IMDRF (Artificial Intelligence/ Machine Learning-enabled) WG Comment Form

Date: August 30, 2024	Document: Good machine learning practice for medical device development: Guiding principles

Name/Organization	Line number	Section	Comments	Proposed change	Resolution
International Society for Pharmaceutical Engineering **(ISPE)**	General	General	ISPE supports IMDRF's efforts to advance the development of GMLPs. As IMDRF advances this topic, we recommend detail and transparency on how IMDRF will collaborate and strategically engage with specific international standards organizations, conformity assessment bodies, and other collaborative bodies on these important issues. This will ensure that IMDRF's stakeholders will be aligned and well-informed in the development of regulatory policies and guidelines for AI/ML-enabled medical devices. Additionally, we recommend IMDRF develop additional guidance that discusses the connection between GMLPs and quality management system (QMS), as well as risk management considerations for medical devices. For instance, the IMDRF/SaMD WG/N23 guidance titled "Software as a Medical Device (SaMD): Application of Quality Management System" provides detailed guidance on implementing established and widely accepted QMS practices for SaMD. Furthermore, we recommend the development of additional guidance that discusses the important considerations for Total Product Lifecycle (TPLC) approaches for AI/ML-enabled medical devices. For example, it would be beneficial for stakeholders to have guidance that outlines the appropriate measures for monitoring model performance and re-training throughout the lifecycle of AI/ML-enabled medical devices. We recommend IMDRF build on the joint efforts between FDA, Health Canada, and MHRA, which published a joint		

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Name/Organization	Line number	Section	Comments	Proposed change	Resolution
			document identifying 5 guiding principles for predetermined change control plans to ensure the ongoing safety and effectiveness of devices throughout the device's TPLC		
ISPE		Introduction	There is no mention of validation in case that GMLP is used in a GxP use case. Include discussion of validation as part of the good software engineering	Amend current text, "Moreover, generative AI may heighten the role of GMLP, including fundamental software engineering practices and validation."	
ISPE		Introduction	Software of Unknown Provenance (SOUP) is introduced as a new concept in this IMDRF document without being adequately defined. We recommend IMDRF clarify the SOUP term and definition and recommend recognizing the IEC 62304 definition for SOUP: "SOFTWARE ITEM that is already developed and generally available and that has not been developed for the purpose of being incorporated into the MEDICAL DEVICE (also known as "off-the-shelf software") or SOFTWARE ITEM previously developed for which adequate records of the development PROCESSES are not available." This will ensure consistency in the use of certain terminology in IMDRF guidance.		
ISPE		Introduction	We recommend IMDRF provide additional clarification in the introduction of the document regarding the expected adoption, implementation, and application of these guiding principles within the regulatory decision-making process. This clarification should provide a discussion on how		

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Name/Organization	Line number	Section	Comments	Proposed change	Resolution
			these principles can be effectively integrated into medical device development programs and existing regulatory processes and procedures. By providing this clarity, stakeholders and regulatory authorities will have a clearer understanding of how to adopt these principles and promote consistent and effective application across different jurisdictions.		
ISPE		1 Introduction	Even if there are scientific advances on the horizon to provide more insights into foundation models, the declaration to determine a Foundation Model should only be allowed in the boundaries of a risk-based approach.	Add to the current text (in red): "AI may also pose a more fundamental challenge with respect to demonstrating device performance. The regulatory science of measuring performance as well as characterizing and detecting errors in these models is maturing to meet this challenge. Nevertheless, the declaration of Foundation Models as SOUP should be based on a risk-based approach."	
ISPE	Line# 18	2	Section-2 References a 'Draft' document (IMDRF/MC/N79 DRAFT: 2023 Guiding Principles to Support Medical Device Health Equity). Document should only reference relevant documents that are approved and list them under Reference section.	Remove reference to unapproved 'draft' document from page 5 of 9, unless "IMDRF/MC/N79 DRAFT: 2023 Guiding Principles to Support Medical Device Health Equity" is approved prior to the approval and provisioning of "IMDRF/AIWG/N73 Good machine learning practice for medical device development: Guiding principles"	
ISPE	Line# 20	2	Section-2 References a 'Draft' document (IMDRF/SaMD WG/N81 DRAFT:2024 Medical Device Software: Considerations for	Remove reference to unapproved 'draft' document from page 5 of 9, unless "IMDRF/SaMD WG/N81 DRAFT:2024	

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Name/Organization	Line number	Section	Comments	Proposed change	Resolution
			Device and Risk Characterization). Document should only reference relevant documents that are approved and list them under Reference section.	Medical Device Software: Considerations for Device and Risk Characterization" is approved prior to the approval and provisioning of "IMDRF/AIWG/N73 Good machine learning practice for medical device development: Guiding principles"	
ISPE	Principle 2	3	User requirements should be included as part of model design fundamentals.	Modify text:"Model design is implemented and maintained with attention to the fundamentals: robust software engineering practices, user requirements and usability…"	
ISPE	Principle 1	3	We recommend removing the phrase "well understood" because it introduces ambiguity and confusion regarding the expectations for the device's intended use/intended purpose. The current wording fails to clarify whether this phrase pertains solely to developers and manufacturers of the device or if it includes both end users and patients. This lack of clarity can lead to misinterpretation of this guiding principle and lead to confusion with its application. We also recommend removing the phrase "clinically meaningful needs" because this expectation is subject to interpretation and may depend on the stakeholder's perspective and their corresponding context of use in the clinical workflow. For example, what may be deemed "clinically meaningful" to a	~~**The device's intended use/intended purpose is well understood, and**~~ **Multi-disciplinary expertise is leveraged throughout the total product life cycle for the device's intended use/intended purpose:** In-depth understanding of a medical device's intended use/ intended purpose including context of use within the clinical workflow, system boundaries, and the desired benefits and associated patient risks, can help ensure that AI-enabled medical devices, address the device's intended use/intended purpose ~~clinically meaningful needs~~ over the total product life cycle of the device. The purpose of AI in the context of the device's intended use should be defined. Multi-disciplinary expertise provides context-	

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Name/Organization	Line number	Section	Comments	Proposed change	Resolution
			clinician may vary significantly from a patient's unique perspective and individual experience. In line with the IMDRF guidance document, "Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices," it is crucial to emphasize a device's intended use rather than describing "clinically meaningful needs." This guiding principle should align with the IMDRF Essential Principles document, which states, "Medical devices and IVD medical devices should achieve the performance intended by their manufacturer and should be designed and manufactured in such a way that, during intended conditions of use, they are suitable for their intended purpose. They should be safe and perform as intended, should have risks that are acceptable when weighed against the benefits to the patient, and should not compromise the clinical condition or the safety of patients, or the safety and health of users or, where applicable, other persons." By removing the phrase "clinically meaningful needs," we can adhere to the principles outlined in the IMDRF Essential Principles document, ensuring clarity and consistency in evaluating the intended use and purpose of medical devices. We also recommend including considerations of "**system boundaries**" in principle 1, which draws upon risk management principles from ISO 14971:2019 - Medical Devices - Application of Risk Management to Medical Devices, where system boundaries include identifying the limits of a	specific insight and experience, informs the intended use/ intended purpose, and enhances the safety and effectiveness of the device.	

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Name/Organization	Line number	Section	Comments	Proposed change	Resolution
			medical device system, including physical and functional interfaces, which can affect risk analysis. Additionally, not only the intended use of the device should be known but also the purpose of AI in the context of the intended use.		
ISPE	Principle 2	3	Explainability is described intrinsically as part of "…communicate decisions and rationale," Based on previous discussion, we would recommend using the word explainability in this context	Add "(i.e., explainability) after "communicate decisions and rationale"	
ISPE	Principles #2.	3,	Data integrity is not listed in that sentence and is actually a separate concept (ALCOA+) than data quality, but they are not the same thing.	Add "data integrity" after "data quality assurance" in the list of fundamentals to ensure both independent concepts are included, especially since integrity is listed in the next sentence.	
ISPE	Principle 3		We recommend this change because there may be circumstances where datasets are not entirely representative of the intended patient population because it is not feasible or necessary based on what is known about the device's intended use/intended purpose. For instance, certain patient characteristics like age, gender, sex, race, ethnicity, geographical location, or specific diseases may be irrelevant to the device's intended use and generalizable performance. Moreover, collecting clinical study data for rare diseases or conditions	**Clinical study participants and/or datasets are representative of the intended patient population:** Data collection protocols should consider ~~ensure~~ ~~that~~ the relevant characteristics of the intended patient population (for example, in terms of age, gender, sex, race, ethnicity, geographical location, disease), intended use environment, and measurement inputs are sufficiently represented in a sample of adequate	

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Name/Organization	Line number	Section	Comments	Proposed change	Resolution
			with limited patient populations poses practical challenges. In such circumstances, it is essential to allow sponsors the flexibility to provide a scientifically and risk-based justification for utilizing a specific dataset for the intended patient population. This approach acknowledges that strict adherence to representative datasets may not always be feasible or necessary, if the sponsor can provide a robust rationale for their chosen dataset. By allowing this flexibility, we can accommodate the unique challenges posed by certain patient populations and rare diseases, while still ensuring the safety and performance of AI/ML-enabled medical devices. We also recommend IMDRF clarify "generalizable performance" to describe how this may apply across the intended patient population. The concept of generalizability and corresponding regulatory expectations for certain AI models in the context of a device's intended use/intended purpose and intended patient population needs to be appropriately defined so developers and manufacturers can make reasonable steps to ensure model safety and performance.	size in the training and test datasets and/or clinical study, so that results can be reasonably generalized to the intended population of interest. These are fundamental for clinical evaluations and important to manage any unintended bias or dataset drift, promote appropriate and generalizable performance across the intended patient population, assess usability, and identify circumstances and subgroups where the model may underperform including over time.	
ISPE	Principle 4	3	The use of generative AI may include foundation models, which are used with prompting strategies only. Here, the meaning of training data sets (on the side of the supplier of the foundation model) and the test set (on the side of the user) may require a comment regarding reasonable efforts to ensure their independence.		

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Name/Organization	Line number	Section	Comments	Proposed change	Resolution
ISPE	Principle 4	3	. The term “external validation” is not universally understood. Recommend deleting “external” since all validation efforts should be proportional to risk.	Training datasets are independent of test sets: Training and test datasets are selected and maintained to be appropriately independent of one another. All potential sources of dependence, including patient, data acquisition, and site factors, are considered and addressed to assure independence. The extent of ~~external~~ validation should be proportionate to risk.	
ISPE	Principle 5	3	Principle #5 should be expanded to include both reference standards and reference methods. For example, if the model is emulating a laboratory test, the test method would be the reference rather than samples from specific patients.		
ISPE	Principle 5	3	We recommend IMDRF clarify what is meant by “accepted methods” and “accepted reference standards” or avoid the terms. As worded, this could be interpreted to be regulatory authority acceptance of specific methods and reference standards. However, it is important to consider a broader perspective that includes acceptance by trained expert communities, standards development organizations, and other third-party organizations. The sponsor should be able to provide a rationale for the choice of a reference standard and the general approach taken to ensure model development and testing provide robustness and	**Selected reference standards are fit-for-purpose:** ~~Accepted methods for developing a fit-for-purpose~~ Reference standards should be fit-for-purpose and ensure that clinically relevant and well characterized data are collected and that the limitations of the reference are understood. This includes documentation of the rationale of the choice of reference standard based on the device’s intended use/intended purpose and assessment of its suitability to address the intended use environment. If available, ~~accepted~~	

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Name/Organization	Line number	Section	Comments	Proposed change	Resolution
			generalizability for the intended patient population. These considerations should be based on an understanding of the device's intended use/intended purpose. The proposed text eliminates "accepted".	appropriate reference standards in model development and testing that promote and demonstrate model robustness and generalizability across the intended patient population are used. The selection of reference standards should be based on broad consensus where available and appropriate expertise.	
ISPE	Principle 6	3	The term "available data" used in this principle is open to ambiguity and interpretation. It is important to note that stakeholders may not have equal access to "available data." Therefore, we recommend IMDRF clarify the definition and corresponding expectations of "available data." This clarification will help stakeholders understand how model choice and design should be appropriately tailored to align with the intended use and purpose of the device. We recommend removing the term "clinical" from discussion of benefits and risks related to the product because there could be certain types of non-clinical benefits and risks that should be considered. We also suggest modifying the discussion of population subgroups to include "***any*** subgroups" instead of assuming all populations contain relevant subgroups. This is because certain population subgroups may not be relevant or may not be adequately understood or identified based on the current understanding of a particular disease or condition for a given device's intended use/intended purpose.	**Model choice and design are tailored to the available data and the intended use/ intended purpose of the device:** Model choice and design are evaluated and shown to be suited to the available data and support the active mitigation of known risks, like overfitting, performance degradation, and security risks. The ~~clinical~~ benefits and risks related to the product are well understood, used to derive clinically meaningful performance goals for testing, and support the product's safety and effectiveness in achieving its intended use/ intended purpose. Considerations include the impact on both the overall intended patient population and *any* ~~its~~ subgroups as well as uncertainty and variability in the device inputs, outputs, and clinical use conditions.	

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Name/Organization	Line number	Section	Comments	Proposed change	Resolution
ISPE	Principle 7	3	The currently worded, “Human factors considerations are addressed,” could be incorrectly interpreted by stakeholders and regulatory authorities to assume that human factors validation testing is needed for all AI/ML-enabled medical devices. However, this consideration should be based on thoughtful and appropriate human factors and usability engineering processes, including use-related risk analyses which may inform whether human factors validation testing is needed. These considerations should be part of a larger risk management framework.	**Performance is assessed with a focus on the human-AI team in the intended use environment:** The performance of the model outputs is assessed in the context of the intended use environment and clinical workflow. Consider appropriate human factors and usability engineering processes, including use-related risk, user skills, user expertise, user understanding of the model outputs and limitations, and user error, for normal use and reasonably foreseeable misuse.	
ISPE	Principle 7 or 9	3	Ideally, the model should signal when input data is outside of the range of the training set. This concept could be incorporated into Principles #7 or #9.		
ISPE	General	Principle 7, 9, 10	Principles 7, 9, and 10 use the term “human-AI team” without being clearly defined. We recommend IMDRF clarify and define the term "human-AI team" as it is not clearly defined in the guiding principles documents nor is it defined in the IMDRF guidance document, "Machine Learning-enabled Medical Devices: Key Terms and Definitions" (IMDRF/AIMD WG/N67, accessible at: https://www.imdrf.org/sites/default/files/2022-05/IMDRF%20AIMD%20WG%20Final%20Document%20N67.pdf). As written in the document, it is not clear if the term “human-AI team” is used synonymously with the concept of “human-in-the-loop.” This term should be clearly defined to avoid confusion. The term “human-AI team” can be confused with other terms like “human-in-the-loop”		

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Name/Organization	Line number	Section	Comments	Proposed change	Resolution
			and it's unclear if these terms are related to one another as written in the document.		
ISPE	Principle 8	3	We suggest modifying the discussion of population subgroups to include "***any*** important subgroups" instead of assuming all populations contain relevant subgroups. This is because certain population subgroups may not be relevant or may not be adequately understood or identified based on the current understanding of a particular disease or condition for a given device's intended use/intended purpose.	Testing demonstrates device performance during clinically relevant conditions: Methodologically and statistically sound test plans are developed and executed to generate clinically relevant device performance information independently of the training dataset. Considerations include the intended patient population, any important subgroups, clinical environment and use by the human-AI team, measurement inputs, and potential confounding factors.	
ISPE	Principle 8	3	Add a sentence highlighting the assessment of infrastructure's and hardware's fitness for purpose, considering the computation power relevant for training or also explainability features.	Add "The infrastructure and hardware are fit for purpose, considering the computation power relevant for training and explainability features."	
ISPE	Principle 9		We recommend IMDRF clarify the meaning of "characteristics of the data used to train and test the model" as it is unclear whether this encompasses aspects such as data collection and analysis methods, data sources, data quality, and patient demographic information. Clarifying this will help stakeholders understand the specific elements that should be considered when disclosing the characteristics of the data used in training and testing the model.		
ISPE	Principle 10	3	The additional controls associated with "monitoring" may create new dependencies, especially for design control and risk management. Device boundaries	**Deployed models are monitored for performance and re-training risks are managed:** Deployed models have the	

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Name/Organization	Line number	Section	Comments	Proposed change	Resolution
			should be considered for these monitoring functions that may be external to the physical device. These monitoring controls need to be considered during design control and risk management activities.	capability for an appropriate level of ongoing monitoring in “real world” use with a risk-based focus on maintained or improved safety and performance. The monitoring systems and/or monitoring plans should also have clearly defined technology and architecture including where they are deployed and if the underlying monitoring system is also AI-based. Additionally, when models are retrained after deployment, there are appropriate controls in place to manage risks of overfitting, unintended bias, or degradation of the model (for example, dataset drift) that may impact the safety and performance of the model as it is used by the human-AI team.	
ISPE	Footnotes	3	Footnote 1 references a ‘Draft’ document (IMDRF/SaMD WG/N81 DRAFT:2024 Medical Device Software: Considerations for Device and Risk Characterization). Document should only reference relevant documents that are approved and list them under Reference section.	Remove reference to unapproved ‘draft’ document from page 6 of 9, unless, “IMDRF/SaMD WG/N81 DRAFT:2024 Medical Device Software: Considerations for Device and Risk Characterization” is approved prior to the approval and provisioning of “IMDRF/AIWG/N73 Good machine learning practice for medical device development: Guiding principles”	
ISPE	Footnotes	3	Footnote 9 references a ‘Draft’ document (IMDRF/MC/N79 DRAFT: 2023 Guiding Principles to Support Medical Device Health	Remove reference to unapproved ‘draft’ document from page 7 of 9, unless, “IMDRF/MC/N79 DRAFT: 2023 Guiding	

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial

Name/Organization	Line number	Section	Comments	Proposed change	Resolution
			Equity). Document should only reference relevant documents that are approved and list them under Reference section.	Principles to Support Medical Device Health Equity" is approved prior to the approval and provisioning of "IMDRF/AIWG/N73 Good machine learning practice for medical device development: Guiding principles"	

1 **MB** = Member body / **NC** = National Committee (enter the ISO 3166 two-letter country code, e.g. CN for China; comments from the ISO/CS editing unit are identified by **)

2 **Type of comment:** **ge** = general **te** = technical **ed** = editorial