ISPE Comments on FDA-2026-N-7874, Considerations for the Regulation of Generative AI-Enabled Medical Devices: Discussion Paper and Request for Feedback

Submit Date: October 14, 2026

Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Room 1061, Rockville, MD 20852

Re: Docket No. FDA-2026-N-7874, *Considerations for the Regulation of Generative AI-Enabled Medical Devices: Discussion Paper and Request for Feedback*

To Whom It May Concern:

The International Society for Pharmaceutical Engineering (ISPE) appreciates the opportunity to comment on the discussion paper titled *"Considerations for the Regulation of Generative AI-Enabled Medical Devices."*

ISPE supports FDA's development of a risk-based framework for GenAI-enabled devices. Our comments emphasize that evidence expectations should be proportionate to risk and context of use, allowing sponsors to justify fit-for-purpose approaches rather than applying uniform requirements.

Because GenAI is evolving rapidly, we also encourage a technology-agnostic, total-product-lifecycle approach that connects premarket evaluation, postmarket monitoring, and change control. Under such an approach, CDRH would specify expected outcomes rather than particular tools or benchmarks, and manufacturers would remain accountable for device performance.

ISPE appreciates FDA's consideration of these discussion points.

Sincerely,

Michael S. Martin, President and CEO
International Society for Pharmaceutical Engineering (ISPE)
6110 Executive Blvd., Suite 600, North Bethesda, MD 20852
Tel. +1 301-364-9201 | www.ispe.org

Enclosure: ISPE Discussion Paper Response

IV. Risk Assessment

Question 1

> ***FDA question:*** *Does the two-axis risk framework, organized around device activity and the consequence of relying on an incorrect output, appropriately capture the dimensions most relevant to the risk of a GenAI-enabled software function? If there are additional dimensions—such as the reversibility of a resulting action, the availability of downstream safeguards, the time pressure of the deployment setting, or the traceability of the output (i.e., to primary source materials)—that should be represented in a risk framework, please describe and provide examples of how they should be represented.*

Position summary

ISPE supports the two-axis framework, similarly to what was presented in FDA's draft guidance Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products

Question 2

> ***FDA question:*** *CDRH seeks input on how to account for the spectrum of GenAI-enabled informational functions that vary in the degree to which they direct a user to a particular action (i.e., between "non-directive" and "action-directing"). What characteristics of an output—such as its wording, specificity, personalization, or context—could be considered as modifiers of the risk associated with an informational function, after accounting for the device's overall functionality and intended use? What additional information would provide manufacturers with sufficient clarity and predictability around risk assessment for informational functions while recognizing that directiveness may exist along a continuum rather than as a binary distinction?*

Position summary

ISPE supports the position that there may be further considerations beyond signaling words to delineate between informational and directional objectives.

Rationale/detail

Where wording, specificity, or the nature of the output materially affects directiveness or risk, these characteristics should be identified during the initial risk assessment and addressed during risk assessment and risk control steps in risk management exercise. Risk control recommendations should be taken by a team of experts familiar with the situation and potentially tested in an evaluation exercise.

Recommendation to CDRH

While the delineation between informational and directional functions is helpful, CDRH may consider emphasizing the de facto impact on the output from the decision-making process in the risk management exercise, which may be influenced by signal words like "recommend."

Question 3

> ***FDA question:*** *CDRH seeks input on whether and when a GenAI-enabled function that results in delivery of clinical information to patients, as opposed to HCPs, could present different or higher risks, while also recognizing the potential benefits associated with improved patient empowerment, engagement, and access to clinical information. What device characteristics, output features, or safeguards might mitigate risks that could arise when a user lacks the domain knowledge to independently evaluate an output, without unnecessarily underestimating patient capability?*

Position summary

ISPE believes that the level of literacy and the ability of critical thinking of the desired output should be reflected in the device design, in concordance with the target population (users/patients) and development, using risk management, i.e., the risk identification step.

Rationale/detail

It is well known, for example, in ICH Q9(R1) (Quality Risk Management), that subjectivity, e.g., bias, can impact all stages of a risk management exercise.

Recommendation to CDRH

CDRH may consider mentioning that the risk management process used during development should be designed to minimize bias during the exercise and in the resulting output.

Question 4

> ***FDA question:*** *CDRH seeks input on whether and how the distinction between generalist and specialist physicians might inform the assessment of risk for HCP-facing GenAI-enabled functions. Under what circumstances, if any, might risk be affected when an HCP who lacks the relevant clinical specialist knowledge receives information that falls within a specialist area of practice? What device characteristics or safeguards might mitigate such risks?*

Position summary

ISPE believes that a distinction between estimated competencies of generalist and specialist physicians should be factored into the risk management exercise.

Rationale/detail

The entire risk management exercise should consider the clinical action that may result from the output and whether the intended HCP has the competence, authorization, and relevant scope of practice to take that action.

Recommendation to CDRH

The CDRH may consider including discussion of factoring HCP competence into device development recommendations.

Question 5

> ***FDA question:*** *For multi-turn conversational GenAI-enabled devices that may migrate from providing "non-directive" information to "action-directing" information over the course of an exchange, how*

> *should risk be assessed across realistic conversational trajectories? How could the intended use of such a device be characterized when its behavior is emergent across a conversation?*

Position summary

ISPE acknowledges the possibility that conversations may drift across the informational to directional spectrum. The intended operational boundaries and context of use should be clearly defined, and the system should explicitly communicate how its output is intended to be interpreted. The associated risk management process should support these boundaries through risk-proportionate test coverage and appropriate control measures.

Rationale/detail

An adaptive approach would accommodate both a) the nature of conversational trajectories and b) sufficient rigor when more directional portions of the spectrum are proposed for use.

Recommendation to CDRH

CDRH may consider risk assessment outcomes that describe a range of modes within the informational vs. directional spectrum, while manufacturers would leverage these risk outcomes through targeted control strategies commensurate with the respective risk to the patient, which may change throughout a conversational trajectory.

Question 6

> ***FDA question:*** *For GenAI-enabled care escalation functions, CDRH is considering how the evaluation may account for both under-escalation and over-escalation. How could manufacturers characterize and weigh these two directions of error, given that they may not be commensurable and that acceptable trade-offs may vary by clinical context?*

Position summary

ISPE believes that under- and over-escalation should be evaluated, controlled, and monitored in the risk management exercise because their clinical consequences and operational impacts may differ. The appropriate trade-off and control strategy should reflect the criticality of the condition, the performance and error modes of the existing care pathway, and the demonstrated effectiveness of human oversight. Users should be able to escalate or override outputs based on additional evidence and clinical judgment, subject to controls proportionate to the resulting risk.

Rationale/detail

Under- and over-escalation should be characterized, measured, and controlled separately because their likelihood, clinical consequences, and operational impacts may differ. The control strategy should be informed by the performance and error modes of the existing care pathway and should reflect the relative severity of each potential error in the intended context. Human oversight may mitigate residual risk, but should receive credit only where its effectiveness has been demonstrated. Users should be able to escalate or override outputs based on additional evidence and clinical judgment, subject to controls proportionate to the resulting risk.

Recommendation to CDRH

CDRH may consider providing guidance that over- and under-escalation should be treated separately in risk management exercises.

V. Premarket Evaluation

Question 7

> ***FDA question:*** *Is the competency-based approach described above, i.e., device benchmarking followed by clinical confirmation, a useful and appropriate framework for evaluating GenAI-enabled devices?*

Position summary

ISPE believes that the competency-based framework is an appropriate approach depending on the purpose of the GenAI system whilst recognizing the challenges of defining competencies.

Question 8

> ***FDA question:*** *How might the two-axis risk framework described in Section IV be considered within the competency-based approach described above to help determine the level of evidence needed for a premarket submission?*

Position summary

ISPE supports using risk management to determine the nature and rigor of evidence required, which should include a proposed, justified control strategy in a pre-market submission. Higher-risk, action-directing or autonomous functions should receive greater scrutiny of robustness, safety-critical recognition and escalation, and adherence to intended-use boundaries.

Rationale/detail

ISPE considers the two-axis framework a useful starting point for the device development risk management exercise.

Recommendation to CDRH

CDRH should present the two-axis framework as a useful starting point for the risk management exercise. The guidance could consider including all important factors received as part of this consultation exercise.

Question 9

> ***FDA question:*** *Would a benchmarking structure such as the one described above be likely to provide adequate evidence of clinical knowledge, analytic capabilities, safety behavior, communication, and generalizability to support a reasonable assurance of safety and effectiveness? Are there elements that are missing, redundant, or inappropriately categorized? Are there externally developed standards that could be leveraged?*

Position summary

ISPE generally supports describing a benchmarking structure; however, care should be taken to allow for changing technology.

Rationale/detail

Since technology is rapidly evolving, the competencies applicable to Agentic AI to ensure acceptable performance may change.

Recommendation to CDRH

CDRH may consider a technology-agnostic approach in benchmarking to cover a broad range of changing technologies.

Question 10

> ***FDA question:*** *CDRH recognizes that publicly available benchmarking assets may be subject to data contamination, saturation, and limited real-world representativeness. How should a sponsor establish that performance on a given benchmark predicts safe and effective real-world behavior for the device's intended use? What evidence should support the construct validity of a benchmark used to gate device evaluation, and what role should sponsor-developed benchmarks play given potential concerns around independence and optimization to the test?*

Position summary

ISPE shares FDA's view on the limitations of publicly available data. For a benchmark on given data - externally sourced or internally - to be reliable, the underlying data and its limitations, strengths, and weaknesses need to be understood to a sufficient extent. In the event of limitations, further sponsor-developed benchmarks may be needed, leveraging targeted data in the context of use of the AI-enabled device where evidence is currently uncertain.

Rationale/detail

ISPE has developed a fitness for purpose concept, comprising four properties

- Reliability,
- relevance
- representativeness
- abundance

These could serve as an orientation for assessing and, where appropriate, producing additional data for benchmarking.

Recommendation to CDRH

CDRH may consider requiring that all benchmarks, whether publicly available or sponsor-developed, are fully justified in relation to the overall risk characterization, perhaps specifically mentioning the need to consider independence and optimization.

Question 11

> ***FDA question:*** *CDRH is considering that clinical confirmation for a GenAI-enabled device might not require a prospective clinical study in every case, and has described above a range of approaches of increasing rigor and patient exposure. How might a sponsor select and justify a confirmation approach tailored to a device's intended use and proportionate to the device's risk profile? Are there device types or risk profiles for which one or more of these approaches would be insufficient or inappropriate? How might the anticipated distribution of real-world inputs be taken into consideration? Are there other methods of clinical confirmation that might help inform the evaluation of GenAI-enabled devices?*

Position summary

ISPE believes that prospective clinical studies should not be universally required for a confirmation approach. The confirmation approach, clinical or non-clinical, should be proportionate to the device's

risk profile, with prospective evidence required to provide sufficient assurance of safety and effectiveness.

Rationale/detail

Selection should consider, for example, system novelty and autonomy; consequences, detectability and reversibility of error; quality and representativeness of existing evidence; anticipated real-world inputs; and the demonstrated effectiveness of proposed risk controls and human oversight.

Recommendation to CDRH

CDRH should permit sponsors to select and justify a fit-for-purpose confirmation approach with appropriate justification.

Question 12

> ***FDA question:*** *How should sponsors achieve statistically meaningful performance measurement for GenAI-enabled devices? Where synthetically generated inputs supplement real patient data, how should sponsors account for differences between the synthetic and real-world distributions when estimating performance, and under what conditions, if any, is it appropriate to combine benchmarking evidence and clinical confirmation evidence to support a single performance estimate?*

Position summary

ISPE believes that statistically meaningful performance will depend on several factors, and the suitability of the data used should be justified accordingly. For instance, the amount of data required to derive statistically meaningful conclusions may depend on the complexity of the model, the scope of the task, and the heterogeneity of data subgroups, each of which may require evidence. When real-world data is scarce or may not fully capture the range of expected real-world scenarios, synthetic data may be used with justification. However, transparency regarding the nature and proportions of the data, as well as the performance of the synthetic data-generating approach, is necessary. Full reliance on synthetic data should require extended reasoning on suitability of the chosen approach in relation to risk.

Rationale/detail

Regarding use of data, and data expectations, trade-offs and compromises may be required. Taking account of risk, various factors such as facilitating data sharing may be supportive measures, as well as use of synthetic data where its adequacy for the device's context can be demonstrated.

Recommendation to CDRH

CDRH may consider a variety of approaches to establish a robust data foundation for deriving a performance estimate, while considering a range of factors that should be considered to satisfy data quality needs; for instance, decision consequence, prevalence of the condition, complexity of the AI approach, heterogeneity of the device's anticipated performance across subgroups and sub-populations, etc.

Question 13

> ***FDA question:*** *For which clinical domains, device functions, or subpopulations is synthetic data particularly well-suited, or particularly inadequate, as a supplement to real-world evidence? What safeguards would mitigate the risk that synthetic data generated by models of the same class as the*

> *device under evaluation reproduces the very performance gaps the evaluation is intended to detect, particularly for underrepresented subgroups?*

Position summary

ISPE believes that the use or restriction of synthetic data should be determined by its fitness for the intended purpose and the level of risk, including the extent and significance of gaps in available real-world data. Its purpose, generation process, and limitations should be specified and assessed, and its relevance, reliability, and representativeness should be demonstrated against independent real-world data, particularly for relevant populations, conditions, and failure modes.

Rationale/detail

ISPE believes that the use or restriction of synthetic data should not be determined solely by clinical domain or device classification. Its suitability should be assessed for the intended purpose and proportionate to the risk, considering the limitations of available real-world data, the role and relative contribution of synthetic data within the evidence package, and the extent to which its relevance, representativeness and reliability can be demonstrated against independent real-world data. Synthetic data should supplement, rather than obscure, evidence gaps and should not be used where its validity for the relevant populations, conditions or failure modes cannot be adequately established.

Recommendation to CDRH

CDRH should avoid prescribing suitability for synthetic data by clinical domain or device classification alone. Sponsors should justify its use based on intended purpose, risk and gaps in real-world data, and demonstrate its relevance, reliability and representativeness against independent real-world evidence, including for underrepresented subgroups and relevant failure modes. CDRH may consider the need for transparency on the purpose of synthetic data generation, its generation process, and a structured assessment of potential deficiencies or weaknesses, to derive the incremental value of such data compared to other evidence sources.

Question 14

> ***FDA question:*** *For open-ended device outputs, how should performance comparators and acceptance criteria be selected? When a panel of qualified clinicians serves as the comparator, how should the applicable standard (for example, the standard of care versus the performance of a median clinician in practice) be defined and justified? Should generalist or specialist physicians be used as a performance standard? When should human-AI team performance, rather than the device operating alone, serve as the basis for evaluation?*

Position summary

ISPE believes that performance comparators and acceptance criteria for open-ended outputs should be selected and justified according to the intended use, clinical context and level of risk. Evaluation methods should be fit for purpose and shown to produce reliable, clinically meaningful and reproducible assessments, while the appropriate comparator (including current standard of care, qualified clinicians, alternative technologies or human-AI team performance) should reflect the decision being supported and the reasonably available alternatives.

Rationale/detail

Where clinicians provide the comparator, their relevant specialty, competence, experience and clinical responsibility should reflect the intended users and decisions supported by the device.

Recommendation to CDRH

CDRH may consider the need to recommend rationales to justify the chosen performance comparators, acceptance criteria, and benchmarks, and to consider alternative options, as well as elevated requirements in cases of higher complexity or risk (e.g., when aiming for a shift from HCP-supervised to autonomous mode). All these considerations should be related to the context of use.

Question 15

> ***FDA question:*** *Are there ways in which performance might be assessed relative to the care, technology, or course of action likely to occur in the absence of the device, rather than to the comparators described in this section? What approaches might be used to identify and justify a comparator such as unaided clinical judgment, delayed specialist review, or no intervention?*

Position summary

ISPE considers that this question should be answered by those with significant clinical expertise and experience in the use of devices, especially devices which include some form of model.

Question 16

> ***FDA question:*** *Should independent third parties be involved in some or all aspects of a competency-based approach, including device benchmarking and clinical confirmation? If so, in what ways might qualified, independent third-party participation contribute to this approach, and what qualifications and independence criteria should apply? Are there aspects of the assessment for which third-party involvement would be impractical or inadvisable? What safeguards or program design features would be critical to prevent third-party participation from limiting competition or preventing innovation?*

Position summary

Third-party evaluation should be available as a risk-based assurance mechanism but should not be universally mandated. Its value depends on the system's risk, novelty, and complexity, as well as the independence or specialist expertise required for a credible assessment.

Rationale/detail

A secondary, independent AI system could support evaluation, for example, by reviewing outputs at scale, detecting inconsistencies or identifying cases requiring expert review. However, it should not automatically be treated as an independent third party. Its independence would depend on separate model provenance, training data, architecture, provider, and incentives.

Recommendation to CDRH

CDRH should avoid requiring third-party evaluation in every case. CDRH should permit risk-based justification of evaluation approaches considering the context of use. CDRH should preserve manufacturer accountability regardless of who performs the assessment.

Question 17

> ***FDA question:*** *CDRH is exploring whether a competency-based approach could be applied to devices that incorporate a variety of underlying model architectures, such as multimodal vision-language models and generative or predictive world models. Are there aspects of the described approach that*

> *might be ineffective or inapplicable to these devices? What additional or distinct considerations might need to be addressed?*

Position summary

ISPE supports an approach, as largely presented in the document, that is technology-agnostic and should avoid specifying benchmarks relying on the clinical context and the technology choice.

Rationale/detail

For example, scope maintenance and boundary adherence may be defined for various modalities of input, demonstrated across a suitable set of data and/or scenarios, while applying to the general logic of this benchmark category.

Recommendation to CDRH

CDRH may consider a competency-based approach that is largely technology-agnostic, depends on technology choice and context of use, and is underpinned by an adequate risk-based rationale.

VI. Postmarket Monitoring

Question 18

> ***FDA question:*** *CDRH is considering whether it may be appropriate to accept greater premarket uncertainty regarding a GenAI-enabled device's benefit-risk profile through greater reliance on postmarket monitoring. Under what conditions might such an approach be appropriate, and what characteristics of a monitoring program would need to be in place to justify reduced premarket evidence? Are there device types or risk profiles for which this approach would not be appropriate?*

Position summary

ISPE supports an approach that accepts greater premarket uncertainty regarding a GenAI-enabled device's benefit-risk profile by placing greater reliance on postmarket monitoring. The design of the postmarketing monitoring program should be based on the estimated residual risk from the premarket development program, particularly when residual risk reduction would be most effective through a carefully designed postmarketing evaluation program.

Rationale/detail

Allowing more premarket uncertainty may lead to faster development with the benefit of devices being available to patients earlier. Use of a postmarket monitoring approach could in some cases allow more real patient data to become available and be used to increase the assurance of an evaluation of the performance of the GenAI-enabled device.

Recommendation to CDRH

While CDRH may generally link expectations on acceptable premarket uncertainty and appropriate depth of monitoring activities to the level of risk and the clinical context, certain circumstances, such as a lack of alternative options and scarcity of data, may admit for higher premarket uncertainty compared to settings in which more treatment options or data are available. Where greater premarket uncertainty is accepted, sponsors should clearly characterize and document the remaining uncertainties and associated risks, communicate them appropriately to users, and explain the postmarket measures and actions intended to address them.

Question 19

> ***FDA question:*** *Please comment on the potential approaches to postmarket performance evaluation, including periodic re-benchmarking, sample-based clinician review, and performance degradation monitoring. What additional approaches should CDRH consider, and how should the cadence and triggering events for reassessment be determined?*

Position summary

Postmarket performance evaluation should build upon the premarket evaluation framework, using consistent and comparable measures where appropriate as discussed in the answer to Question 18. A coherent lifecycle approach supports detection of performance changes over time while improving the efficiency of evaluation design and implementation.

Rationale/detail

Further coverage of relevant elements (data, adaptive behavior) may allow for greater control and earlier detection of deficiencies, rather than the lagging measure of device performance. Continuity with premarket evaluation improves comparability over time and enables an efficient, coherent lifecycle approach.

Recommendation to CDRH

CDRH may consider additional dimensions that may be relevant for postmarket performance, including data and adaptive behavior. CDRH should encourage postmarket evaluation frameworks that build on premarket methods and measures.

Question 20

> ***FDA question:*** *Please comment on whether the proposed approaches to postmarket monitoring can be facilitated by machine-based supervisory agents. What considerations, including the evaluation and reliability of the supervisory agent itself, should CDRH take into account for such an approach?*

Position summary

ISPE welcomes the possibility to employ machine-based supervisory agents for postmarket monitoring purposes. The performance of such agents should be verified themselves, as they form part of the control strategy. As a matter of scalability, a portion of instances may be retained for (additional) human review to allow for a monitoring loop for the supervisory agent itself.

Rationale/detail

With the supervisory agent becoming part of the control system and simultaneously exhibiting model characteristics, such functionality would be subject to a control strategy in itself.

Recommendation to CDRH

CDRH may consider the use of machine-based supervisory agents with appropriate demonstration of their suitability and adequate performance monitoring in place.

Question 21

> ***FDA question:*** *What roles might clinicians, healthcare institutions, professional societies, standards-setting bodies, and other stakeholders appropriately play in postmarket monitoring, and how can these roles be structured without diffusing manufacturer accountability?*

Position summary

ISPE believes that postmarket monitoring can be supported through user education, incident reporting, clinical review, performance-data sharing, and the identification of emerging risks. These complementary roles should be clearly defined and should not transfer or dilute the manufacturer's accountability for device performance and regulatory compliance.

Rationale/detail

Training, clear documentation, and accessible user support can help users evaluate outputs, recognize limitations, and report potential deficiencies. Stakeholder feedback should be collected through defined channels, assessed systematically, and incorporated into the manufacturer's risk management, corrective action, preventive action, and change management systems.

Recommendation to CDRH

CDRH should encourage defined stakeholder roles, reporting pathways, and information-sharing arrangements while maintaining clear manufacturer ownership of monitoring, investigation, and corrective action.

Question 22

> ***FDA question:*** *CDRH envisions that a manufacturer's premarket competency-based assessment might serve as a baseline against which post-deployment modifications could be re-evaluated. How might the extent of re-benchmarking or other evidence be scaled to the nature and expected impact of a given modification? For example, are there categories of change that might not significantly affect safety or effectiveness of a GenAI-enabled device, or might be appropriately managed within a sponsor's quality management system versus requiring FDA premarket review and authorization before implementation? Are there categories of changes that might be appropriate for inclusion in a PCCP?*

Position summary

ISPE generally supports the possibility of including modifications to GenAI-enabled medical devices in PCCP. However, even seemingly routine changes, such as prompt alterations, may cause unintended performance shifts, and establishing sufficient evidence for such changes may be challenging. Such changes should therefore be supported by rigorous verification against predefined benchmark expectations and acceptance thresholds, including, where relevant, measures of performance shifts between model versions.

Recommendation to CDRH

CDRH may consider inclusion of modifications to AI-enabled medical devices in a PCCP, in due consideration of effective assurance of performance of such changes, whether carried out manually or automatically.

Question 23

> ***FDA question:*** *How might PCCP concepts or other change-control approaches be adapted for GenAI-enabled devices when the nature or scope of future modifications cannot be fully prespecified?*

Position summary

ISPE believes that sponsors should maintain effective configuration and change control across all device components, consistent with established practices for externally sourced software.

Rationale/detail

Controls should provide traceability of component versions, dependencies and changes, together with impact assessment, predefined acceptance criteria, regression testing and rollback capability. Sponsors should use effective and current configuration-management methods appropriate to the device, its architecture and level of risk.

Recommendation to CDRH

CDRH should specify expected control outcomes without prescribing configuration-management tools or technologies.

Question 24

> ***FDA question:*** *For devices built on third-party foundation models, changes to the underlying model may be initiated by the third-party model developer rather than the device manufacturer. How can a manufacturer detect, evaluate, and respond to such changes in a timely manner? What mechanisms—for example, contractual, technical, or through a PCCP—could provide reasonable assurance that third-party developer-initiated changes do not compromise the safety or effectiveness of the device?*

Position summary

Generally, notifications of changes should be expected, similar to changes performed by cloud providers. In addition, technical measures such as periodic re-benchmarking against a fixed reference set may unveil potential shifts in performance in the context of use, while considering potential probabilistic behavior of models. As described above, a PCCP with performance controls may provide the basis to efficiently accommodate such changes.

Rationale/detail

The sponsor's model selection should consider the third-party provider's configuration and change management capabilities. Where these controls do not provide sufficient notification, traceability and assurance for the intended use and risk, the model or supplier should be considered unsuitable.

Recommendation to CDRH

CDRH may specifically consider third-party initiated changes such as those to foundation models as applicable within the PCCP framework under appropriate control conditions, while considering technical means such as periodic re-benchmarking. CDRH should expect sponsors to assess third-party providers' configuration and change-management capabilities as part of supplier qualification and model selection.

VII. Other Topics

Question 25

> ***FDA question:*** *Would voluntary Foundation Model MAFs be practical for foundation model developers to provide and to keep current, and what content might they include? Given that participation would be voluntary and that model developers may have limited incentive to disclose safety-relevant information, what would make such a program sufficiently useful for premarket review? Are there alternative mechanisms CDRH should consider for obtaining information about underlying foundation models?*

Position summary

Voluntary measures, such as codes of conduct, are implemented under the EU AI Act (see references), while specific requirements are also imposed for certain general-purpose models. ISPE would welcome alignment with such considerations to leverage existing transparency efforts.

Rationale/detail

For instance, Annex XII of the EU AI Act demands technical documentation for providers of general-purpose AI models to downstream providers that integrate the model into their AI system, including tasks, acceptable use, versions, architecture, modalities, information used on the data for training, testing and validation, and where applicable the type and provenance of data and curation methodologies.

Recommendation to CDRH

CDRH may consider comparable frameworks encouraging transparency of foundation model providers, such as those implemented under the EU AI Act.

Supporting evidence/citations

https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Question 26

> ***FDA question:*** *Are there additional considerations that inform the premarket and postmarket evaluation of agentic GenAI-enabled devices, beyond those applicable to non-agentic GenAI-enabled devices? How should the elevated risk associated with autonomous multi-step action, tool use, and reduced opportunity for human review be reflected in acceptance criteria and oversight?*

Position summary

ISPE considers the same principles as for general Generative AI applicable; this includes specification and targeted selection of contextual data, use of tools and choice of human quality gates, alongside guardrails (e.g., input controls, controls on multi-agent behavior), while demonstrating their effectiveness and suitability. These principles should stress the use of risk management when applying these principles.

Recommendation to CDRH

CDRH may consider using, in guidance, the same principles as for general Generative AI, using the same or very similar language and terms. It is recognized that providing illustrative examples is a challenge in guidance.

End of comments