4 April 2025

Dockets Management Staff (HFA-305)
Food and Drug Administration
5630 Fishers Lane, Rm. 1061 Rockville, MD 20852
via online submission to https://www.regulations.gov/

RE: Docket No. ***FDA-2024-D-4488 FDA Draft Guidance: Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations***

Dear Sir or Madam,

The International Society for Pharmaceutical Engineering (ISPE) appreciates the opportunity to comment on the above-referenced draft guidance.

ISPE appreciates the FDA's efforts to continue to expand regulatory guidance on AI.

ISPE finds the FDA's draft guidance on AI in regulatory decision-making well-structured, detailed, and comprehensive. ISPE appreciates the document's clear subchapters, technical appendices (especially on performance validation metrics), and real-world examples. However, ISPE highlights several areas that have an opportunity for improvement:

- Data Management Clarifications – A clearer definition of data types (training, tuning, validation) and their positioning within the TPLC (Total Product Life Cycle) is requested.
- Patient Data Protection – ISPE requests the FDA to mandate anonymization, security, and consent measures for patient data, including strict data deletion protocols and U.S.-only processing unless subject to stringent agreements.
- AI Model Transparency – Concerns exist regarding the feasibility of detailing complex models like Large Language Models, particularly due to supplier restrictions.
- AI-Enabled Device Approvals – Clarity is needed on first-time AI device submissions for specific populations and whether Model Cards are mandatory.
- AI-Specific Risks – Sections on safety, effectiveness, and cybersecurity risks for Large Language Models require expansion.
- Validation and Performance Monitoring – Greater alignment with Predetermined Change Control Plans (PCCP) guidance and transparency in training/validation datasets is recommended.
- Human-AI Team Concept – A clearer definition and examples of 'Human-AI team performance evaluation' are requested.

ISPE is a not-for-profit organization of individual members from pharmaceutical companies, contract manufacturing organizations, suppliers and service providers, and health authorities. ISPE's 22,000+ members lead scientific, technical, and regulatory advancement throughout the entire pharmaceutical lifecycle in more than 90 countries around the world. ISPE does not take a political position or engage in lobbying activities or legislative agendas.

We appreciate the opportunity to submit these comments for your consideration. Please do not hesitate to contact me if you have any questions.

Respectfully,

Mike Martin
ISPE President and CEO
International Society for Pharmaceutical Engineering (ISPE)
North Bethesda, Maryland 20852 USA
Mmartin@ispe.org

Response to a request for comments ***FDA Draft Guidance: Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations FDA-2024-D-4488***

Comments submitted by the International Society for Pharmaceutical Engineering (ISPE), regulatorycomments@ispe.org

GENERAL COMMENTS ON THE DOCUMENT
In the opinion of ISPE, the guideline is very well-written and clear. ISPE members have found this guideline more detailed and complete for the audience. ISPE appreciates in particular the structure in terms of: • The subchapters' division in: - '*Why should it be included in a submission for an AI-enabled device'* - '*What sponsors should include in a submission'* - '*Where sponsors should provide it in a submission'* • The Appendices with more technical points or submission addressed (in particular, Appendix C with advice on performance validation metrics and methods) • Insertion of some real-world examples, like in data management for overfitting and bias problems (lines 755-775)
ISPE suggests adding a clear definition and application of concepts around data management throughout the guidance. For instance, training data, tuning data, tuning evaluation data, test data, and clinical validation data are mentioned, where it is not always clear at what exact position within the total product lifecycle (TPLC) the data should be used.
In today's data-driven healthcare landscape, protecting patient information is essential, as such, it is ISPE's opinion that medical device manufacturers should be required to anonymize, secure and store data, obtain explicit consent, establish data deletion protocols, ensure secure data transfers, and uphold transparency and accountability. Specifically, manufacturers must: • Anonymize Data: Ensure that all patient data is de-identified to prevent the re-identification of individuals, thereby protecting patient privacy while still allowing for valuable data analysis. • Secure and Store Data: Adopt state-of-the-art security protocols—including encryption, access controls, and secure storage solutions—to safeguard patient information against unauthorized access and breaches. • Obtain Explicit Consent: Require that patients provide clear, informed consent before their data is used for secondary purposes, including training machine learning models or other advanced analytics. This step is crucial in upholding ethical standards and ensuring that patients maintain control over how their information is utilized.

We urge the FDA to integrate these provisions into its regulatory framework to ensure that as medical devices become increasingly sophisticated, patient rights and data security remain a top priority.
The management of Model Cards is not addressed in the guideline, and whether this will be mandatory or optional to provide updates to FDA (e.g., will be an addition to the 510k summary).

Specific Comments on the Text

ISPE indicates text proposed for deletion with ~~strikethrough~~ and text proposed for addition with **<u>bold and underlining</u>**.

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
171-174	The AI-enabled device does not introduce different questions of safety and effectiveness compared to the non-AI-enabled device and meets other requirements for a determination of substantial equivalence by section 513(i) of the FD&C Act.	The AI-enabled device does not introduce different questions of safety and effectiveness compared to the non-AI-enabled device and meets other requirements for a determination of substantial equivalence by section 513(i) of the FD&C Act, **<u>or the difference in term of safety and effectiveness is supported by validation evidence and related considerations on human oversight, depending on the system itself.</u>**	Depending on the AI system, human oversight needs to be considered especially when a substantial equivalence is claimed with non-AI enabled device.
189-192	While the proposed recommendations are intended to be broadly applicable to AI-enabled devices, many of these recommendations may be specifically relevant to devices that incorporate the subset of AI known as machine learning, particularly deep learning and neural networks. Additional considerations may apply for other forms of AI.	Please see the Comment column.	Since Neural Networks should belong to Deep Learning Models, we would suggest updating the sentence to read “machine learning and deep learning” instead of “machine learning and particularly deep learning and neural networks.” Please specify if ‘Additional forms of AI’ concepts could include Generative AI (GEN AI) (and Natural Language Processing (NLP)) models.

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
287	verification and validation activities, also known as the testing process, and is not used to	Please see the Comment column.	ISPE suggests distinguishing data that is used for verification of the model quality (test data set) and data that is used for acceptance testing as well as those for (clinical) validation of the device. These are distinct steps, where the testing would fall more into the data science area (demonstrating sufficient generalizability of the model), while the acceptance testing and (clinical) validation data sets would fall more into the area of Life Science or device-specific use of data.
311-315	Development – Risk Assessment, Data Management, and Model Description and Development • Validation – Data Management and Validation • Description of the Final Device – Device Description, Model Description and Development, User Interface and Labeling, Public Submission Summary	Please see the Comment column.	ISPE suggests including further aspects as elements of the TPLC, like design and design controls, determination of data's fitness for use and fitness for purpose, as well as the design of change management procedures (incl. considerations on dynamic or online-learning systems). ISPE additionally suggests either indicating the relevance of data management throughout the phases, or rendering this as an overarching topic, as it also serves important aspects in the post-market management area.
340 - 344	In addition, under 21 CFR 820.30(i) a manufacturer must establish and maintain procedures to identify, document, validate or where appropriate verify, review, and approve of design changes before their implementation ("design changes") for all devices, including those automated with software.	Please see the Comment column.	ISPE suggests that consideration is given to establishing a connection with the recently published Predetermined Change Control Plan guidance on medical devices.

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
362	Operational sequence of the device	Operational sequence of the device **use**	Lacking the term “use”.
365	How a user may interact with it,	How a user may interact with it, **what outputs are expected,**	
372 - 400	*What sponsors should include in a submission:*	Please see the Comment column.	ISPE also suggests adding as helpful for good practice: • a description of the limitations of the model and the measures that are foreseen to mitigate risks from these limitations (like warning brought to the attention of users, with a clear relation to data and modelling aspects) • Ongoing monitoring and the role of the user in terms of effective monitoring, including feedback as a key aspect to ensure ongoing control of the device. • a description of the learning behavior and change management aspects, as well as incident management processes and training of staff.
374	A statement that AI is used in the device	A statement that AI is used in the device, **the type of AI model(s) used and why is AI used**	Providing early on the type of AI model and the rationale for using AI compared to non-AI devices will allow the evaluator to assess more directly the related potential added value and risks.
389	A description of the intended use environment(s) (e.g., clinical setting, home setting)	A description of the intended use environment(s) (e.g., clinical setting, home setting)**, including other usage condition(s) (e.g., night use)**	Additional information about the context of use
397-400	A description of any calibration and/or configuration procedures that must be regularly performed by users in order to maintain performance, including when calibration must be performed and how	A description of any calibration, **setting** and/or configuration procedures that must be regularly performed by users in order to maintain performance, including when calibration must be performed and how	It is ISPE’s opinion that if some concepts related to hardware and/or physical connections of measuring devices (directly linked to Device System File (DSF) performances) could fall into the guideline

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
	users can identify if calibration is needed again or is incorrect, as applicable.	users can identify if calibration is needed again or is incorrect, as applicable. **It is recommended to highlight if, during the setting phase, the device should be set-up (e.g., for sensor devices it could means the electrodes placements) by an expert operator or autonomously by the user (with or without the need of a pre-training).**	principles, ISPE suggests an additional note to be more inclusive of the different configuration cases.
421	A description of the potential impact of the configurable elements on user decision making.	Please see the Comment column.	ISPE suggests adding guidance when the device is fine-tuned on the user's own data (e.g., in a patient setting, or when fine-tuning a model to the exact circumstances of a lab in a diagnostic setting). These aspects are like configuration, i.e., chosen by the user, though would have various consequences on the quality assurance of data and models maintaining fitness for use.
534 - 536	Where sponsors can provide it in a submission: The user interface information should be included in the "Software Description" in the Software Documentation section of the marketing submission.	Please see the Comment column.	ISPE suggests as a potential addition that warnings and indications of limitations may be added to the user interface, including measures of uncertainty when feasible or applicable, as key means to allow for informed decision-making of users.
577-578	Model Output Explanation of what the model output means and how it is intended to be used.	Model Output • Explanation of what the model output means and how it is intended to be used. • **Description of algorithms applied to manage the model output in cases of bad quality, missing or not conforming data in input, as applicable.**	ISPE suggests recommending, when applicable, a description of algorithms applied to manage the model output in cases of missing or bad quality data inputs. Commonly, in these cases, the model could be unable to present a valid and/or acceptable result alone, and some algorithms are implemented to apply data transformations (e.g., filling Not a Number (NaN) with the

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
		<u>Model Output interpretability is considered a key factor for AI-driven solutions.</u>	average of previous values or applying a moving average to smooth output). These solutions usually make the results more readable by users. Summarizing them can also be useful to assure model compliance and reliability for the intended COU.
583 - 585	Model Architecture • High-level description of the methods and architecture used to develop the model(s) implemented in the device.	Please see the Comment column.	ISPE suggests that regulated organizations are reminded to demonstrate transparency on the use of third-party models, as applicable, when describing the model architecture.
580	Automation	Automation **<u>and degree of human oversight</u>**	ISPE suggests adding not only the degree of automation, but also the counterpart, which is human oversight, to get the full vision of how the device is working and controlled.
595 - 604	Performance Data • Description of the performance validation data, including: o The source(s) of data; o Study sites; o Sample size; o Other important study design and data structure information (e.g., randomization schemes, repeated measurements, clinical reference standard); o Primary endpoints of the validation study, including pre-specified performance criteria; and o Criteria/expertise used for determining clinical reference standard data.	Add another bullet • If consideration is given to the inclusion of real-world data post-marketing, the sponsor makes a suggestion.	It is ISPE's opinion that it could be helpful to the readership to provide further considerations on the use of additional information like real-world data or insights from post-market monitoring activities. As such information becomes available, it may augment the information gathered in the clinical setting, which is currently the focus of this section. The language here could refer to future intent or be moved to the post-approval section.
628-633	Limitations • Description of all known limitations of the AI-enabled device, AI-DSF(s), or model(s).	Please see the Comment column.	ISPE suggests adding further limitations to be considered, such as the cohort of patients on which the model was developed. The model may have only been validated for certain sub

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
	Some limitations of a model may not reach the degree of severity that would warrant a contraindication, warning, or precaution, but they may still be important to include in labeling. For example, the training dataset may have only included a few patients with a rare presentation of a disease or condition; users may benefit from knowing the limitations of the data when that rare presentation is suggested by the model as a diagnosis.		populations of patients according to age or other stratification therefore it should not be applied to patients outside of the designated cohort, unless otherwise justified.
645 - 651	Customization Description of and instructions on any customizable features, including: • When users or healthcare systems can configure the operating points for the device; • When it is appropriate to select different configurations; and • When operating points are configurable, how end users can discern the operating point the device is currently operating at.	Customization Description of and instructions on any customizable features, including: • When users or healthcare systems can configure the operating points for the device; • When it is appropriate to select different configurations; and • When operating points are configurable, how end users can discern the operating point the device is currently operating at. • **Which effects could provide a wrong customization for the COU, as applicable**	Additional description for custom features. ISPE suggests adding considerations on fine-tuning models on own data, and their implication on model and version management. To our understanding, such scenarios require careful decision-making, when to iterate on a foundation model version, and when to improve the fine-tuned version via further data gathered.
72 and 672	VII Risk Assessment	Risk Assessment and Control	This terminology is to align with ICH Q9(1) since the section discusses risk control (line 728) as well as risk assessment.
704-716	Risks Related to Information in AI-Enabled Devices.	Risks Related to Information in AI-Enabled Devices.	ISPE notes a connection to the comment made on line 578 on outputs interpretability.

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
	One aspect of risk management that can be particularly important for AI-enabled devices is the management of risks that are related to understanding information that is necessary to use or interpret the device, including risks related to lack of information or unclear information. Misunderstood, misused, or unavailable information can impact the safe and effective use of a device. For example, for devices that utilize complex algorithms, including AI-enabled devices, the performance in different disease subtypes may not be apparent to users, or the logic underlying the output information may not be easily understandable, which can negatively affect user understanding and use of the device. Lack of, or unclear information can also make it difficult for different users to understand whether a device is not performing as expected, or how to correctly follow instructions. FDA recommends that consideration of risks related to understanding information should be one part of a comprehensive approach to risk management for an AI-enabled device.	One aspect of risk management that can be particularly important for AI-enabled devices is the management of risks that are related to understanding information that is necessary to use or interpret the device, including risks related to lack of information or unclear information. Misunderstood, misused, or unavailable information can impact the safe and effective use of a device. For example, for devices that utilize complex algorithms, including AI-enabled devices, the performance in different disease subtypes may not be apparent to users, or the logic underlying the output information may not be easily understandable, which can negatively affect user understanding and use of the device. Lack of, or unclear information can also make it difficult for different users to understand whether a device is not performing as expected, or how to correctly follow instructions. FDA recommends that consideration of risks related to understanding information should be one part of a comprehensive approach to risk management for an AI-enabled device. **FDA encourages the use of solutions to increase interpretability of information and results. For example, a metric to display the percentage of output accuracy can be able to quantify the reliability of the result shown to the user – especially in real-time measurements.**	ISPE notes a connection to the concepts expressed in the guideline from line 1131 to 1134 and from 1785 to 1796.

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
719 - 724	What sponsors should include in a submission: Sponsors should provide a “Risk Management File” that includes a risk management plan, including a risk assessment. In addition to other considerations, the risk assessment should consider user tasks and knowledge tasks that occur throughout the full continuum of use of the device, including, for example, the process of installing the device, maintaining performance over time, and any risks associated with user interpretation of the results of a device, as appropriate.	Please see the Comment column.	A clarification on the term "knowledge task" would be helpful to contextualize the guidance provided. In case higher-level tasks on the cognitive level like interpretation of model results - as opposed to standard steps in applying the device, activating a function, etc. - are meant, guidance on how to interpret "knowledge tasks" would be helpful.
755-759	Data management is also an important means of identifying and mitigating bias. The characterization of sources of bias is necessary to assess the potential for AI bias in the AI-enabled device. AI bias is a potential tendency to produce incorrect results in a systematic, but sometimes unforeseeable, way due to limitations in the training data or erroneous assumptions in the machine learning process	Please see the Comment column	ISPE suggests being more generic on model category involved speaking about ‘models’ learning processes’ instead of ‘machine learning processes’
765	which can impact the AI-enabled device performance in the underrepresented population.	which can impact the AI-enabled device performance in the underrepresented population. **It may happen that for a new device submission, only a specific population is included in the training model and related device performance; in this case, a disclaimer should be integrated into the User Manual.**	ISPE believes that it is important to underline that AI-model training can occur over time, that during early training that only limited population data (e.g., one ethnicity) may be available, which may then evolve.

Section or Line Number	Current Text	Proposed Change	Rationale or Comment

813-819	Data Collection
	
	• A description of how data were collected (e.g., clinical study protocols with inclusion/exclusion criteria), including:
	  ○ The names of clinical sites or institutions involved. Sites should be uniquely identified, and they should be referred to consistently throughout the submission.
	  ○ The time period during which the data were acquired.
		Data Collection
		
		• A description of how data were collected (e.g., clinical study protocols with inclusion/exclusion criteria **and reasoning**), including:
		  ○ **The place/country in which data were acquired, with specifications of a particular population features and eventual constraints about the choice.**
		  ○ The names of clinical sites or institutions involved. Sites should be uniquely identified, and they should be referred to consistently throughout the submission.
		  ○ **Acquisition protocol approval by ethical committee, when applicable**
		  ○ The period during which the data was acquired.
			According to the example made at line 765, ISPE suggests adding to the description the country in which the data were acquired to be sure to include consideration about the population's site-related features.
			
			ISPE suggests additional points about data collection approval that could benefit both sponsors and reviewers.

843-852	Data Cleaning/Processing To provide optimum training results, it may be important to clean data used for development, such as by removing incorrect, duplicate, or incomplete data. These processing steps should be described, including data quality factors used, data inclusion/exclusion criteria, treatment of missing data, and whether the steps are internal or external to the AI-DSF. Testing data, on the other hand, should only be processed in a manner that is	Data Cleaning/Processing To provide optimum training results, it may be important to clean data used for development, such as by removing incorrect, duplicate, or incomplete data **or providing data filtering and feature selection.** These processing steps should be described, including data quality factors used, **normalization/standardization/scaling, data transformations,** data inclusion/exclusion criteria **(e.g., outlier rejection),** treatment of missing data, and	ISPE suggests additional points be included to assist sponsors in their application, particularly when normalization has been used.

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
	representative of the RWD the model will encounter in its intended use. Any such data processing, data quality factors used, data inclusion/exclusion criteria, and treatment of missing data should be justified as aligned with pre-processing implemented in the final AI-DSF.	whether the steps are internal or external to the AI-DSF. Testing data, on the other hand, should only be processed in a manner that is representative of the RWD the model will encounter in its intended use. Any such data processing, data quality factors used, data inclusion/exclusion criteria, and treatment of missing data should be justified as aligned with pre-processing implemented in the final AI-DSF. **For example, when applied, data/features normalization (useful to avoid performances degradation for most ML models) is usually performed both on training and test sets.** **While the training values are often prepared for models scaled with respect to the range of training set (available and static), during test phases that involve real-time data this approach could be not applicable. Due to the nature of RWD, indeed, the test data normalization process usually considers other reference scale (commonly are the training set values themselves to act as reference range for test set normalization).**	
897-906	Management and Independence of Data • A description of the development data, including how the development data were split into training, tuning, tuning evaluation, and any additional	Management and Independence of Data • A description of the development data, including how the development data were split into training, tuning, tuning evaluation, and any additional subsets,	ISPE notes that same suggested additions made also in commenting lines 349-351 and 418-420 of FDA guideline with title: Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
	subsets, and specification of which model development activities were performed using each dataset. • A description of the controls in place to ensure the data used for testing is sequestered from the development process. • A justification of why the data used for validation provides a robust external validation. For example, a description of the sites from which test data originates from, because, in general, test data should come from sites different from those used to develop the AI-DSF.	and specification of which model development activities were performed using each dataset. **In addition:** ○ **A reporting of the percentage in which validation and test sets are split and the method for dataset splitting according to the specific COU (random choice of instances or splitting according to a rigorous sub-division – e.g., by patients in a clinical context or by batches in manufacturing).** ○ **An explanation of the cross-validation methods used to evaluate the model's performances (e.g., hold out, leave one out, K-fold cross validation, etc.), as applicable. In the case of cross-validation use, the sponsor could also focus on the modalities of the iterative evaluation on different combination of training and test set.** • A description of the controls in place to ensure the data used for testing is sequestered from the development process. • A justification of why the data used for validation provides a robust external validation. For example, a description of the sites from which test data originates from, because, in general, test data should come from sites different from those used to develop the AI-DSF.	Making for Drug and Biological Products - Guidance for Industry and Other Interested Parties - DRAFT GUIDANCE Additions include: - Percentage of splitting for training/validation and test sets - mode of splitting data – randomly or according to a specific rule (e.g., by patients) - cross-validation technique

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
937-939	FDA understands that, depending on the source of the patients and/or samples used in the training and test data, some relevant patient characteristic information may not be available	FDA understands that, depending on the source of the patients and/or samples used in the training and test data, some relevant patient characteristic information may not be available **(including differences in the ethnicities of the data used for the training model)**	It is important to note that an AI-model is training over time. This means that in the beginning, there may be the situation that only one ethnicity is available and then during time and progress with future studies the ethnicities range may be updated.
1012	Model Development	Please see the Comment column.	ISPE suggests adding a requirement that the libraries used to construct and build the model should be included in the model development description. These are critical to the construction of the model and referring to them will facilitate the description of the model development from a technical perspective.
1022 - 1024	• An explanation of any pre-trained models that were used, as applicable. o If a pre-trained model was used, specify the dataset that was used for pre-training and how the pre-trained model was obtained.	Please see the Comment column.	ISPE suggests adding further description on the requirement that the model provenance should be included in the description. Model architecture can be constructed from code or can be downloaded from repositories. These models can also be pre-trained to enhance their abilities – sometimes known as transfer learning.
1036 - 1046	For an AI-enabled device, validation includes ensuring that the device, as utilized by users, will perform its intended use safely and effectively, as well as establishing that the relevant performance specifications of the device can be consistently met. For AI-enabled devices, manufacturers should demonstrate users' ability to interact with and understand the device as	Please see the Comment column.	ISPE recommends adding a reminder that such activities should be based on having received appropriate training or user information.

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
	intended in addition to ensuring the device itself meets relevant performance specifications. To this end, it can be helpful to consider both performance validation (including human factors validation) and an evaluation of usability. Note that, for the purposes of this guidance (in the context of risk controls in the absence of human factors validation), usability describes whether he device can be used safely and effectively by the intended users, including whether users consistently and correctly receive, understand, interpret, and apply information related to the AI-enabled device.		
1089	Performance Validation	Please see the Comment column.	ISPE requests further guidance on the acceptance testing of interfaces, in particular for groups like the elderly, which would be helpful, in so far as they need to conclude from model results and explanatory factors being presented. An example may be a mental health or obesity control application that supports the patient with recommendations on activities supporting clinical benefit. Here, the patient would be the direct decision-holder acting on the device's recommendations. ISPE also suggests adding aspects around feedback provided by users. Feedback may involve corrections to model results or qualitative feedback on the model's behavior, which can play an important role in post-market monitoring and further model development. If the feedback is a critical control factor, care should be given to the quality assurance of such data and training or raising awareness

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
			with users on the relevance of their activities and input, including risks (e.g., thinking about insulin treatment and the expected quality of feedback when a patient exhibits low sugar levels). Also, such feedback may need to be assessed in terms of periodic reviews.
1099	How the device performs overall in the intended use population	How the device performs overall in the intended use population **and based on what version of the model**	ISPE believes it will be important to evaluate the device performance by knowing the data that has been used for training and the associated potential model version.
1156-1173	For some devices, more emphasis may be placed on the model's standalone performance (i.e., Did the actual output match the expected output?). For others, a focus may be assessing the performance of the human-AI team, beyond just the performance of the model in isolation (i.e., Did the intended user working with the new device perform the same or better than the operator alone or with another device?). Sponsors should consider that, in certain scenarios, both standalone and human-device team performance evaluations may support the overall performance evaluation of the AI-enabled device. Performance evaluation of AI-based medical image analysis systems is an illustrative example of how the clinical study approaches may change as the intended use of the device moves along the spectrum of human-device interactions. Standalone assessments measure the model's performance	Please see Comment column	ISPE suggests providing a definition of Human-AI team to make clearer the concept ISPE also suggests defining "**reader studies"** before mentioning them. For example, we have found the definition made in the abstract of this work to be useful: https://pubmed.ncbi.nlm.nih.gov/29653758/

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
	independently of human interaction, whereas reader studies compare the performance of the intended user both with and without the AI-enabled device (i.e., comparing the human vs. human-device team performance). Reader studies typically serve as the primary performance evaluation for AI-enabled devices that aid in clinical decision-making in medical imaging applications, because they allow sponsors to evaluate the tool's clinical benefit in the hands of the intended user.		
1181-1196	Validation methods differ depending on the intended use of a device. For example: • Devices estimating defined measurements otherwise performed by accepted reference methods may need a precision study to adequately assess their repeatability and reproducibility. • Devices monitoring time-series patient data and needing periodic re-calibrations may need a stability study and a change tracking study to assess their dynamic responses. • Devices similar to survey instruments measuring less well-defined patient parameters may need additional evidence of construct validity (i.e., the extent to which a test measures what it is proposed to measure). • Prognostic clinical decision support devices may need longitudinal data	Validation methods differ depending on the intended use of a device. For example: • Devices estimating defined measurements otherwise performed by accepted reference methods may need a precision study to adequately assess their repeatability and reproducibility. • Devices monitoring time-series patient data and needing periodic re-calibrations may need a stability study and a change tracking study to assess their dynamic responses. • Devices similar to survey instruments measuring less well-defined patient parameters may need additional evidence of construct validity (i.e., the extent to which a test measures what it is proposed to measure). • Prognostic clinical decision support devices may need longitudinal data with survival analysis, calibration analysis, and/or discrimination analysis (e.g., risk	ISPE suggests adding a note for criticality based on the COU (e.g., clinical critical parameters monitored).

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
	with survival analysis, calibration analysis, and/or discrimination analysis (e.g., risk stratification analysis), among other methods. Depending on the specific AI-enabled device, this evidence could come from non-clinical bench or analytical studies, pre-clinical animal studies, clinical performance studies, clinical outcome studies, or some combination thereof.	stratification analysis), among other methods. Depending on the specific AI-enabled device, this evidence could come from non-clinical bench or analytical studies, pre-clinical animal studies, clinical performance studies, clinical outcome studies, or some combination thereof. **In the case of devices that estimate parameters critical for patients' survival, the submission may require additional details about validation to ensure safety and reliability.**	
1263-1276	Study Results To support performance validation, sponsors should include information regarding the study results. Important aspects for these documents to cover include: • An explanation of the pre-specified results for each test, including subgroup analyses. • An explanation of the results with adequate subgroup analyses for relevant subgroups as described above. ○ If demographic information is not available for the study data, an explanation of the reasons it is not available,	Study Results To support performance validation, sponsors should include information regarding the study results. Important aspects for these documents to cover include: • An explanation of the pre-specified results for each test, including subgroup analyses. • **A detailed description on how performances and results are validated (e.g., comparison with gold standard devices, with expert annotations or others)** • An explanation of the results with adequate subgroup analyses for relevant subgroups as described above.	ISPE suggests the addition of a description of the performance validation methods and the reference chosen.

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
	why performance evaluation can be supported without demographic subgroup analysis, and how risks associated with the lack of demographic subgroup analyses have been controlled. • When feasible, and appropriate, an evaluation of the device repeatability and reproducibility. The specifics of how these studies are conducted will depend on the specific device being evaluated, and may include phantom, simulated, contrived or clinical data	○ If demographic information is not available for the study data, an explanation of the reasons it is not available, why performance evaluation can be supported without demographic subgroup analysis, and how risks associated with the lack of demographic subgroup analyses have been controlled. • When feasible, and appropriate, an evaluation of the device repeatability and reproducibility. The specifics of how these studies are conducted will depend on the specific device being evaluated, and may include phantom, simulated, contrived or clinical data	
1287	XI. Device Performance Monitoring	Please see the Comment column.	ISPE recommends the addition of a comment on this section on the implications and considerations of performance monitoring for online-learning systems and the need to mitigate the risk of degraded model performance through the ingestion of additional training data.
1290	…may change or degrade over time, presenting a risk to patients.	…may degrade over time **as the nature of the input data drifts and deviates from the data used in training and testing the model**, presenting a risk to patients.	ISPE recommends clarifying this statement to avoid misleading the reader into thinking that the model itself deteriorates internally, instead, it is the model's fitness or alignment with the real-world relationship between the input it receives and output it is expected to predict that may decline.
1357-1363	• Monitoring potential causes of undesirable changes in performance, such as:	• Monitoring potential causes of undesirable changes in performance, such as:	ISPE recommends the following:

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
	○ Changes in patient demographics or disease prevalence; ○ Shifts in input data; ○ Changes to input data due to corruption in the data pipeline (input data integrity), such as missing values, duplicate records, data type mismatches; and ○ Changes in users' behavior or in user demographics.	○ Changes in patient demographics or disease prevalence; ○ Shifts in input data; ○ Changes to input data due to corruption in the data pipeline (input data integrity), such as missing values, duplicate records, and data type mismatches; and ○ Changes in users' behavior or in user demographics. ○ **Network connection drop (e.g., when cloud platform data are used as inputs)** ○ **Changes in input data quality due to an incorrect use of the device** ○ **Break/malfunctioning of physical connections of sensors" and "incorrect setting of device (e.g., incorrect placement of sensors)**	Cases in which no data comes in inputs could be counted among the sources of changes in device performance. Bad quality data could be provided due to usages different from those indicated. For example, in cases of extreme activities for devices designed to be used at rest. It may be a subpoint of 'shifts in input data'. If it is Device Hardware functioning that influences DSFs (device software functions) that could fall into the guideline principles, we would suggest adding also data shifts.
1525	A statement that AI is used in the device	A statement that AI is used in the device, **its added value and human oversight**	ISPE believes this information may of interest and trust for public summaries.
1529 - 1531	• A description of the class of model (e.g., convolutional neural network, recurrent neural network, support vector machine, transformers) and limitations of the model within the device description;	Please see the Comment column.	The bullet point considers a single model; however, as mentioned earlier in the guidance, models may be combined to a structured approach (e.g., a tree model on patient characteristics, a convolutional neural network on image data, and a regression models that combines indications from the other two models to a single score). Therefore, ISPE recommends expanding this point to make

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
			provision for more complicated modelling architectures and settings.
1644	• Logic of operation of each device component and of the user interface system as a whole.	Please see the Comment column.	ISPE recommends adding transparency to any limitation of the model within the device relevant and understandable by the user.
1713 - 1721	Study Reports All performance and usability assessments should be objective, and the model should not be tweaked opportunistically in light of the test data results (i.e., no post-hoc adjustment). In general, proceeding to execute the study protocol only after a sound validation plan (study protocol and statistical analysis plan) is documented and finalized helps avoid these post-hoc adjustments. Execution of the plan includes collecting the required data, conducting the pre-specified analysis, and reporting the study results. Validation study reports should specify the associated protocol version and adequate justifications should be provided for any repeated tests or tests with deviations from the pre-specified plans.	Please see the Comment column.	ISPE members follow the thinking that opportunistic post-hoc modification of models may be unacceptable. However, ISPE would like to recommend adding further guidance on more constructive use of insights from applying the model to test data sets or during clinical validation. For instance, if crucial insights are gained on the modelling device and its safety (e.g., on the limitations of its generalizability), changes may be applicable in spirit of promoting patient safety. Such activities, when comprehensively safeguarded by thorough testing to mitigate potential risks in these changes, should be beneficial to the device.
1882	As such, AI-enabled devices can be prone to errors of device use and information interpretation.	As such, **it may be challenging to explain how AI-enabled devices arrive at a decision which can lead to errors** of device use and information interpretation.	ISPE suggests clarifying that the root cause of the error is user understanding and interpretation.
1930-1933	In general, model cards can be adapted to the specific needs and context of each AI-enabled device. However, for the public summary, we encourage sponsors	In general, model cards can be adapted to the specific needs and context of each AI-enabled device. However, for the public	ISPE recommends including addition of explanation.

Section or Line Number	Current Text	Proposed Change	Rationale or Comment
	to follow the general principles for creating model cards outlined in this guidance. Some elements may not be available for some devices.	summary, we encourage sponsors to follow the general principles for creating model cards outlined in this guidance. Some elements may not be available for some devices **(only applicable parts should be considered to provide a general overview of AI-enabled devices)**	

End of Comments