

October 19, 2026

Re: Docket Number FDA-2026-N-7874; Considerations for the Regulation of Generative AI-Enabled Medical Devices: Discussion Paper and Request for Feedback

The Radiological Society of North America (RSNA) appreciates the opportunity to provide input on the Food and Drug Administration (FDA) Center for Devices and Radiological Health discussion paper on generative artificial intelligence (GenAI)-enabled medical devices [Docket Number FDA-2026-N-7874].

RSNA is a leading global organization representing more than 53,000 radiologists and medical imaging professionals across 150 countries. Radiology is among the most data-intensive fields in medicine and has extensive experience evaluating, implementing, and monitoring AI-enabled technologies in clinical practice.

RSNA supports FDA's effort to develop a nimble, risk-based, total product lifecycle approach for GenAI-enabled devices. These technologies may improve the quality, timeliness, and accessibility of care, but their stochastic outputs, broad task capabilities, dependence on foundation models, and potential to evolve after deployment create risks that cannot be addressed through one-time premarket review alone. Patient safety should remain the central organizing principle, and regulatory flexibility should be paired with clear, enforceable safeguards that promote transparency, accountability, clinical validation, and continuous real-world monitoring.

Our comments focus on six priorities: (1) broadening the proposed risk framework; (2) pairing competency-based benchmarking with clinical confirmation; (3) establishing minimum, risk-based continuous monitoring expectations; (4) assigning complementary responsibilities without diffusing manufacturer accountability; (5) strengthening change control, version transparency, and foundation-model governance; and (6) ensuring scalable implementation across healthcare settings.

I. Summary of Recommendations

- FDA should retain the proposed activity-by-consequence framework as a useful starting point, but it should not be the sole basis for risk classification. Risk also depends on the user's ability to independently review an output, the reversibility and time sensitivity of resulting actions, the availability of downstream safeguards, the breadth of the intended use, and the traceability of outputs to source information.

- A competency-based approach can complement existing device evaluation, provided it assesses the final deployed configuration and is paired with clinical confirmation that reflects the intended population, workflow, users, modalities, and clinical setting. Benchmark performance alone is insufficient to establish safe and effective real-world use.
- FDA should not accept materially greater premarket uncertainty for high-consequence, autonomous, action-taking, or poorly reviewable functions. Where a limited shift toward postmarket evidence is permitted, it should depend on an enforceable monitoring plan, validated detection thresholds, timely escalation and corrective-action requirements, transparent versioning, and FDA access to meaningful real-world evidence.
- Continuous monitoring should evaluate the clinical AI system, not the model alone. This includes model performance, subgroup performance, human-AI interaction, workflow effects, appropriate use, model and system changes, and cumulative effects when multiple AI tools operate in the same environment.
- Manufacturers must remain accountable for device safety and effectiveness. Healthcare organizations, clinicians, professional societies, and standards bodies have complementary roles based on what each can observe and control, but those roles should not substitute for manufacturer obligations.
- FDA should strengthen Predetermined Change Control Plans (PCCPs) for GenAI-enabled devices, require timely and detailed communication of changes to models, prompts, and system configurations, and establish mechanisms for end users to report performance deviations and request review of change-control plans based on real-world experience. Manufacturers should provide sufficient transparency to support safe, validated customization for local patient populations and clinical workflows while protecting proprietary intellectual property.

II. Risk Assessment Should Reflect Clinical Context and Reviewability

RSNA agrees that the degree of device activity and the consequence of relying on an incorrect output are central dimensions of risk. However, a two-axis framework may understate risk where outputs appear informational but are difficult for users to independently evaluate, are presented with high fluency or authority, or influence downstream decisions indirectly. FDA should therefore apply the framework as an organizing heuristic within a broader, multidimensional assessment.

The framework should explicitly account for reviewability, action reversibility, time pressure, availability of downstream safeguards, output traceability, intended-user expertise, and the breadth and variability of the device's functions. A radiology output that cannot be independently verified without specialist expertise may warrant greater oversight even if it is labeled non-directive. Similarly, an output used in an urgent stroke, trauma, or oncology workflow may pose greater risk because the opportunity to detect and correct an error is limited.

When clinical oversight is relied upon to mitigate risk, the intended end user should have the qualifications and training necessary to independently perform the equivalent clinical task and assess

the validity of the AI output. FDA should consider the qualifications of the intended end user when evaluating whether human oversight provides a meaningful safeguard.

FDA should assess directiveness based on the substance, specificity, personalization, presentation, and foreseeable effect of an output, rather than terminology alone. Repeated suggestions, ranked options, confident recommendations, workflow prioritization, or default selections may influence action even when the device avoids words such as “recommend.” For multi-turn systems, intended use and risk should be evaluated across realistic conversational trajectories, including foreseeable migration from information retrieval to recommendation or action.

FDA should require manufacturers to establish safeguards that identify when user requests or system behavior move beyond a device's validated indications for use. These safeguards should provide clear warnings about the limitations of the device's validated performance without unnecessarily restricting appropriate clinical judgment. For conversational systems, safeguards should operate across multi-turn interactions rather than evaluating individual prompts in isolation. Manufacturers should validate these safeguards and provide auditable records of detected deviations, warnings, and subsequent system actions.

Patient-facing and generalist-facing functions should not automatically be treated as unsafe, because they may improve access and engagement. Risk mitigation should instead reflect the user's ability to understand limitations and independently evaluate the output. Safeguards may include plain-language explanations, traceable source information, calibrated uncertainty, explicit limits on scope, clear escalation pathways, specialist review for higher-risk use, and testing across health literacy levels and communication styles.

III. Competency-Based Evaluation Must Be Coupled with Clinical Confirmation

RSNA supports further development of a competency-based approach that combines structured device benchmarking with clinical confirmation. This approach is well suited to systems with open-ended inputs and outputs, but it should supplement, not replace, application-specific analytical and clinical validation. Evaluation should address the final user-facing device in its deployed or representative configuration, including prompts, retrieval components, guardrails, orchestration logic, user interface, and relevant third-party dependencies.

Benchmarking should be risk-based and multidimensional. In addition to clinical knowledge and task fidelity, FDA should require evaluation of reproducibility, stochastic variation, hallucination and confabulation risk, boundary adherence, calibration and uncertainty communication, subgroup performance, robustness to distribution shift, workflow integration, automation bias, usability, and foreseeable failure modes. For radiology applications, testing should reflect variation in modalities, manufacturers, scanners, imaging protocols, acquisition conditions, patient populations, disease prevalence, and clinical practice settings.

Public benchmarks are useful but may be affected by data contamination, saturation, and limited representativeness. Sponsors should demonstrate construct validity and explain how benchmark

performance predicts safe and effective behavior for the intended use. FDA should encourage sequestered datasets, independent expert adjudication, multi-institutional testing, predefined acceptance criteria, and transparent reporting of dataset composition and limitations. Sponsor-developed benchmarks may be appropriate for specialized functions but should be subject to independent review and safeguards against optimization to the test.

Evidence requirements for GenAI-enabled devices should reflect the differences between AI and traditional therapeutic interventions. For many diagnostic and workflow applications, clinical benefit depends not only on device performance but also on how clinicians interpret and act on AI outputs. Moreover, performance may vary across patient populations, imaging equipment, and clinical settings, limiting the generalizability of a single clinical trial. FDA should therefore support flexible, risk-calibrated evidentiary approaches that incorporate real-world evidence, multi-institutional validation, and prospective clinical studies as appropriate to the intended use and potential consequences of error.

Clinical confirmation should be proportionate to risk. Lower-risk assistive functions may be supported by retrospective multi-site studies or silent prospective evaluation. Higher-risk, action-directing, autonomous, or poorly reviewable functions should require stronger evidence, including prospective evaluation when appropriate. FDA should assess both device-alone performance and human-AI team performance when the intended use depends on clinician interaction. A technically accurate system may still create harm through poor workflow integration, alert fatigue, automation bias, or unclear communication.

IV. Continuous Monitoring Should Be Standardized, Actionable, and System-Focused

RSNA strongly supports robust postmarket monitoring but cautions against using postmarket monitoring as a general substitute for adequate premarket evidence. Greater premarket uncertainty may be acceptable only for lower-risk, reviewable functions when the expected benefits are meaningful and a mature, enforceable monitoring infrastructure is in place. It should not be accepted for high-consequence, autonomous, action-taking, or difficult-to-review functions where an error may cause serious or irreversible harm before it can be detected.

Continuous monitoring should be an ongoing lifecycle process that can detect signals, review their clinical significance, intervene when appropriate, and reassess after intervention. It should evaluate the full clinical AI system, including local model performance, human-AI interaction, patterns of appropriate and inappropriate reliance, workflow effects, intended use, off-label use, subgroup performance, model and system versions, and changes over time.

FDA should establish minimum manufacturer expectations that are adaptable to clinical risk, technological characteristics, and degree of reviewability. Monitoring plans should identify measures, data sources, cadence, thresholds, triggering events, responsible parties, escalation pathways, and actions that may follow a significant signal. Triggering events should include model or software updates, changes to an underlying foundation model, changes in imaging equipment or firmware,

significant shifts in patient or data distributions, material workflow changes, cybersecurity events, and observed changes in error patterns.

Periodic re-benchmarking, sample-based clinician review, and performance-degradation monitoring are complementary and should be combined where appropriate. FDA should also consider concordance and discordance analysis, review of near-misses, structured end-user feedback, subgroup performance surveillance, audit-log review, and monitoring of processing failures or incomplete analyses. Machine-based supervisory agents may assist with monitoring, but they should themselves be validated, versioned, auditable, and monitored. Automated supervision should not become an opaque second layer of risk.

V. Responsibilities Should Follow Observability, Capability, and Control

No single stakeholder can fully observe or manage all risks in a deployed GenAI-enabled device. FDA should establish complementary responsibilities while preserving clear manufacturer accountability:

- Manufacturers should monitor technical performance, model integrity, processing failures, updates, known limitations, and safety signals; investigate product-related degradation; provide monitoring-enabling information and safeguards to identify and communicate use beyond validated indications for use; and support timely corrective action.
- Healthcare organizations should maintain governance, data capture, model and version tracking, audit trails, escalation processes, and the ability to restrict, pause, or discontinue use when warranted.
- Clinical AI and specialty teams should define clinically meaningful measures and thresholds, interpret signals, evaluate human-AI performance, and recommend revalidation or intervention.
- End users should understand the device's intended use and limitations, exercise appropriate clinical oversight, and have easily accessible mechanisms to report suspected errors, unexpected behavior, and workflow or safety concerns.
- Professional societies and standards organizations can develop consensus measures, independent benchmarks, implementation guidance, education, and multi-institutional monitoring infrastructure.

These complementary roles should not diffuse accountability. Manufacturer responsibility should remain primary for the safety and effectiveness of the marketed device, including performance changes caused by components or third-party models incorporated into the device. FDA should establish a dual-reporting pathway that allows manufacturers and healthcare organizations or clinicians to report significant performance deviations directly to the agency, especially when vendor response is delayed or inadequate.

FDA should provide greater clarity regarding accountability when AI-enabled devices contribute to patient harm. As generative AI systems assume increasingly autonomous and near-autonomous functions, responsibility should be proportional to each stakeholder's ability to observe, evaluate, and

control system behavior. Clinicians should not become the default bearers of liability for errors arising from AI systems they cannot reasonably oversee or control.

VI. Transparency and Monitoring-Enabling Information Are Essential

Healthcare organizations cannot meet monitoring expectations without timely access to information needed to identify, investigate, and act on safety signals. FDA should require manufacturers to provide model and system identifiers; complete version histories; descriptions of changes and expected performance effects; known limitations and failure modes; relevant input and output data specifications; processing status and error information; subgroup performance information; and audit information sufficient for root-cause analysis.

For generative AI-enabled devices, manufacturers should provide healthcare organizations with sufficient transparency regarding the prompts, system instructions, and configurations that govern device behavior. This transparency is necessary to support clinical evaluation, ongoing monitoring, and appropriate adaptation to local patient populations and workflows. Manufacturers should support controlled customization and validation of these components when appropriate, while maintaining protections for proprietary intellectual property. Changes that could materially affect clinical performance should be documented and subject to appropriate revalidation.

When clinically meaningful, point-of-care transparency should indicate whether eligible inputs were successfully analyzed, which images, series, text, or other inputs were used, whether exclusions or processing errors occurred, and, where validated and appropriate, provide output-specific uncertainty or confidence information for the individual patient rather than relying solely on general statements in model documentation. Because GenAI systems can produce convincing but incorrect outputs, this information should help users recognize when additional review or verification is warranted. Static documentation is not sufficient when safe interpretation depends on information available within the workflow.

FDA should require standardized model cards and dataset reporting elements that are updated with each material version. For foundation models supporting multiple applications, documentation should include a core model description and application-specific sections that distinguish validated uses from exploratory or off-label uses.

VII. PCCPs and Foundation Model Dependencies Require Stronger Change Controls

PCCPs can support iterative improvement, but future changes to GenAI-enabled devices may be difficult to fully pre-specify. FDA should permit flexibility only within defined safety boundaries and require sponsors to identify the types of changes, methods of evaluation, acceptance criteria, communication procedures, and circumstances that require a new submission. Changes affecting intended use, autonomy, user population, clinical domain, safety-critical behavior, or validated performance boundaries should generally require heightened review.

FDA should require timely notice to deploying organizations before material updates, except when urgent security or safety action is necessary. Notices should explain what changed, why it changed, which functions or populations may be affected, how performance compares with the prior version, and whether local revalidation is recommended. Healthcare organizations should retain access to prior version information for diagnostic consistency, troubleshooting, and root-cause analysis.

For devices that rely on third-party foundation models, the device manufacturer must remain responsible for detecting, evaluating, and responding to upstream changes. Contracts, technical controls, version pinning, change notifications, audit logs, fallback procedures, and re-benchmarking requirements should support this responsibility. A voluntary “Foundation Model Device Master File” could improve regulatory efficiency, but participation should not reduce the device sponsor’s obligation to demonstrate safety and effectiveness for each intended use. FDA should consider incentives for participation and minimum update-notification commitments so that the files remain current and useful.

VIII. Agentic Systems Warrant Heightened Oversight

Agentic systems may plan and execute multi-step actions, interact with external tools, and propagate errors across interconnected workflows. FDA should evaluate not only individual outputs but also decision pathways, tool selection, sequencing, autonomy boundaries, handoff quality, error recovery, and cumulative downstream effects. Acceptance criteria should be more stringent as autonomy and consequences increase.

High-consequence or irreversible actions should include validated human-authorization checkpoints unless the sponsor provides compelling evidence that autonomous operation offers a favorable benefit-risk profile. Systems should maintain sufficiently detailed, reviewable audit logs of prompts, retrieved information, actions, tool outputs, failures, overrides, and human interventions to support clinical oversight, incident investigation, and accountability.

IX. FDA Should Support Scalable Implementation and Harmonization

Meaningful monitoring should not depend on the resources of an individual healthcare organization. Less resourced and rural institutions may be unable to build independent monitoring systems, creating uneven safety oversight. FDA should coordinate with other Department of Health and Human Services agencies to support interoperable data and reporting standards, shared monitoring tools, multi-institutional validation networks, quality datasets, professional education, and research on human-AI interaction.

FDA should work with clinical experts, professional societies, standards organizations, patients, healthcare organizations, and industry to define common data elements for AI identification and versioning, provenance, processing status, outputs, audit information, and safety reporting. Harmonized standards should reduce duplicative local burden and support cross-institutional learning while preserving privacy and proprietary protections. These efforts should build upon existing interoperability standards, including DICOM, FHIR, and established common data element frameworks, wherever appropriate.

Conclusion

RSNA supports FDA's effort to modernize oversight for GenAI-enabled medical devices. A successful framework should be risk-based and adaptable, but also sufficiently rigorous to address variable outputs, broad capabilities, foundation-model dependencies, and increasing autonomy. Premarket evaluation and postmarket monitoring should function as complementary safeguards, not substitutes.

RSNA encourages FDA to center patient safety; evaluate the final clinical AI system rather than the model alone; require clinically meaningful transparency and traceability; align responsibilities with stakeholder observability, capability, and control; preserve manufacturer accountability; and support scalable monitoring infrastructure. These principles can enable innovation while maintaining the public and professional trust necessary for GenAI-enabled devices to improve patient care.

RSNA appreciates the opportunity to provide input and looks forward to continued collaboration with FDA. For additional information, please contact Libby O'Hare, Director of Government Relations, at eo hare@rsna.org.

Sincerely,

A handwritten signature in black ink, appearing to read 'Carolyn C. Meltzer', with a long horizontal line extending to the right.

Carolyn C. Meltzer, MD
Chair of the Board
Radiological Society of North America