

ARTIFICIAL INTELLIGENCE IN HEALTHCARE GUIDELINES (AIHGle)

Developed by:



Endorsed by:



Published Oct 2021

CONTENTS

1	FOREWORD	4
2	INTRODUCTION	5
2.1	Objectives	5
2.2	Guiding Principles	6
2.3	Definition, Scope and Interpretation of Guidelines	7
2.4	How to use these Guidelines?	9
2.5	Other Relevant Legislation and Guidelines	9
3	SETTING CLEAR RESPONSIBILITIES – DEVELOPERS & IMPLEMENTERS	12
3.1	Establishing Service Level Agreements (SLA)	12
4	DEVELOPMENT	13
4.1	KEY RECOMMENDATIONS	14
	DESIGN	15
4.2	Clinical Inputs	15
4.3	End-User Inputs	16
4.4	Understanding the Current Clinical Practice	17
4.5	Data	17
4.6	Cybersecurity	18
4.7	Explainability	20
	BUILD	21
4.8	Development Standards	21
4.9	Self-Validation	22
	TEST	
4.10	Evaluation and Monitoring of AI-MD	23

4.11 Intended Use and Workflow	26
5 IMPLEMENTATION	27
5.1 KEY RECOMMENDATIONS	28
USE	
5.2 Clinical Governance	29
5.3 Operational Workflows and Processes	31
5.4 End-User Communication	33
MONITOR & RESPOND	35
5.5 Post-Deployment Monitoring	35
REVIEW	36
5.6 Review and Tracking	36
6 EMERGING DEVELOPMENTS IN AI	38
6.1 Continuous Learning AI-MD and Deployment	38
6.2 Synthetic Data	40
7 SHARING YOUR FEEDBACK	41
8 CASE STUDY – SINGAPORE EYE LESION ANALYSER PLUS (SELENA+)	42

1 Foreword

Across the world and in Singapore, Artificial Intelligence (AI) is increasingly being used throughout the healthcare continuum – from training, research, administration, clinical decision support to direct patient care, so as to identify previously unseen insights, increase system efficiency and improve patient care and outcomes. However, alongside the benefits are also risks and ethical concerns if AI is not properly designed and implemented.

To improve the understanding, codify good practice and support the safe growth of AI in healthcare, the Ministry of Health (MOH), the Health Sciences Authority (HSA) and the Integrated Health Information Systems (IHIS) co-developed the **Artificial Intelligence in Healthcare Guidelines (AIHGle)**. The AIHGle (*pronounced as “agile”*) serves as a guide for developers and implementers of AI in healthcare and complements the existing HSA regulatory requirements of AI Medical Devices (AI-MDs). These guidelines aim to share good practices with the healthcare community to guide the safe development of AI in healthcare.

As a “living” document, the AIHGle will be periodically updated to incorporate good practices in the rapidly developing AI landscape.

I hope you find these guidelines useful as you embark on your AI journey.



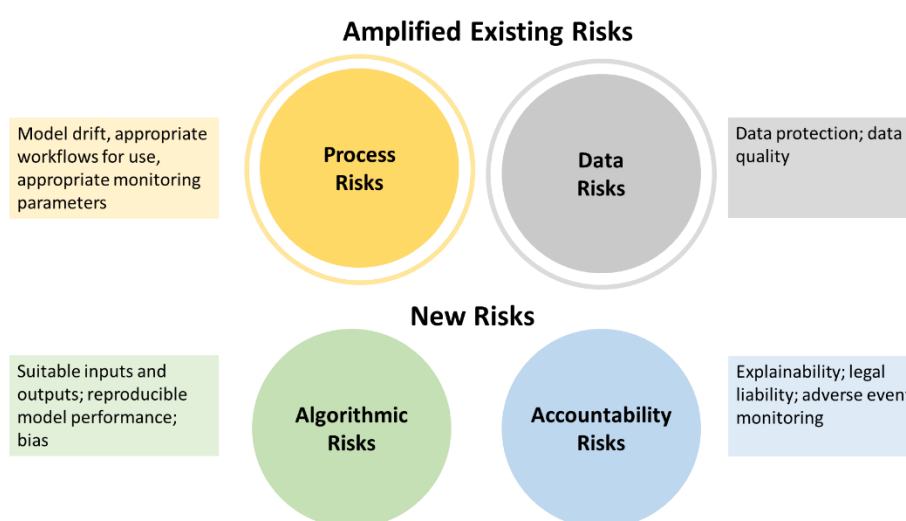
ASSOCIATE PROFESSOR (DR) KENNETH MAK
DIRECTOR OF MEDICAL SERVICES
MINISTRY OF HEALTH

2 Introduction

2.1 Objectives

Artificial Intelligence (“AI”) technologies have the potential to improve healthcare efficiency, accessibility, quality and affordability. However, AI also amplifies existing process and data risks, and creates new accountability and algorithmic risks (**Figure 1**) which, if not managed systematically, may lead to poor patient outcomes and erode clinician and patients’ trust in the use of AI – limiting the potential benefits of the technology.

Figure 1: AI’s Amplification of Existing Risks and Creation of New Risks



With the increasing use of healthcare AI in Singapore, the intent of the AIHGle is to improve clinical and public trust in the technology by:

- a. Providing a set of recommendations to encourage the safe development and implementation of primarily AI-Medical Devices (“AI-MDs”), and secondarily any other AI implemented in healthcare settings; and
- b. Signposting HSA's AI-MD registration requirements.

WHAT IS AN AI-MEDICAL DEVICE (AI-MD)?

Definition: An AI-MD as defined by the Health Sciences Authority (HSA) refers to AI solutions which are intended to be used for investigation, detection, diagnosis, monitoring, treatment or management of any medical condition, disease, anatomy or physiological process. AI-MDs typically have a direct impact to patient safety.

Examples: AI tools for diagnosis of sleeping disorders and cancers, image detection of diabetic retinopathy, and management of Type 1 Diabetes.

Aside from what is legislated under HSA's Health Products Act (HPA) and its subsidiary legislations and guidance documents, **the AIHGle complements HSA's regulatory requirements, and provides a set of good practices for developers and implementers. Healthcare institutions developing in-house AI-MDs are strongly encouraged to also refer to the AIHGle for guidance. The AIHGle is meant to be a "living" document that will be updated alongside the rapid development of AI.**

2.2 Guiding Principles

The recommendations in these guidelines are based on principles adapted from the AI guidance provided by the Personal Data Protection Commission (PDPC) and the Monetary Authority of Singapore (MAS).¹ The adoption of these principles serves to ensure the safe provision of AI services, and for building trust in the dependability and efficacy of AI in healthcare. Sections 2.2.1-2.2.5 briefly describe each guiding principle. These are further illustrated in the subsequent *Development* and *Implementation* sections.

2.2.1 Fairness

The development and implementation of AI-MD should not result in discriminatory or unjust clinical impact on patients across different demographic lines (e.g. race, gender, etc.).

2.2.2 Responsibility

While developers should be responsible for the proper design of algorithms used in the AI-MD, organisations using AI-MD to deliver care will be responsible for the decision to implement the AI-MD and the clinical outcomes arising from the use of AI-MD in ensuring that safe care is delivered. Similar to the implementation of any other MD, the use of AI-MD does not change the liability of the implementing institution or the individual medical professional in their provision of appropriate and safe care.

2.2.3 Transparency

End-users of AI-MD (e.g. medical practitioners, patients) should be informed that they are interacting with an AI-MD (*further details on end-user communication are in Section [5.4](#)*).

2.2.4 Explainability

The decisions or recommendations from an AI-MD should endeavour to be explainable and reproducible. The level of explainability is dependent on the varying expectations of the end-user and the risks of the AI-MD. End-users should be consulted during the development or adoption of the AI-MD to ensure the explainability meets their expectations. For example, this can include, but are not limited to, end-users knowing the data sets, testing protocols, and algorithmic model uses (*examples of ways to achieve explainability are in Section [4.7.2](#)*).

¹ PDPC – Model AI Governance Framework (2020); MAS – Principles to Promote FEAT in the Use of AI and Data Analytics in Singapore's Financial Sector (2018).

2.2.5 Patient-Centricity

Safeguards in the design, development, and implementation of AI-MD should be put in place to ensure that patients' interests, including their safety and well-being, are protected.

2.3 Definition, Scope and Interpretation of Guidelines

2.3.1 What is AI?

These guidelines take a broad view of the definition of AI, and considers AI to be a set of general purpose technologies intended to allow machines to (i) model and optimise, (ii) automate, (iii) forecast and (iv) classify/detect a required result.

Recent growth in AI applications has been driven by developments in Machine Learning (ML) and Deep Learning (DL) algorithms due to increased computational power.

Machine Learning refers to a set of algorithms and statistical models (e.g. Linear Regressions, Support Vector Machines, Random Forests) that allows machines to perform specific tasks without using explicit instructions, by relying on patterns in the input data.

Deep Learning² refers to a subset of Machine Learning techniques, using multiple processing layers to learn representations of data with several levels of abstraction.

2.3.2 What types of AI do these guidelines focus on?

AI in healthcare has many use-cases: such as for clinical, administrative, research, and in policy development. These guidelines focus on higher risk medical and clinical use, specifically AI-MDs.

While most of the AIHGLE focuses on “locked” AI-MDs³ which are the prevalent form of AI-MD today, there are algorithms that change over time and continuously ‘learn’. [Part 6](#) provides early guidance on these continuous learning AI-MDs.

These guidelines focus on the development and implementation of AI-MD. **However, similar principles may also broadly apply to any AI application in healthcare (e.g. training, wellness, administration, etc.),** as illustrated in [Figure 2](#).

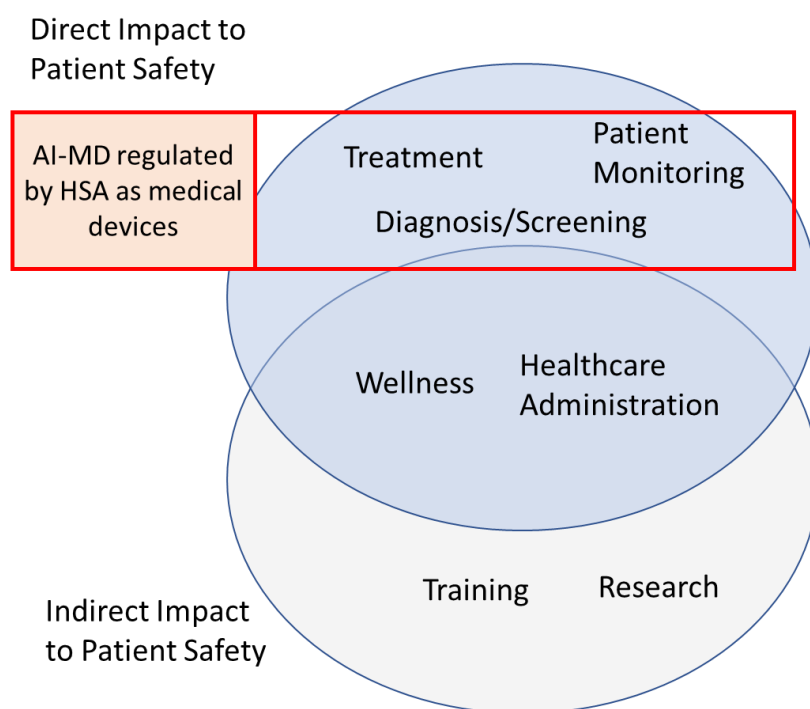
As for AI research, researchers intending to develop AI for eventual clinical use should refer to the recommendations in this document to facilitate the eventual operationalisation of their research in clinical settings. For the avoidance of doubt, researchers conducting AI-related research must still comply with all applicable laws and regulations, such as the Human Biomedical Research Act (HBRA).⁴

² Nature, Deep Learning, Yann LeCun et. al (2015).

³ The USA Food & Drug Administration defines ‘locked’ as an algorithm that provides the same result each time the same input is applied to it and does not change with use. Examples of locked algorithms are static look-up tables, decision trees, and complex classifiers.

⁴ Please see Section [2.5](#) for further details.

Figure 2: AI Use-Cases in Healthcare



2.3.3 Who do these guidelines apply to?

These guidelines apply to two key stakeholders in the development and implementation of AI-MD:

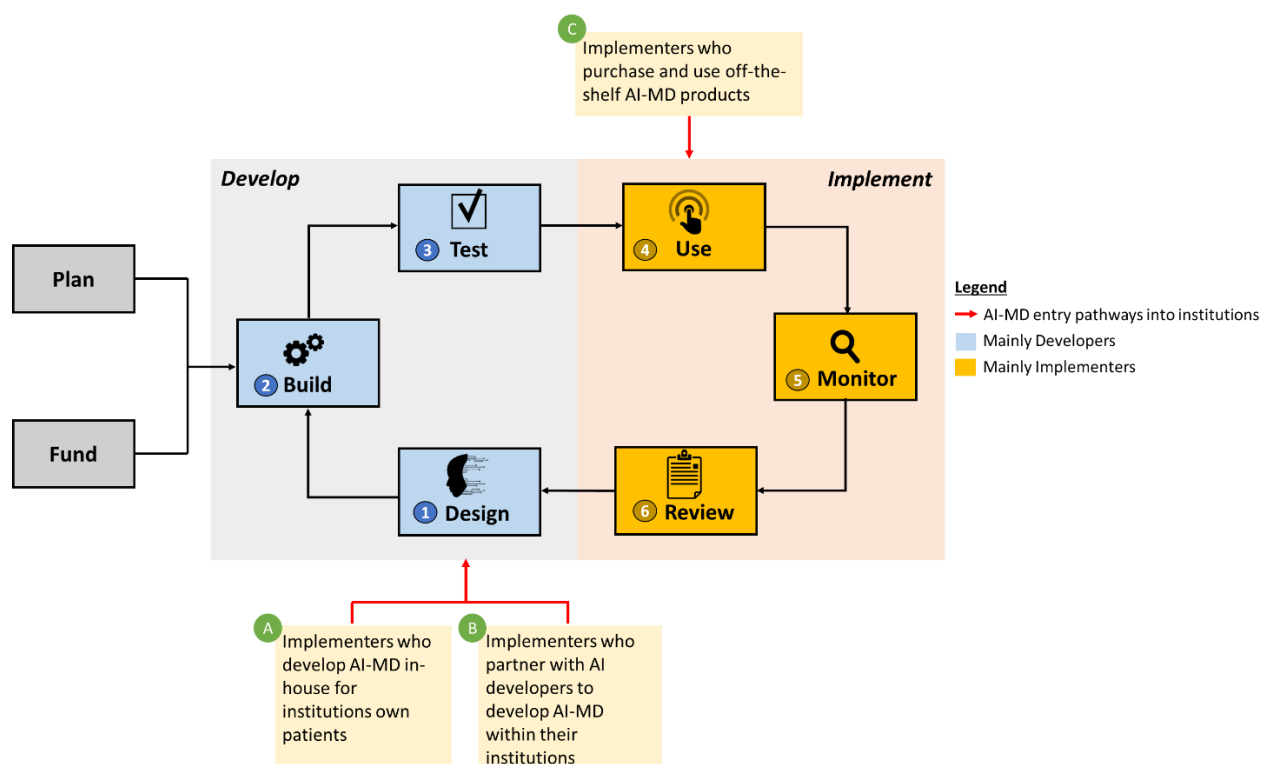
- a. **“Developers”** refer to organisations or individuals who **plan, fund, develop and/or maintain** AI-MD, including standalone software medical devices that can interact with patients directly, or AI-MD intended to be used as part of healthcare service provision by organisations or individual healthcare professionals.
- b. **“Implementers”** refer to organisations or individuals **who use AI-MD to deliver healthcare services** [e.g. those regulated under the Private Hospitals and Medical Clinics Act (PHMCA) or the future Healthcare Services Act (HCSA)].⁵

The two groups of stakeholders are not mutually exclusive. For example, some organisations can be both developers and implementers of AI-MD, such as hospitals that develop AI-MD in-house for use on their patients.

These guidelines recognise that there are a variety of methods of developing and implementing AI-MD (**Figure 3**). Nevertheless, developers and implementers will need to work together continuously and iteratively to successfully apply the guidelines.

⁵ The provision of healthcare services is not limited to services provided within licensed premises, and could include services provided at the patients’ home, or in mobile conveyances.

Figure 3: Different Ways of Developing and Implementing AI-MD – 3 Common Entry Pathways of AI-MD into Institutions



2.4 How to use these Guidelines?

These guidelines are structured into two sections: (1) recommendations for **developers**, followed by (2) recommendations for **implementers**. Key recommendations in each section are illustrated upfront for ease of reference. As many of these recommendations are part of HSA's existing AI-MD product registration requirements, these regulatory requirements are indicated in **red**. For an actual example of the applicability of the key recommendations, please see [Part 8](#)).

2.5 Other Relevant Legislation and Guidelines

In addition to complying with HSA's Health Products Act (HPA) regulations, organisations or individuals involved in the development or implementation of AI-MD should also consider the requirements set out in other related legislations/guidelines covering the provision of healthcare services, professional responsibilities, product safety and data protection, amongst others. Examples of the relevant legislation and guidelines are listed in **Table 1**.

Table 1: Examples of Key Legislation and Guidelines⁶

S/N	Legislation and Guidelines	Regulatory Context
1	Personal Data & Protection Act (PDPA) and Guidelines <ul style="list-style-type: none"> • Specific Guidance for the Healthcare Sector (2017) 	Obligations over the protection, collection, use, disclosure, and access of personal data used in AI-MD.
2	Private Hospitals & Medical Clinics Act (PHMCA) and Guidelines <i>(to be transitioned to the Healthcare Services Act (HCSA) starting 2022)</i> <ul style="list-style-type: none"> • National Guidelines for Retention Periods of Medical Records (2015) 	Healthcare Institutions licensed under the PHMCA, including those which use AI-MD, must comply with its requirements (e.g. controls on the maintenance and security of medical records) and all licensing terms and conditions.
3	Professional Registration Acts and their respective Ethical Code and Guidelines <ul style="list-style-type: none"> • Singapore Medical Council Ethical Code and Ethical Guidelines (2016) • Singapore Dental Council Ethical Code and Ethical Guidelines (2018) 	All registered healthcare professionals, including those who use AI-MD to deliver healthcare services (e.g. clinical decision support tools to read Computer Tomography (CT) scans) must comply with the requirements under their respective professional Acts and guidelines.
4	Civil Law (Amendment) Bill (2020)	Sets out the statutory test to determine the healthcare professional's duty in giving medical advice to a patient.
5	Health Products Act (HPA) <ul style="list-style-type: none"> • Regulatory guidelines for Telehealth products • Regulatory Guidelines for Software Medical Devices – A Lifecycle Approach 	Regulatory controls over medical devices (e.g. dealer's licensing, product registration, change management for registered medical devices, notification of medical devices used in clinical trials), and post-market surveillance.
6	Human Biomedical Research Act (HBRA) (2015) Ethical Guidelines for Human Biomedical Research (2015)	Researchers conducting healthcare research involving AI, which fall under the definition of human biomedical research, must comply with the requirements under the HBRA and/or refer to the ethical guidance document for researchers and those involved in human biomedical research in Singapore.
7	National Telemedicine Guidelines (2015)	Specific guidelines for the provision of telemedicine services.

⁶ References to Acts include subsidiary legislation.

S/N	Legislation and Guidelines	Regulatory Context
8	Healthcare Cybersecurity Essentials Guidelines (2021)	Provide guidance to all healthcare providers on basic cybersecurity measures that they can adopt to ensure the security and integrity of their IT assets, systems, and patient data.
9	PDPC Model AI Governance Framework (2nd Edition)	The Model Framework supports the translation of ethical principles into practical recommendations that organisations could readily adopt to deploy AI responsibly.
10	PDPC Implementation and Self-Assessment Guide for Organisations (ISAGO)	A companion to complement the voluntary Model AI Governance Framework (Model Framework) and aims to help organisations assess the alignment of their AI governance processes with the Model Framework, identify potential gaps in their existing processes and address them accordingly.
11	<p>Compendium of Use-Cases: Practical Illustrations of the Model AI Governance Framework</p> <ul style="list-style-type: none"> • Volume 1 • Volume 2 	A Compendium of Use-Cases demonstrating how various organisations across different sectors – big and small, local and international – have either implemented or aligned their AI governance practices with PDPC’s Model AI Governance Framework, or have effectively put in place accountable AI governance practices and benefit from the use of AI.
12	Singapore Computer Society (SCS) AI Ethics & Governance Body of Knowledge (BoK)	The Body of Knowledge (BoK) is an industry-led effort that provides a reference document for business leaders, Information and Communications Technology (ICT) professionals and Professionals, Managers, Executives, Technicians (PMETs) on the ethical aspects related to the development and deployment of AI technologies.

3 Setting Clear Responsibilities – Developers & Implementers

3.1 Establishing Service Level Agreements (SLA)

3.1.1 Unlike Commercial-Off-The-Shelf (COTS) AI-MDs which come with an End User Licensing Agreement (EULA) from the developer, and self-built AI-MDs where responsibility is fully borne by the implementer, **there are potential gaps, overlaps and a general lack of clarity over responsibility, when AI-MDs are developed in collaboration between developers and implementers.** To mitigate these gaps, **developers and implementers should enter into Service Level Agreements (SLAs)⁷** to set clear and mutually agreed responsibilities.

3.1.2 **Figure 4** summarises the recommended areas that an SLA could cover. Detailed information on these areas will also be covered in the subsequent sections.

Figure 4: Recommended Areas to be included within an SLA (non-exhaustive)

Design <i>e.g. seeking clinical inputs relevant to AI-MD's intended use, relevance of training datasets, setting performance baselines.</i>	Build <i>e.g. documentation of AI-MD development protocol and reference standards.</i>	Test <i>e.g. evaluation and validation of AI-MD model to ensure patients would be "no worse off".</i>
Use <i>e.g. appropriate approving authority for implementing AI-MD, operational workflow and staff training.</i>	Monitor <i>e.g. "ground-truthing" of AI-MD's performance, consistent and continued performance evaluation.</i>	Review <i>e.g. ad-hoc review of patient safety issues, annual performance review.</i>
Intellectual Property (IP) <i>e.g. access to specific info on algorithmic design.</i>		

⁷ For collaborations that are research-based in nature, Research Collaboration Agreements (RCAs) should be signed between developers and implementers, and the collaborative scope should be limited to the Design, Build and Test phases of an AI-MD. Terms and conditions within RCAs can take reference from the SLA's recommendations.

4 Development

*This section provides guidance for clinical use AI-MD developers, and is intended to serve as a general guide for the safe development of AI-MD. For avoidance of doubt, all AI-MD intended for use and commercial distribution in Singapore must be registered with HSA. AI-MDs developed by institutions for their own patients are not subject to HSA's current HPA registration requirements. **However, as a good practice, institutions are encouraged to register such AI-MDs with HSA to reassure institutions and patients of the AI-MD's quality, safety, and efficacy.***




In addition to these guidelines, developers should also refer to [HSA's website](#) for more details on registration requirements. Given the variety of use-cases of AI-MD, developers are encouraged to engage early with HSA under their [pre-market consultation scheme](#) for specific guidance on individual AI-MD applications.

4.1 KEY RECOMMENDATIONS

**Note: Recommendations (Figure 5) in red are part of HSA's existing AI-MD product registration requirements.*

Figure 5: Key Recommendations for AI-MD Developers

Developers should:



1 DESIGN
<ul style="list-style-type: none">• Seek clinical inputs relevant to the intended use of the AI-MD when developing the AI-MD. [4.2.1]• Seek end-user input for a holistic AI-MD design and development process. [4.3.1]• Determine the current clinical practice baseline to ensure that the AI-MD's performance is minimally no worse off than current practice. [4.4.1]• Ensure representativeness of datasets to reduce unintended bias. [4.5.1]• Document all biases and/or limitations identified in an AI-MD and rectify them if possible. [4.5.4]• Ensure AI-MD can prevent, detect, respond and where possible, recover from foreseeable cybersecurity risks. [4.6.1]• Demonstrate effectiveness of the AI-MD and endeavour to ensure a sufficient level of explainability based on what their intended end-user requires. [4.7.1]
2 BUILD
<ul style="list-style-type: none">• Adhere to HSA's regulatory guidelines for software medical devices. [4.8.1]• Adopt appropriate development standards (e.g. ISO 13485, ISO 14971, IEC 62304). [4.8.2]• Document properly all changes to AI-MD and ensure all software versions are reproducible. [4.8.3]
3 TEST
<ul style="list-style-type: none">• Evaluate and validate periodically the AI-MD's performance to ensure it minimally meets the clinical practice baseline. [4.10.1]• Document the intended use of the AI-MD. [4.11.1]• Define clearly and document how the AI-MD should be incorporated into clinical workflows. [4.11.2]

DESIGN

4.2 Clinical Inputs

4.2.1 **Developers should seek clinical inputs relevant to the intended use of the AI-MD in the development of their AI-MD.** These may be from individuals with relevant expertise⁸, and include but are not limited to inputs on:

- a. The clinical problem statement
 - i. Examples of areas to be covered: clarity on the issue, intended use, current clinical practice baseline, patient inclusion/exclusion criteria, possible clinical workflows, integration into clinical workflows;
- b. The representativeness⁹ of training and testing datasets which, beyond fair/unbiased outcomes, can have an impact on patient safety;
 - i. Examples of factors affecting representativeness: demographics, clinical context, heterogeneous expression of diseases or conditions in subgroups, known/existing biases, appropriate input type (image, text, numbers);
- c. Algorithm testing approach, including but not limited to:
 - i. Appropriate testing scenarios and validation methodology;
 - ii. Testing input quality (e.g. type and resolution of images); and
 - iii. Testing at the "boundary conditions" between valid/invalid input (e.g. between patient inclusion/exclusion criteria).
- d. Identifying causal relationships between inputs and outputs of the AI-MD, i.e. a precursor to explainability; and
- e. Developing the user manual(s), i.e. bringing a clinical lens to sharing how the AI-MD should be implemented safely and taking into consideration the understanding of existing clinical workflows.

4.2.2 Developers should **take ownership over the clinical inputs obtained for the development of the AI-MD.** For avoidance of doubt, an AI-MD development team should include clinicians (or relevant domain experts) to guide and lead the seeking of the necessary clinical inputs.

⁸ The relevant type of expertise depends on the use-case for the AI-MD and does not need to be limited to only medical practitioners.

⁹ Representative data refers to the selection of datasets that accurately reflects the characteristics (e.g. age, gender, socioeconomic status, profession, etc.), insights, and observations of a target population group.

4.3 End-User Inputs

- 4.3.1 In addition to seeking clinical inputs, **developers should seek end-user inputs (e.g. from medical practitioners, patients) as part of a holistic AI-MD design and development process.** Aspects of end-user inputs that developers can seek are listed in **Table 2**.

Table 2: Aspects of End-User Input for AI-MD Development

Type of End-user	Aspect
Medical Practitioner , in addition to clinical inputs obtained <i>(e.g. doctors who will be using the AI-MD in the future)</i>	<ul style="list-style-type: none">i. Clinical opinions on the safety and efficacy of the AI-MD.<ul style="list-style-type: none">a. What is required to build trust and confidence in adopting the AI-MD as opposed to current clinical methods.ii. Integration of AI-MD into the clinical workflow.iii. Usability (e.g. UX, UI) of the AI-MD, including phrasings of notifications for (i) changes/errors pertaining to input parameters and (ii) clarity of the user manual in guiding medical practitioner in output interpretation.
Patient <i>(i.e. those who will be interacting directly with the AI-MD)</i>	<ul style="list-style-type: none">i. What is required to build trust and confidence in the care rendered through AI-MD (e.g. quality of care expected vis-à-vis care from physical doctors).ii. Usability (e.g. UX, UI) of the AI-MD.iii. Clarity of the wording of notifications in the user manual to advise patients on whether they should/should not use the AI-MD for their symptoms/conditions.

4.4 Understanding the Current Clinical Practice

Current Clinical Practice Baseline

- 4.4.1 Before an AI-MD is developed and trained, **developers should determine the current clinical practice baseline to ensure that the AI-MD's performance is minimally no worse off than current practice in how it may impact patient safety/welfare** (see Section [4.10.1](#) on validation of AI-MD's performance to meet current clinical practice baseline). This includes establishing appropriate performance indicators, tracking and measuring existing clinical performance in terms of accuracy and specificity. In cases where there is no established medical knowledge and processes, the AI-MD can be compared with the status quo (i.e. no intervention).

4.5 Data

- 4.5.1 Recent developments in AI-MD have been driven by the growth of ML/DL models that move away from pre-defined rules, allowing machines to determine complex statistical relationships from data provided. Since ML/DL models are entirely dependent upon the integrity of training datasets, **developers should ensure that the datasets used in the development stage are representative to reduce unintended bias.**
- 4.5.2 **Developers should assess and document the following features of all data used in any production version of their AI-MD:**
- a. Credibility of the source of the dataset;
 - b. Period when the dataset was compiled or updated;
 - c. Accuracy of the dataset, including the accuracy of any annotated/labelled features;
 - d. Completeness of the dataset;
 - e. Medical relevance of the dataset;
 - f. Demographic representativeness (e.g. age, ethnicity, gender, sociodemographic stratum and location); and
 - g. Biases in the dataset.

- 4.5.3 Naturally occurring biases in datasets may represent the true nature of a population (e.g. prevalence of a chronic condition within a specific population sub-group), and therefore may not always be negative. However, unintended biases arising from incomplete datasets, poor labelling, and/or hidden variables are always negative and will result in inaccurate algorithmic decisions. Management of such biases should be done iteratively over the AI-MD's development process, and developers should use the following methods to assess the presence and validity of biases:
- a. Compare the representativeness of the inputs to the intended outputs of the AI-MD; and
 - b. Check for variable performance (in both output and error rates) of the AI-MD across demographic sub-groups and assess the clinical relevance of any disparity.
- 4.5.4 **Developers should comprehensively document all biases and/or limitations identified in an AI-MD (whether in, for example, the underlying dataset or algorithm) and rectify them if possible.** If not possible (or when investigations into the biases or limitations are still ongoing), developers should document these as limitations to the intended use of the AI-MD and reflect these in the user manual.
- 4.5.5 Once a representative dataset has been determined, developers should divide this into training and testing datasets and use appropriate methods to do so (e.g. stratified sampling). This is to ensure that both datasets are equally representative.

4.6 Cybersecurity

For cybersecurity guidance applicable to AI-MD implementation, please see Section [5.3.3](#).

- 4.6.1 The risk of cybersecurity threats lie in two key aspects of AI-MD development: (1) the AI-MD's design and (2) the data collected and used for an AI-MD's algorithmic decisions, which can in turn affect the AI-MD's outputs. **Developers should ensure that the AI-MD can prevent, detect, respond and where possible recover from foreseeable cybersecurity risks.** Further guidance on cybersecurity approaches can be found in [HSA's](#) and [International Medical Device Regulators Forum \(IMDRF\)'s](#) guidelines.

Security by Design

4.6.2 To ensure that patient care is not compromised due to security breaches, developers should secure the AI-MD's design by taking into account cybersecurity considerations, including but not limited to the following:

- a. Preventing unauthorised use of AI-MD (e.g. restricted role-based user access, user authentication controls);
- b. Detecting potential cybersecurity risks (e.g. continuous monitoring via regular security or antivirus scanning, incident logging feature for tracing of any cybersecurity attacks, self-validation of the AI-MD's robustness against cybersecurity attacks);
- c. Responding to cybersecurity incidents (e.g. notification feature to alert end-users of cybersecurity attacks, mitigating impact of cybersecurity attacks using anti-malware or firewall); and
- d. Recovering from cybersecurity incidents (e.g. in-built systems to deploy cybersecurity patches and updates efficiently).

Data Protection

4.6.3 Data protection is necessary to safeguard personal data from misuse and ensure patients' trust in how their data is managed. Further guidance on data protection provisions specific to the healthcare sector (e.g. seeking patients' consent on collection, use and disclosure of personal data) are available in [PDPC's advisory guidelines](#).

4.6.4 Developers should ensure that all data (i.e. training, testing, and 'live' clinical data) used in their AI-MD are secured by design. One way of protecting data and reducing the risk of sensitive personal information being exposed is to de-identify that data. In situations where certain individual characteristics need to be retained within the AI-MD algorithm, developers can also consider anonymisation techniques to protect this data (e.g. data masking, pseudonymisation, data perturbation).

4.6.5 However, anonymisation and de-identification of data do not remove all risks, such as the re-identification of the data. Developers should consider implementing other relevant data protection measures such as (but not limited to):

- a. Logging users' access;
- b. Limiting access to data via pre-specified user roles;
- c. Consistently reviewing the data access logs to pick up discrepancies and being aware of where all data collected are being transmitted and stored; and

- d. Applying Privacy-Enhancing Technologies (PETs) during data analysis and modelling (e.g. Homomorphic Encryption, Differential Privacy, Federated Learning, etc.).

4.7 Explainability

4.7.1 Explainability is important in fostering end-user trust in the AI-MD, during the testing and deployment stages. However, at this point, current techniques to interpret and explain algorithmic decisions are uneven and may hinder complete AI explainability. **Developers should therefore demonstrate effectiveness of the AI-MD and endeavour to ensure a sufficient level of explainability based on what their intended end-user requires:**

- a. **For Medical Practitioners:** Clarity in recommendations, algorithmic decisions of the AI-MD, and concurrence that these are in-line with their current clinical practices; and
- b. **For Patients:** Trust that the care rendered via the AI-MD is safe and that they will be “no worse off”.

4.7.2 Regardless of the specific AI-MD use-case (e.g. treatment, diagnosis/screening, patient monitoring), developers could fulfil the following criteria for explainability:

- a. Direct explanation through documentation of the AI-MD algorithm development process, covering: selection process of AI algorithm/ensemble for a particular AI-MD model, choice and characterisation of the training and validation dataset;
- b. Implicit explanation through descriptions of the intended inputs and outputs of the AI-MD, covering: range of possible AI-MD inputs (e.g. type/quality of medical images), and outputs (e.g. feature identification, heat maps); and
- c. Independent explanation through causality-based models illustrating how different inputs leads to different outputs: for instance, the Local Interpretable Model-Agnostic Explanations (LIME)¹⁰ model is useful in illustrating positive and negative causal relationships between different features, which is used in feature importance for AI medical imaging.

¹⁰ The Local Interpretable Model-Agnostic Explanations (LIME) is a technique used to explain the predictions of any AI classifier in an interpretable and faithful manner, by learning an interpretable model locally around the prediction. For more information, please refer to [“Why Should I Trust You?”: Explaining the Predictions of Any Classifier, Marco et al., 2016.](#)

BUILD

4.8 Development Standards

- 4.8.1 Proper development standards are important to ensure the safe, transparent and reproducible development of any AI software. **Developers should adhere to [HSA’s regulatory guidelines for software medical devices](#)**, which highlight the key components of maintaining proper development standards in AI-MD.
- 4.8.2 **Developers should also refer to the following examples of standards to ensure that they adopt appropriate development standards:**
- a. Risk Management Approach – ISO 14971
 - b. Quality Management System – ISO 13485
 - c. Software Development – IEC 62304

Versioning

- 4.8.3 Each time changes are made to any component of the AI-MD (e.g. within training datasets, decision-making processes, or output formats), **developers should ensure all changes to their AI-MD are properly documented and all software versions are reproducible**. One way is to assign different version numbers to the AI-MD under development (“versioning”).¹¹ This ensures that the development process remains structured, transparent and that outputs are reproducible.
- 4.8.4 Developers should ensure that their versioning records sufficiently allow the replication¹² of each version of the AI-MD, documenting changes to aspects such as (but not limited to):
- a. Parameters and hyper-parameters;
 - b. Training datasets, to capture any amendments to the data and/or how it is collected; and
 - c. Changes in training/testing environments (e.g. differences in lab vs. clinical settings).

All versioning records should be retained and made available to regulatory authorities upon request.

¹¹ For the avoidance of doubt, versioning of the AI-MD may continue even after the AI-MD has been deployed / implemented (e.g. after AI software patching).

¹² Complete replication may not be always possible for versions of AI-MD developed with stochastic algorithms (e.g. stochastic gradient descent) that explicitly use randomness during model optimisation and learning.

4.8.5 Developers of AI-MD should also comply with change notification requirements for registered AI-MD.¹³

4.9 Self-Validation

4.9.1 Developers can consider incorporating self-validation mechanisms into the AI-MD. This is so that the AI-MD can detect anomalous performance to trigger the appropriate escalations (e.g. human interventions, reverting to an earlier validated pathway, and/or to shut down the AI-MD), and to necessitate any further reviews of the AI-MD. Suggested ways to facilitate self-validation within the operational thresholds set include but are not limited to:

- a. Employing statistical models to self-validate the AI-MD; and
- b. Building in independent self-diagnostic routines.

¹³ For more details on the types of change notifications which need to be provided to HSA, please refer to flowchart 2.3 in HSA's guidance on [change notifications](#).

TEST

Once a model is developed, it needs to be validated using discrete testing datasets and protocols. This sub-section covers key points on this process.

4.10 Evaluation and Monitoring of AI-MD

Validation

- 4.10.1 **Developers should periodically evaluate and validate their AI-MD's performance to ensure it minimally meets the clinical practice baseline** (see [Section 4.4](#)), and verify the accuracy, and reproducibility of the AI-MD's algorithmic decisions.
- 4.10.2 The clinical performance of an AI-MD involves more than just technical measures of its algorithm's performance (e.g. Area under the Curve (AUC) – Receiving Operating Characteristic Curve (ROC) or Precision-Recall Curve (PRC), True Positive Rate, Positive/Negative Predictive Value, Cohen's Kappa Score, etc.). Developers should work with implementers to ensure that the actual clinical outcomes¹⁴ of the AI-MD (i.e. impact on patients when the AI-MD is introduced to their care) are measured and assessed.
- 4.10.3 **Table 3** sets out a suggested stepwise AI-MD validation approach, and the types of risks it assesses. Developers should compare results from each step with the current clinical practice baseline.

¹⁴ Clinical association of the outcomes of an AI-MD to its intended use can be established through existing evidence (e.g. literature, original clinical research, guidelines), or generating new evidence (e.g. data analysis, clinical trials) – SaMD: Clinical Evaluation, IMDRF, 2017.

Table 3: Stepwise Validation of AI-MD

Validation Steps	Risks Assessed
<p>Step 1 - Testing on Retrospective Data:</p> <ul style="list-style-type: none"> i. Ensure testing data is representative and <u>not</u> the same data used to train the AI-MD. ii. Testing the AI-MD's performance across performance variables¹⁵ (e.g. AUC score, sensitivity and specificity with minimum of 95% confidence interval). A combination of performance variables (<i>see Section 4.10.2 for reference</i>) should be considered if the AI-MD's performance is complex and cannot be evaluated via a single variable alone.¹⁶ iii. Use clearly defined testing protocols that consider: <ul style="list-style-type: none"> a) Tests for Bias – e.g. gender, image quality. b) Tests on “Boundary Conditions” – between patient inclusion/exclusion criteria, and between valid/invalid inputs. 	<ul style="list-style-type: none"> i. Performance of the AI-MD outside of training data.
<p>Step 2 - Testing on Actual Data (in parallel to ‘live’ clinical settings or actual use in the case of a clinical trial):</p> <ul style="list-style-type: none"> i. Put in place clear workflows to ensure security of the data collected and used to test the AI-MD. ii. Prepare clearly defined testing protocols as stated in Step 1. iii. Ensure appropriate approval(s) from the organisation(s) in which the testing will be carried out and seeking of patient consent. 	<ul style="list-style-type: none"> i. Performance of the AI model on a live population in a clinical setting(s) with actual input data. ii. Robustness¹⁷ of the AI model in handling actual volumes. iii. Change in actual clinical outcomes as a result of introducing the AI-MD.

¹⁵ Outcome-based performance variables should be considered if the outputs of the AI-MD is not directly measurable (i.e. classifiable outcomes, performance scores). Examples of such AI-MDs include those that are usually designed for clinical interventions (e.g. surgical AI robots, AI-based cancer treatment, etc.).

¹⁶ It has become commonplace to evaluate machine learning algorithms based on overall measures like accuracy or Area under the Curve (AUC). However, one evaluation metric may not always capture the complexity of performance. As an extreme illustration, an algorithm designed to predict a rare condition found in only 1% of a population can be extremely accurate by labelling all individuals as not having the condition. This tool is 99% accurate, but completely useless. Yet, it may “outperform” other algorithms if accuracy is considered in isolation - *STAT Report (2021): Promise and Peril - How artificial Intelligence is transforming healthcare*.

¹⁷ Ability of an AI to cope with and operate correctly (as a system or components of it) in the presence of invalid/erroneous inputs or stressful environment conditions (e.g. high volume of inputs, adversarial attacks on AI models).

4.10.4 As a good practice, developers should also consider subjecting validation results of the AI-MD and the associated testing methodology for peer-review (e.g. via peer-reviewed journals).¹⁸

- a. Developers should ensure the following components of validation are documented:
 - i. Description of test population(s) and testing dataset(s);
 - a) Similar to the training data, the testing data used should be as representative and robust for the intended use of the AI-MD.
 - ii. Description of testing protocol (including how the AI-MD deals with cases on the boundaries of the case inclusion and exclusion criteria);
 - iii. Validation results proving that the AI-MD performs no worse than established medical knowledge and processes in respect of its intended purpose; and
 - iv. Testing dates.

¹⁸ Peer-reviews of AI-MDs (especially those trained on demographic data comprising e.g. racial, gender, etc.) can mitigate against compounding bias. Peer-reviewers may include internal and external care providers, researchers, educators, and diverse groups of data scientists other than AI algorithm developers.

After successful testing, the AI-MD is ready for implementation. However, what the AI-MD should/should not be used for, and where it sits in the clinical workflows needs to be documented clearly to ensure fidelity of implementation (e.g. via user manuals and/or EULAs).

4.11 Intended Use and Workflow

Intended Use

4.11.1 Developers should clearly document the intended use of their AI-MD and provide updates to implementers (e.g. via revisions to user manuals), which should address the following information:

- a. Specific clinical use-cases which the AI-MD is designed for;
- b. Benefits and limitations of the AI-MD;
- c. Alternative options to using the AI-MD;
- d. Required inputs for the AI-MD's algorithmic decisions;
- e. Intended outputs of the AI-MD; and
- f. Boundaries of the inputs and outputs.

Intended Clinical Workflow

4.11.2 Apart from the parameters and intended use of the AI-MD, **developers should also clearly define and document how the AI-MD should be incorporated into clinical workflows.** This is important to ensure that implementers can use the AI-MD appropriately.

4.11.3 A well-defined clinical workflow should include the following information:

- a. At what stage of the clinical workflow the AI-MD is intended to be used;
- b. How implementers should respond to the variety of AI-MD's outputs (i.e. within, or beyond expected output range);
- c. The recommended degree and point(s) of human oversight for the AI-MD; and
- d. How data flows to and from the AI-MD.

5 Implementation


*This section provides guidance for AI-MD implementers, with a focus on organisational implementers¹⁹, and apply regardless of whether the AI-MD was developed in-house, in collaboration with a partner, or purchased off-the-shelf. **Figure 6** summarises the key recommendations which AI-MD implementers should take note.*

¹⁹ As mentioned in [2.3.3b](#), implementers refer to organisations or individuals who use AI-MD to deliver healthcare services (e.g. those regulated under the Private Hospitals and Medical Clinics Act (PHMCA) or future Healthcare Services Act (HCSA)).

5.1 KEY RECOMMENDATIONS


Figure 6: Key Recommendations for AI-MD Implementers

Implementers should:




4 USE

- Exercise clinical governance and oversight over the adoption and implementation of AI-MD to ensure responsible and safe implementation. [5.2.1]
- Seek approvals from their Organisational Leadership and properly document the decision to implement the AI-MD. [5.2.2]
- Track the AI-MD at the point of deployment (i.e. “ground-truthing”), to determine the “deployment baseline”. [5.2.4]
- Introduce appropriate oversight based on the intended use, workflows of the AI-MD and the clinical context. [5.3.1]
- Adopt appropriate cybersecurity policies around the AI-MD to protect and respond to threats and vulnerabilities commensurate with the AI-MD’s intended use and risks. [5.3.3]
- Train staff to operate and interpret results from the AI-MD. [5.3.4]
- Ensure that end-users (i.e. medical practitioners, patients) are clearly informed that they are interacting with an AI-MD in the delivery of care, and are provided sufficient information to make informed decisions. [5.4.1]



5 MONITOR & RESPOND

- Ensure that the AI-MD continues to perform at/above the deployment baseline, and have appropriate triggers and escalation pathways if the AI-MD’s performance falls below this baseline. [5.5.2]
- Be prepared to receive, respond to, and investigate any reports of adverse events or other device issues resulting from the use of the AI-MD. [5.5.3]



6 REVIEW

- Undertake an ad-hoc review when there are errors resulting from the use of the AI-MD, and regular reviews (e.g. yearly) to ensure the AI-MD continues to have clinical relevance and meets organisational needs. [5.6.1]
- Perform maintenance on the AI-MD at least once a year to ensure continued functionality. [5.6.2]

USE

5.2 Clinical Governance

- 5.2.1 Implementers (i.e. licensed healthcare service providers) are to meet all licensing requirements set out under the Private Hospitals and Medical Clinics Act (PHMCA) or the future Healthcare Services Act (HCSA) to ensure patient safety and welfare for all services provided by the licensee. As such, **implementers should exercise clinical governance and oversight over the adoption and implementation of AI-MD to ensure responsible and safe implementation.**

Decision to Implement

- 5.2.2 **Implementers should seek approvals from their Organisational Leadership²⁰ and properly document the decision to implement the AI-MD.** Organisational Leadership should have, or be provided with, the relevant clinical, operational and technical knowledge, to make an informed decision over the adoption and implementation of AI-MD in the institution. The decision to implement an AI-MD should be properly documented to ensure accountability.
- 5.2.3 The decision to implement an AI-MD should involve the following considerations, especially when the implemented AI-MD is purchased off-the-shelf (i.e. not developed in-house by an institution for its patients):
- Intended use and purpose of the AI-MD;
 - Expected clinical impact of using the AI-MD in terms of the outcomes of care provided, including impact on efficacy, safety and quality of care provided;
 - Perform a risk²¹ assessment of the AI-MD;
 - Alignment with clinical services planning;
 - Representativeness of AI-MD training dataset and validation methodology in relation to the patient population that the AI-MD would be used for;

²⁰ Organisational Leadership refers to those who are responsible for the overall leadership and governance of the healthcare service and varies based on the organisational size and structure. For large healthcare organisations (i.e. hospitals, nursing homes, laboratories) this could be the Board of Directors, Clinical Director, and Chairman, Medical Board, or equivalent. For solo practitioner clinics/organisations, this could be the business owner and/or the clinical lead.

²¹ Risk is defined as a function of (a) Impact and (b) Likelihood; (a) Impact – severity of patient harm if AI-MD is erroneous, and how quickly errors could be discovered and rectified; (b) Likelihood – probability for errors to occur depending on the AI-MD model and level of human oversight.

- f. Whether the AI-MD would align with the organisation’s internal ethical frameworks²² (e.g. handling of biases);
- g. Existing regulatory approvals for the AI-MD (including ensuring that the AI-MD is registered with HSA, and the AI-MD distributors or developers are in compliance with HSA’s regulatory requirements²³);
- h. Any known patient safety issues with the AI-MD;
- i. Proposed implementation plans and how the AI-MD would fit into the current clinical workflows based on the accompanying user manual/communication materials; and
- j. Identified risks of implementing the AI-MD and the mitigating measures for these risks (e.g. ability to quickly switch back to fully human care, or earlier validated AI pathways).

5.2.4 Implementers should track the AI-MD at the point of deployment (i.e. “ground-truthing”), to determine the “deployment baseline”. This is to determine if there are any differences in performance from what has been indicated by the developer, or due to the particular context that the AI-MD is deployed in (e.g. certain target group demographics, clinical workflows, quality of image resolutions). The processes and outcomes of “ground-truthing” should also be documented.

- a. If the deployment baseline is lower than the current clinical practice baseline, this may result in the AI-MD’s output making patients “worse off”, and in these situations, the AI-MD should not be deployed in its current form.

²² Organisations looking to develop/explore ethical frameworks surrounding the use of AI-MD can refer to the following papers for discussion on the ethics of data and AI in healthcare: *An Ethics Framework for Big Data in Health and Research*, Xafis et.al 2019. *AI-Assisted Decision-making in Healthcare*, Lysaght et.al 2019.

²³ Manufacturers and distributors of AI-MD registered with HSA for use in Singapore must comply with all registration requirements, including [reporting of Adverse Events](#) (AE) and [Field Safety Corrective Actions](#) (FSCA) to HSA, as well as submitting [notifications of changes](#) to registered Medical Devices.

5.3 Operational Workflows and Processes

Appropriate Oversight

- 5.3.1 **Implementers should introduce appropriate oversight based on the intended use, workflows of the AI-MD and the clinical context.** Implementers could draw reference from manufacturer's recommendations on the necessary oversight as presented in the user manual of the registered AI-MD.
- 5.3.2 If the oversight indicated in the user manual is insufficient (e.g. based on level of training and comfort of the implementer's staff working with the AI-MD), the implementer should introduce additional human oversight where necessary.

Cybersecurity

- 5.3.3 Ensuring cybersecurity requires a holistic approach, and simply securing the design of the AI-MD is insufficient. The [Healthcare Cybersecurity Essentials \(HCSE\)](#) guidelines offer basic cybersecurity measures to ensure the security, confidentiality, integrity, and availability of IT assets, systems and patient data. However, in determining the right cybersecurity posture, **implementers should adopt appropriate cybersecurity policies around their AI-MD to protect and respond to threats and vulnerabilities commensurate with the AI-MD's intended use and risks.** These policies can include, but are not limited to:
- Enhancing technology and processes to effectively protect against cybersecurity threats (e.g. safeguards to prevent unauthorised access to data processed by AI-MD);
 - Strengthening processes to detect and respond to cybersecurity vulnerabilities (e.g. conduct of regular penetration testing, use of up-to-date antivirus software and firewalls); and
 - Enhancing internal frameworks and processes to improve cybersecurity regimes (e.g. regular reviews of cybersecurity posture, establish clear channels for staff reporting of suspicious activities).

Manpower Resources

5.3.4 For AI-MDs that are intended to be used alongside healthcare professionals, **implementers should ensure that staff are trained to operate and interpret results from their AI-MD.** These include, but are not limited to:

- a. Train staff on how to use the AI-MD (e.g. how to properly provide input data, identifying signs of failures/alerts from the AI-MD);
- b. Train staff in working alongside the AI-MD (e.g. how to interpret outputs from Clinical Decision Support (CDS) tools, how to incorporate these tools into individual clinical decisions);
- c. Educate staff on the risks and limitations of the AI-MD (e.g. if the AI-MD is built for specific use-cases only, such as interpreting images from specific optical equipment and with specific image orientations/views, or only for specific patients);
- d. Educate staff on how and when to activate contingency *plans* (see [Section 5.5.2](#));
- e. Improve staff's cybersecurity awareness and competencies (e.g. regular staff training to educate on appropriate handling of datasets/platforms, email hygiene and password protection, etc.); and
- f. Educate staff on communicating the use of the AI-MD to medical practitioners/patients, as appropriate (see [Section 5.4](#) on *End-User Communication that includes resolving queries on the AI-MD use and output*).

5.4 End-User Communication

Transparency

- 5.4.1 **Implementers should ensure that end-users (i.e. medical practitioners, patients) are clearly informed that they are interacting with an AI-MD in the delivery of care, and are provided sufficient information to make informed decisions** of whether to continue using the AI-MD, or if due to patient/condition exclusion, care should be with a clinician instead. **Table 4** details the suggested information that implementer should communicate to end-users.
- 5.4.2 For the avoidance of doubt, implementers of AI-MD must continue to comply with all applicable laws and requirements (etc.), including the seeking of appropriate patient consent²⁴, with respect to the healthcare professional's duty in giving medical advice, and the applicable ethical code and guidelines.

²⁴ Consent required when using AI-MD should be no different from consent taken for other medical procedures performed by actual physicians. Implementers can refer to Section C6 of the Singapore Medical Council's *Ethical Code and Ethical Guidelines* for details on the principles of consent.

Table 4: Suggested Information for End-User Communication

Type of End-User	Suggested Information to share with the End-User	Examples
Medical Practitioner	<ul style="list-style-type: none"> i. Clarity that they are interacting with an AI-MD. ii. Limitations of the AI-MD (e.g. inclusion/exclusion criteria) where required. iii. Date of most recent AI-MD audit. iv. Contact for institution POC (implementer) so they may obtain specific AI-MD performance information if required (e.g. intended function, risks vs. benefits, clinical-causal-relationships between inputs and outputs, accuracy range). 	Medical imaging AI that is used by medical practitioners as clinical decision support to identify cancerous lesions in CT/MRI scans.
Patient	<ul style="list-style-type: none"> i. Clarity that they are interacting with an AI-MD. ii. Limitations of the AI-MD (e.g. inclusion/exclusion criteria) where required. iii. Name of the specific institution that has deployed the AI-MD (for accountability). iv. Contact for institution POC (implementer) for adverse events, to seek clarification on the use of the AI-MD, or request for an in-person interaction. 	App-based AI that provides patients with personalised guidance on lifestyle changes for weight loss and management of obesity-related chronic diseases.

MONITOR & RESPOND

5.5 Post-Deployment Monitoring

5.5.1 Apart from validating the AI-MD's performance pre-market (during the development phase of the AI-MD), implementers also need to monitor their AI-MD's performance post-deployment to ensure the continued safety, efficacy, and robustness of the AI-MD's model.

5.5.2 **Implementers should ensure that the AI-MD continues to perform at/above the deployment baseline and have appropriate triggers and escalation pathways if the AI-MD's performance falls below this baseline.**

This could be caused by changes in the target population or input collection methods (e.g. if tools used to collect the inputs have been updated). Monitoring ensures that longer-term performance continues to remain in-line with this established baseline. To do this, implementers should:

- a. Identify key monitoring outcomes and monitoring frequency;
- b. Set input/output thresholds for these outcomes; and
- c. Put in place self-validation mechanisms into the AI-MD to trigger the following possible escalation pathways where thresholds are breached:
 - i. Initiating human intervention (i.e. "safe-fails");
 - ii. Reverting to an earlier validated pathway; and/or
 - iii. Shutting down the AI-MD.

5.5.3 **Implementers should also be prepared to receive, respond to, and investigate any reports of adverse events or other device issues resulting from the use of the AI-MD.²⁵**

- a. Immediately respond: Implementers should develop processes and contingency plans to ensure that staff respond timely and appropriately during AI-MD failure, where patients' safety and welfare are potentially compromised. Contingency plans should **include shutting down the AI-MD and switching to analogue protocols** (i.e. where the AI-MD is not involved) so that patients would still receive safe and appropriate care.
- b. Investigate and understand: Implementers should also aim to identify and investigate any differences in outputs. One reason could be drifts in either model performance or clinical impact, which will require a review. Such reviews might investigate whether datasets continue to be representative of the current patient population, any changes in existing care processes, or patient demographics.

²⁵ Please refer to HSA's guidance on [Adverse Event](#) reporting.

REVIEW

5.6 Review and Tracking

- 5.6.1 **Implementers should undertake an ad-hoc review when there are errors resulting from the use of the AI-MD, and regular reviews (e.g. yearly) to ensure the AI-MD continues to have clinical relevance and meets organisational needs.** Implementers should also actively update developers when significant issues have been identified from these reviews, and if any mitigation measures have been taken. When doing so, implementers should consider the following components in **Table 5**.

Table 5: Recommended Components for Ad-hoc and Regular Reviews of AI-MD

Frequency	Recommended Components
Ad-hoc	Having a “Morbidity and Mortality” (M&M) conference when there are: <ul style="list-style-type: none">i. Patient safety issues arising from the use of AI-MD.ii. Inconsistencies in the recommendations between the Medical Practitioner and AI-MD.
Regular (e.g. yearly)	<ul style="list-style-type: none">i. Having a performance review (e.g. sampling 10% of outputs from a Diabetic Retinopathy screening AI-MD and comparing it to the deployment baseline).ii. Reviewing past ad-hoc M&Ms arising from the use of the AI-MD, and considering the potential of model drifts due to:<ul style="list-style-type: none">a. Changes to patient inclusion/exclusion criteria.b. Changes to clinical workflows.c. Changes in actual patient population compared to patient population of AI-MD’s training dataset.iii. Based on the above, consider if the current implementation mitigations are sufficient, need to be enhanced, or if the AI-MD needs to be taken offline and re-built.

Maintenance

5.6.2 Implementers should ensure that maintenance is performed on their AI-MD at least once a year to ensure continued functionality. This should include:

- a. Identifying potential cybersecurity vulnerabilities of their AI-MD;
- b. Reviewing the information flow within their AI-MD (e.g. how information gets transmitted from image collection devices to AI algorithms, and subsequently to the report provided to medical practitioners) to ensure that the AI-MD processes the right inputs to provide the output(s) as expected;
- c. Updating their AI-MD versions (if applicable); and
- d. Verifying of model performance and assessment.

6 Emerging Developments in AI

This section provides recommendations for emerging developments in AI which are applicable to both the development and implementation aspects.

6.1 Continuous Learning AI-MD and Deployment

- 6.1.1 While most AI-MDs today are static/locked (i.e. do not automatically incorporate new data into the algorithm), one advantage is the introduction of the AI-MD's ability to continuously learn and adapt during its deployment. This could result in immediate updates to the inputs and outputs of the AI-MD. Appropriate controls for this type of AI-MD have been widely discussed by both local and international regulators.^{26, 27}
- 6.1.2 Implementers should ensure that the risks of deploying AI-MD with continuous learning capabilities are identified and reasonably mitigated beforehand. These risks include, but are not limited to:
 - a. Inappropriate²⁸ initialisation parameters;
 - b. Biased or unrepresentative input data affecting the algorithms upon which the AI-MD's model is built;
 - c. A possible inability to fully validate updates to the model's algorithms (in order to ensure clinical validity and accuracy), due to the AI-MD's continuous learning capabilities;
 - d. Abnormal behaviour (e.g. maliciously introduced data), and/or end-user manipulations (e.g. introducing rare yet valid and important data).
- 6.1.3 Before implementing AI-MD capable of continuous learning, developers and implementers are encouraged to consider using the following alternative:
 - a. Deploy locked AI algorithms at user facilities while allowing the AI algorithm to learn in parallel, and only implement the updated algorithm post-learning, once sufficient rigorous checks and validation have been conducted.²⁹

²⁶ The US Food and Drug Administration (FDA) distributed a discussion paper titled *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning [AI/ML] – Based Software as a Medical Device [SaMD]*, which provides suggestions for a software lifecycle approach to regulating AI.

²⁷ The Singapore Health Sciences Authority (HSA) has also discussed continuous learning in a software lifecycle approach in their *Regulatory Guidelines for Software Medical Devices*.

²⁸ Initialisation parameters are known to auto-calibrate based on the model's 'learning' to converge on the loss-minimising point (i.e. the lowest point on the loss function so that few predictions would have deviated from real-life results). Inappropriate initialisation parameters can lead to divergence instead of convergence, where the model does not operate at the most accurate point and affects the outputs produced.

²⁹ HSA requires change notifications to be submitted for any changes which affect the safety, quality, or efficacy of the AI-MD. Further information can be found [here](#).

- 6.1.4 Interested implementers and developers of continuous learning AI-MD should ensure there are sufficient safeguards to maintain the quality, safety and efficacy of the AI-MD during deployment. These safeguards should also be submitted to HSA as part of AI-MD product registration and when submitting change notifications. Developers should consider safeguards such as using ensemble methods to enhance the overall accuracy of AI-MDs and mitigate the risks of poor decisions made by the AI-MD, by adding the decisions made by the retrained AI-MD to an ensemble of other AI algorithms with validated performance.
- 6.1.5 Introduce protocols to log the factors that changed the model, and ensure that the retrained AI-MD and any subsequent versions are assigned new version numbers, are traceable, and can be reverted to a previous version when necessary. In addition, for every version, to also log the results of the new version, if there is a difference compared to the validated and approved model.
- 6.1.6 Developers should introduce controls to review the newly trained and deployed AI-MD at high frequencies. This is to ensure the AI-MD:
- a. Is within a specified performance range, taking reference from the deployment baseline (*see Section 5.2.4*); and
 - b. Alerts end-users when performance drops below the range, so that they can employ mitigation measures (e.g. reverting to an earlier proven version, stop operations, etc.).
- 6.1.7 Some controls include to:
- a. Automate validation;
 - b. Record any input changes which led to a drop in performance; and
 - c. Frequently review the input data to monitor (e.g. by sampling) and be able to investigate as necessary any instances of model drift.
- 6.1.8 Developers and implementers should agree on appropriate triggers and escalation pathways should the AI-MD be no longer performing at or above the deployment baseline.

6.2 Synthetic Data

- 6.2.1 The creation and usage of synthetic data is another technique in obtaining necessary inputs to train AI algorithms. The implementation of algorithms such as Generative Adversarial Nets (GAN) provides an increasingly accessible way for developers to build hyper-realistic synthetic datasets that may no longer be distinguishable from real data.
- 6.2.2 In addition to the other recommendations on data in these guidelines, developers who wish to use synthetic data in the training and development of their AI-MD algorithms should ensure that the synthetic data is:
- a. Labelled as synthetic data, with the method and date of creation clearly documented, and are reproducible;
 - b. Equally de-identified as the original datasets used to generate the synthetic data;
 - c. Adequately reproducible; and
 - d. Verified by a clinical practitioner(s) to be clinically viable.
 - i. This verification can be done on a representative subset of synthetic data, rather than the entirety of the dataset.
 - ii. Clinical viability is based on the clinical practitioner's judgement on whether such data can be observed in a real-world setting.
 - e. Deployed to complement an existing data set (but not replacing it), with the limitations of the data set clearly defined. Synthetic data typically can reproduce only very specific attributes of a disease accurately and are unlikely to reproduce all aspects of a condition correctly. If AI is trained only with the limited range of synthetic reproductions, the AI is at risk of being inherently biased.

7 Sharing your Feedback

- 7.1.1 The AIHGle sets out good practice recommendations and complements the current HSA regulations for the safe and responsible development and implementation of AI in healthcare. Given the rapid developments in AI, the AIHGle is also meant to be a “living” document that will be updated in parallel with the developments of this field.
- 7.1.2 For feedback pertaining to the AIHGle, please access the feedback form by clicking on [this link](https://go.gov.sg/aihgle-feedback) or scanning the QR code below.



8 Case Study – Singapore Eye Lesion Analyser Plus (SELENA+)

What is the SELENA+?

- SELENA+ is an AI-based **deep learning system** used to analyse retinal images for signs of **3 major diabetic eye diseases**:
 - 1) **Diabetic Retinopathy (DR)**;
 - 2) **Glaucoma**, and
 - 3) **Age-related Macular Degeneration**.
- SELENA+ **potentially improves the productivity of human graders by ~50%**, with results produced within shorter turn-around times
 - SELENA+ can **provide results within minutes for patients with non-referable conditions**, while **patients with referable conditions will be escalated to human graders for confirmation**.

*Early
2000s*

Co-developed by a team of researchers and clinicians from the Singapore National Eye Centre (SNEC)'s Singapore Eye Research Institute (SERI) and National University of Singapore (NUS)'s School of Computing.

2017

Published in the Journal of American Medical Association (JAMA) - [*Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images from Multi-ethnic Populations with Diabetes*](#)


2019



Approved by HSA as a [**Class B Medical Device**](#).


*2021
onwards*


As part of the Singapore Integrated Diabetic Retinopathy Programme (SiDRP), SELENA+ will be **deployed to all local polyclinics**, and its **deep learning algorithms** may also be **extended to develop a predictive risk assessment model for cardiovascular disease** in the near future.


The SELENA+ **development and implementation processes were well-governed:**


Phase	Development	Measures Taken to Develop SELENA+
 DESIGN	<ul style="list-style-type: none"> Seek clinical inputs relevant to the intended use of the AI-MD when developing the AI-MD. 	<ul style="list-style-type: none"> Ophthalmologists (Retina Specialists) were included in the development team, to advise on the clinical problem statement, representativeness of AI-MD training and test datasets, algorithm testing approach, and user manual.
	<ul style="list-style-type: none"> Seek end-user inputs for a holistic AI-MD design and development process. 	<ul style="list-style-type: none"> Inputs on the development of SELENA+ was sought from end-users (i.e. Ophthalmologists and human graders) during SELENA+ workgroup meetings, User Acceptance Testing (UAT) sessions, etc.
	<ul style="list-style-type: none"> Determine the current clinical practice baseline to ensure that the AI-MD's performance is minimally no worse off than current practice. 	<ul style="list-style-type: none"> Current clinical practice baseline is equivalent to the assessed standard of care provided by professional non-physician human graders.
	<ul style="list-style-type: none"> Ensure representativeness of datasets to reduce unintended bias. 	<ul style="list-style-type: none"> SELENA+ was trained on images collected in the SiDRP and validated in more than 10 external datasets (a total of ~71,000 retinal images from ~14,000 patients) with various ethnicities (e.g. Local (i.e. Chinese, Indian, Malay), Latino, African-Caribbean, and Caucasian).
	<ul style="list-style-type: none"> Document all biases and/or limitations identified in an AI-MD and rectify them if possible. 	<ul style="list-style-type: none"> SELENA + is mainly trained on local Singaporean population, and thus, the performance will be skewed towards the Asian and pigmented population. For the disease classification, SNEC has divided that into 'Referable' versus 'Non-referable'; however, SELENA+ does not have severity-specific classification. Thus, should patients require a specific diagnosis, they are advised to visit an ophthalmologist still.


Phase	Development	Measures Taken to Develop SELENA+
 DESIGN	<ul style="list-style-type: none"> • Ensure AI-MD can prevent, detect, respond and where possible, recover from foreseeable cybersecurity risks. 	<ul style="list-style-type: none"> • SELENA+: <ul style="list-style-type: none"> ○ Relies on its host system (SiDRP) to detect/alert potential cybersecurity vulnerabilities. The host system also complies with MOH's HealthTech Instruction Manual-ICT Security Policies (HIM-ISP). ○ Does not retain input data nor does it keep residual data (i.e. no Database Management System (DBMS) that potentially could be exploited). ○ Encrypts its AI models.
	<ul style="list-style-type: none"> • Demonstrate effectiveness of the AI-MD and endeavour to ensure a sufficient level of explainability based on what their intended end-user requires. 	<ul style="list-style-type: none"> • SELENA+ was published in the international peer-reviewed Journal of American Medical Association (JAMA) (Ting et al. 2017). • While SELENA+ has the capability to demonstrate visualisation techniques, the different visualisation techniques in deep learning are not yet mature enough to consistently demonstrate and support AI explainability. • Thus, SNEC is adopting the universally and locally acceptable clinical standards in terms of screening performance (i.e. minimum of ~80% sensitivity and ~80% specificity).
 BUILD	<ul style="list-style-type: none"> • Adhere to HSA's regulatory guidelines for software medical devices. 	<ul style="list-style-type: none"> • Approved by HSA in 2019 for meeting the regulatory requirements of a Class B Medical Device.
	<ul style="list-style-type: none"> • Adopt appropriate development standards (e.g. ISO 13485, ISO 14971, IEC 62304). 	<ul style="list-style-type: none"> • Obtained ISO 13485 certification in Oct 2019.
	<ul style="list-style-type: none"> • Document properly all changes to AI-MD and ensure all software versions are reproducible. 	<ul style="list-style-type: none"> • Change notifications to SELENA+ were documented, submitted and approved by HSA.

Phase	Development	Measures Taken to Develop SELENA+
 TEST	<ul style="list-style-type: none"> Evaluate and validate periodically the AI-MD's performance to ensure it minimally meets the clinical practice baseline. 	<ul style="list-style-type: none"> Evaluated the diagnostic performance/accuracy, specificity and sensitivity of SELENA+ against the manual grading assessment by professional non-physician human graders.
	<ul style="list-style-type: none"> Document the intended use of the AI-MD. 	<ul style="list-style-type: none"> Documented in SELENA+'s user manual and Standard Operating Procedures (SOPs) of selected polyclinics with SELENA+ deployed.
	<ul style="list-style-type: none"> Evaluate and validate periodically the AI-MD's performance to ensure it minimally meets the clinical practice baseline. 	<ul style="list-style-type: none"> Evaluated the diagnostic performance/accuracy, specificity and sensitivity of SELENA+ against the manual grading assessment by professional non-physician human graders.

Phase	Implementation	Measures Taken to Implement SELENA+
 USE	<ul style="list-style-type: none"> Exercise clinical governance and oversight over the adoption and implementation of AI-MD to ensure responsible and safe implementation. 	<ul style="list-style-type: none"> Trial Implementation of SELENA+ was done in partnership with all polyclinics who hold existing Private Hospitals & Medical Clinics Act (PHMCA) licences that require proper clinical governance structures.
	<ul style="list-style-type: none"> Seek approvals from Organisational Leadership and properly document the decision to implement the AI-MD. 	<ul style="list-style-type: none"> Proof-of-Concept and evaluation of SELENA+ were done in consultation with the Project Directors of SiDRP (Medical Directors of SNEC and Tan Tock Seng Hospital (TTSH)) and Clinical Service Directors of the polyclinic clusters. Final approval for implementation of SELENA+ was signed off by both Project Directors of SiDRP.
	<ul style="list-style-type: none"> Track the AI-MD at the point of deployment (i.e. “ground-truthing”), to determine the “deployment baseline”. 	<ul style="list-style-type: none"> SELENA+’s post-implementation diagnostic performance is audited against the current assessed standard of care.
	<ul style="list-style-type: none"> Introduce appropriate oversight based on the intended use, workflows of the AI-MD and the clinical context. 	<ul style="list-style-type: none"> SELENA+’s primary assessments were passed to secondary human graders to check accuracy and validity, and to determine if further medical intervention is required (e.g. referrals to eye specialists).
	<ul style="list-style-type: none"> Adopt appropriate cybersecurity policies around the AI-MD to protect and respond to threats and vulnerabilities commensurate with the AI-MD’s intended use and risks. 	<ul style="list-style-type: none"> IHiS cybersecurity policies were in-place as required by the polyclinics for implementing SELENA+ (i.e. MOH’s HealthTech Instruction Manual-ICT Security Policies (HIM-ISP)).

Phase	Implementation	Measures Taken to Implement SELENA+
 USE	<ul style="list-style-type: none"> Train staff to operate and interpret results from the AI-MD. 	<ul style="list-style-type: none"> Prior to using SELENA+, polyclinic staff were trained and briefed on its benefits, risks and limitations, how to correctly interpret outputs and appropriate next steps, when/how to activate contingency plans, etc. In view of the prevailing COVID-19 restrictions, staff trainings/briefings were conducted both physically and remotely e.g. training sessions for Reading Centre Staff conducted by team leads at SNEC's Reading Centre, email circulars sent to Polyclinic Directors (i.e. Nursing leads were subsequently briefed in turn) to communicate the impact of SELENA+ on current clinical workflow, dedicated SELENA+ hotline for Nurses, etc.
	<ul style="list-style-type: none"> Ensure that end-users (i.e. medical practitioners, patients) are clearly made aware that they are interacting with an AI-MD in the delivery of care, and are provided sufficient information to make informed decisions. 	<ul style="list-style-type: none"> Doctors were informed upfront on the use of SELENA+ as part of the care delivery and the appropriate point-of-contact to seek further performance-related information on SELENA+.

Phase	Implementation	Measures Taken to Implement SELENA+
 MONITOR	<ul style="list-style-type: none"> Ensure that the AI-MD continues to perform at/above the deployment baseline, and have appropriate triggers and escalation pathways if the AI-MD's performance falls below this baseline. 	<ul style="list-style-type: none"> Continuous monitoring ensures SELENA+'s diagnostic performance would not be compromised by potential/undetected model drifts (e.g. due to changes in profiles of intended patient populations). This is done by comparing SELENA+'s performance against the current assessed standard of care (i.e. professional non-physician human graders) at 2 frequencies: <ul style="list-style-type: none"> Ongoing monitoring (daily): A selected proportion of SELENA+ cases reported as 'Normal' will be escalated to human graders for confirmation. This escalation process is automated as part of SELENA+'s design. Periodic monitoring (3-6 months intervals): Manual extraction of SELENA+'s outputs and assessments of human graders for comprehensive review and analysis. Key monitoring outcomes, thresholds, and frequencies are well-defined to activate human intervention, revert to an earlier validated AI-MD pathway, or shut down the AI-MD, when necessary.
	<ul style="list-style-type: none"> Be prepared to receive, respond to, and investigate any reports of adverse events or other device issues resulting from the use of the AI-MD. 	<ul style="list-style-type: none"> Staff are trained to identify signs of failure of alerts from SELENA+ and escalation protocols for reporting of any adverse events to HSA.

Phase	Implementation	Measures Taken to Implement SELENA+
 REVIEW	<ul style="list-style-type: none"> Undertake an ad-hoc review when there are errors resulting from the use of the AI-MD, and regular reviews (e.g. yearly) to ensure the AI-MD continues to have clinical relevance and meets organisational needs. 	<ul style="list-style-type: none"> Regular reviews are conducted to ascertain that the clinical workflow, diagnostic performance, and risk mitigation measures continue to be effective and relevant, and that SELENA+ supports the national objectives of chronic disease prevention and management.
	<ul style="list-style-type: none"> Perform maintenance on the AI-MD at least once a year to ensure continued functionality. 	<ul style="list-style-type: none"> IHiS engaged technical vendor of SELENA+ to perform technical maintenance on SELENA+ at regular intervals. IHiS will also conduct checks on the AI platform and SiDRP.