinical trial approvals. While AI holds the promise of a revolution in healthcare, its application must be grounded in rigorous scientific validation and clinical evidence. We encourage institutional Helsinki committees to examine the quality and reliability of the data underlying AI-based interventions, ensuring transparency and reproducibility.

The issues below seek to foster a balanced approach to evaluating AI-based products in clinical trials, prioritizing patient safety, scientific integrity, and ethical considerations. By adopting these guidelines, committees can promote high standards of scientific rigor and ethical practice, ultimately advancing the collective mission of improving patient care through innovative technologies.

DRAFT

Key Principles for Evaluating AI-Driven Interventional Trials

In light of the rapid pace of development of AI technologies, the continues unrevealing of ethical and safety aspects, as well as the development of the international regulatory infrastructure, we consider this document to be an adaptive document subject to updates, and we would be happy to receive responses and feedback to: samd@moh.gov.il.

DRAFT

Key Principles for Evaluating AI-Driven Interventional Trials

Question	Significant issues
Does the title of the study state that the clinical intervention is based on AI, machine learning or an algorithmic model?	It is important to consider the degree of autonomy/automation of the tool (fully autonomous, partially autonomous or non-autonomous).[8]
Does the informed consent form state that the technology being tested in the clinical trial is based on AI?	Make sure that the wording in the informed consent form is in understandable language and that there is a concise explanation of the significance of using AI in the context of the study.
What is the version of the model/tool used?	
What is the intended use/purpose of the AI tool?	In the intended use/purpose, the following points should be addressed: • The medical purpose. • Intended users (clinicians, patients, the general public, etc.). • The intended clinical environment. • The role of outcomes in the clinical process. • Mode of use (including the degree of autonomy of the tool). • Characteristics of learning and update of the model (proactive or continuous).
What are the inclusion and exclusion criteria at the participant and input data level? What is the reasoning for these criteria? - Do these criteria correspond to the designated population for the tool?	• The reasoning for the criteria can relate to the limitations of the model, risk management of bias, etc. • The entrepreneur should justify the sample size with the ability to reach statistical significance, and take into account the possible risk to patients.
Are there any previous regulatory approvals for the tool?	If there are previous approvals: • Do the approvals include the AI component in the product? • Have there been any changes to the product or model since approval? • What is the indication indicated in the certificate?
Is there prior information and evidence of intervention using the specific AI tool or other AI tool for the same clinical indication/purpose?	

[8] See definition in IMDRF's N81: IMDRF SaMD WGN81 DRAFT 2024, Medical Device Software Considerations for Device and Risk Characterization - final draft.pdf

DRAFT

Key Principles for Evaluating AI-Driven Interventional Trials

Question	Significant issues
What is the process of collecting/acquiring input data?	
What mechanisms and tools exist to maintain data privacy?	• Processes of de-identifying data should be addressed, and cybersecurity principles should be upheld. • If external tools are used for preserving privacy, it is necessary to specify which tools.
Is there reference to analyzing the model's performance in subpopulations (e.g., gender, age, ethnicity, etc.)?	
What are the factors to which the AI tool is compared to in the study?	It is necessary to consider whether the comparison is made to human best practice, and/or to the accepted professional practice for the clinical procedure.
How was the model trained?	• Training process: what data was used and how they were chosen, the algorithm and techniques chosen and the optimization process. • Human involvement: who developed the model, their expertise and their role in the process. • Model evaluation: how the model's performance was evaluated (performance on different sets of data/comparison with other models/expert evaluation/other). The training process should be considered with an emphasis on the data sets and their contents, and how they were chosen. • The limitations of the model must be addressed, inter alia, in relation to biases, sub-populations, sensitivities, etc. • There is room to attach basic statistics as well.
How will the performance of the model be tested? Is there a plan for continuous evaluation of model performance throughout the study?	It is important to address, inter alia, who is responsible for examining the performance of the model, and to address performance in subpopulations as applicable.

DRAFT

Key Principles for Evaluating AI-Driven Interventional Trials

Question	Significant issues
Does the tool require integration with other systems in the clinical environment?	• It is essential to specify which systems are interacting with the AI tools. • If there are specific versions of these systems, this should be noted. • Consideration must be given to the structure of the information, the types of information transmitted, the frequency of updating/online updating. • In addition, consideration should be given to the level of risk to the ongoing functioning of the systems, including the issue of information security.
Is there an interaction between a human and the AI tools in handling the input data? - If so, is a level of expertise required from the participants?	Consideration must be given to the involvement of the human factor in entering the data into the AI system, and the expertise required for this purpose.
Is there a program for evaluating and handling poor-quality, inaccessible, or incorrectly entered input data?	
Is there a risk management plan to prevent biases in the input data, information discontinuity (problems with availability and flow of information), etc.?	
What does the output of the AI intervention look like?	Results should be defined in two ways: model performance, and model and human joint performance.
How does the output contribute to the decision-making process or any other clinical activity?	• How does the degree of autonomy of the tool affect the clinical process? • Emphasis should be placed on the form of interpretability and analysis of the output. • Whenever possible, a breakdown of the rationale and an explanation of the model's output (explainability) should be incorporated, as well as an indication of the level of security. • Within the framework of the results of the model, metrics should be defined in various aspects: accuracy relative to the equator, saving time/resources, user satisfaction (medical staff, patients, etc.) and the rate of user intervention in system decisions.

DRAFT

Key Principles for Evaluating AI-Driven Interventional Trials

Question	Significant issues
What is the training process for the interface of the AI tools?	• Is there a reference to how usability is tested? • It is important to indicate whether digital literacy is required on the part of the participants in the experiment.
What are the criteria for stopping the experiment?	Emphasis should be placed on these criteria from the perspective of the risk management plan, for example: difficulty in translating the outputs into proper clinical activity, gaps in input data.
Is there any reference to reporting substantial changes to the model to the Committee ?	• Anticipated points for changes in versions of the model (including third-party software used), usage label or research environment should be noted. • It should be made clear if these are material changes that will require reconsideration by the institutional committee.
How can errors in model performance be identified and analyzed? - How is the risk around these errors managed?	• The risk of low performance of the model (as well as alarm fatigue) and the responses to these risks should be assessed. • Processes should be defined to identify and document errors in the model (for example, discrepancies between the user's decision and the system), processes that will allow recording the performance of the system for reverse investigation (for example, keeping a history of the prediction that performs the model with all the data (snapshot) for investigation or rerun if required.

DRAFT