

October 19, 2026

Michelle Tarver
Director
Center for Devices and Radiological Health
Food and Drug Administration
10903 New Hampshire Ave.
Silver Spring, MD 20993

Submitted Electronically

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

Dear Director Tarver,

On behalf of our nearly 5,000 member hospitals, health systems and other healthcare organizations, our clinician partners — including more than 270,000 affiliated physicians, 2 million nurses and other caregivers — and the 43,000 healthcare leaders who belong to our professional membership groups, the American Hospital Association (AHA) appreciates the opportunity to provide feedback to the Food and Drug Administration (FDA) on regulatory considerations for generative artificial intelligence (GenAI)-enabled medical devices.

AI-enabled healthcare applications are transforming care delivery in innumerable ways, holding tremendous promise for improving healthcare outcomes and reducing costs. As GenAI technologies advance, policymakers will grapple with novel challenges related to model bias, hallucinations and model drift, which are not yet fully accounted for in existing medical device frameworks. AI tools are inherently designed to be agile and adaptive, taking in new data points, discerning patterns and continually updating to improve model accuracy. This is especially true for GenAI. The FDA will face questions about which GenAI applications constitute medical device versus non-device functions, and how the agency's regulatory frameworks should account for these tools.

In general, the AHA supports a flexible, risk-based regulatory framework for GenAI-enabled medical devices that protects patients without impeding responsible innovation. To achieve this balance, the AHA recommends that the FDA:



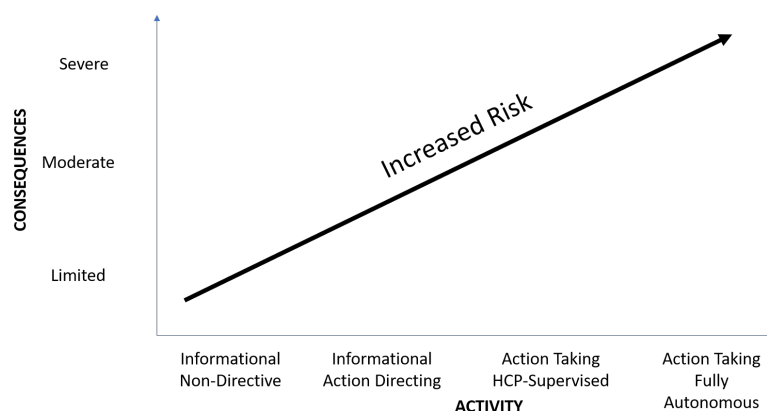
- Refine the proposed risk heuristic for evaluating the risk of GenAI applications.
- Provide labelling guidance on “informational action directing” devices versus “action taking” devices.
- Align premarket clearance and post-deployment evaluation and monitoring activities under a total product lifecycle approach.
- Clarify that GenAI-enabled clinical decision support and administrative tools are not subject to FDA oversight.
- Assign primary responsibility for premarket and post-deployment monitoring and evaluation activities to developers.

Below are our detailed comments.

CONSIDERATIONS FOR RISK ASSESSMENT

FDA’s discussion paper proposes a risk-based approach to generative AI-enabled medical device evaluation. The request for information (RFI) seeks feedback on a conceptual framework for assessing risk, which the FDA has structured as a heuristic comprised of two variables. The first variable reflects the “activity level” of the GenAI-enabled device, ranging from providing informational but non-directive outputs to a device user to fully autonomous action. The second variable reflects the “consequences” of the GenAI-enabled medical device’s actions, ranging from limited to severe. Figure 1, drawn from the RFI, includes a diagram of this risk-based approach.

Figure 1. FDA’s proposed 2-variable risk heuristic for generative AI-enabled medical devices



Below is our detailed feedback on the proposed risk heuristic.

Expand Risk Heuristic to Account for Variables Like Probability of Consequence

The AHA generally supports the pursuit of risk-based approaches for the evaluation of generative AI-enabled medical devices. However, we recommend the FDA add

variables to its proposed heuristic, such as reversibility, downstream safeguards and output traceability in its risk-based evaluation. The AHA also recommends that the FDA add variables for likelihood or probability of consequences, which are currently missing from the risk matrix and discussion paper. The likelihood of a negative outcome, error (type I or II) or other consequence is a vital component of risk evaluation, especially in the context of statistical tools like GenAI.¹

Update or Remove Proposed “Activity Level” Variable

The FDA’s proposed framework correctly recognizes that the activity level of GenAI devices exists on a spectrum. However, activity level alone may not sufficiently predict the risk of particular devices, and the relationship to risk and activity level is not as linear as the FDA’s heuristic suggests. As such, the AHA encourages the FDA to reconsider the structure of this variable.

For example, the discussion paper suggests that “informational action directing” activities may include a “strongly worded patient-facing recommendation about whether to seek emergency care.” Recent studies of certain commercial products on triaging efficacy demonstrate a range of risks or consequences for these tools. A recent Nature article found that certain commercial AI tools over-triaged non-emergent cases and under-triaged emergency conditions.² Specifically, the tools had a 52% mis-triage rate for certain patients that required escalation to an emergency department (e.g., diabetic ketoacidosis patients or patients with impending respiratory failure). At the same time, the tools over-triaged 64% of non-emergent cases. As the article states, “The 51.6% undertriage rate for true emergencies represents the most concerning finding, as missed emergencies can result in patient harm, while the 64.8% overtriage rate for nonurgent cases, although less dangerous, risks unnecessary healthcare utilization at scale.”³

In the context of the FDA’s heuristic, this study demonstrates that “informational action directing” applications, in and of themselves, do not necessarily present lower consequences or risk than “action taking HCP-supervised applications.” As the article states, “The undertriaged cases—rising pCO₂ signaling respiratory failure, metabolic acidosis in DKA—are presentations that no experienced clinician would delay.”⁴ For these reasons, the FDA’s heuristic should consider variables, such as likelihood of error, negative outcome or consequence, as more heavily weighted than whether an activity is informational versus action-directing.

¹ In statistical models, a “type I” error denoted by α occurs when there is a false positive result (i.e., incorrectly rejecting a null hypothesis). A “type II” error denoted by β occurs when there is a false negative result (i.e., incorrectly failing to reject a null hypothesis).

² <https://www.nature.com/articles/s41591-026-04297-7>

³ Ibid.

⁴ Ibid.

Clarify Labeling Guidance to Distinguish Medical Devices and Non-medical Devices, and Informational and Action-taking Devices

The example above related to triaging efficacy also underscores the importance of clear product labelling. As we have previously commented, for consumers, it can be confusing to distinguish between medical and non-medical device applications, or devices intended to be informational versus action-directing.⁵ This is particularly true for models that rely on similar types of physiological data and generate similar recommendations or outputs, yet differ in their intended use, validation processes and regulatory status.

We encourage the FDA to update guidance on labelling in collaboration with the Federal Trade Commission (FTC). Such guidance should describe how the FDA and FTC would distinguish between informational and action-taking devices, and suggest disclaimer language for informational products. Additionally, this should specify the level to which healthcare professionals are in the decision loop and the type of clinician (general versus specialist) involved. Transparency in labelling can better inform consumers of the extent to which a tool is autonomous and the type of clinician in the loop. The FTC has a broad mandate to prevent unfair or deceptive acts or practices, and the FTC and FDA share jurisdiction over marketing of certain health-related products (including devices).

TOTAL PRODUCT LIFECYCLE APPROACH TO REGULATION OF GENAI-ENABLED MEDICAL DEVICES

The discussion paper requests feedback on the applicability of total product lifecycle (TPLC) approaches to the oversight of GenAI-enabled medical devices, an approach the FDA has long promoted for medical device regulation. In general, the AHA agrees that risk-based evaluation should be integrated into a TPLC approach for GenAI-enabled medical device oversight. A risk-based approach to measurement and evaluation aligned with TPLC could focus scarce resources, such as time, personnel and cost, on the highest-risk applications and also coincide with the approach of domestically accredited and international standards development groups.

GenAI is perhaps one of the most prominent examples where TPLC approaches, and alignment of premarket clearance requirements with post-deployment monitoring activities based on risk, would be most beneficial because GenAI models are constantly evolving. The potential for model drift and hallucinations underscores the importance of post-deployment monitoring activities in ways that perhaps previous static devices did not necessitate.

⁵ <https://www.aha.org/letterscomments/2026-08-13-aha-responds-fda-request-input-development-21st-century-cures-act-section-3060-required-report>

COMPETENCY-BASED APPROACH FOR PREMARKET EVALUATION

The discussion paper also requests feedback on a two-part competency-based approach for premarket evaluation comprised of non-clinical device benchmarking and clinical confirmation. Device benchmarking elements would determine whether a GenAI-enabled device demonstrates the clinical knowledge, analytic capabilities, safety behavior, communication and generalizability to support reasonable assurance of device safety and effectiveness for its intended use. The agency acknowledges that benchmarking alone may not fully account for real-world conditions, and therefore proposes integration of clinical confirmation activities in premarket evaluation to provide review of clinically representative conditions.

The AHA agrees in concept with a competency-based approach for premarket evaluation, but encourages clarity of sponsor and developer requirements. Specifically, when it comes to clinical confirmation, providers often lack visibility into an AI model's underlying weights, making it difficult to determine alignment with real-world patient populations. Therefore, the agency should clarify device benchmarking and clinical confirmation activities that are the responsibility of the sponsor/device developer.

The AHA also encourages the FDA not to unintentionally restrict clinicians in utilizing clinical judgment for tools. The discussion paper references standards of performance and proposes comparing performance to a “panel of qualified clinicians whose consensus reflects the applicable standard.” Instead of relying on a panel, we encourage the agency to leverage existing consensus-based clinical practice guidelines. Standards should be transparent and should be denoted in labelling.

The FDA also explores the role of third parties in evaluation and benchmarking. If this were pursued, we encourage the agency to explore models where multiple third parties could support instead of a single third party. Having a single third-party evaluator could limit competition and artificially inflate pricing.

POST-MARKET MONITORING AND EVALUATION

The discussion paper also requests input on post-market monitoring for GenAI-enabled devices. Our detailed feedback is below.

Synchronize Post-deployment Monitoring Activities With TPLC to Reduce Pre-clearance Barriers

As previously noted, opportunities exist to better synchronize post-deployment measurement and evaluation activities in ways that could reduce unnecessary premarket barriers.⁶ For example, currently over 96% of AI-enabled medical devices are

⁶ <https://www.aha.org/lettercomment/2025-12-01-aha-letter-fda-ai-enabled-medical-devices>

approved under the 510k clearance process, but the 510k process caps the number of indications for which applicants can seek approval at a given time. By its very nature, GenAI is adaptable, resulting in new capabilities that could augment the number of clinical indications for which a particular AI application is suited. Yet, the current 510k framework generally requires vendors and other developers to submit multiple clearance applications for substantially similar technologies. This approach may limit the ability to expand otherwise safe and effective uses of AI and result in added costs to vendors and providers alike, as application fees have often been passed down to end users through increased prices.

The AHA encourages the FDA to explore clearance pathways whereby vendors and other developers can provide detailed post-market evaluation and monitoring plans, including additional special controls, to streamline 510k clearances under fewer applications. Such an approach could shorten the time to get evidence-based, safe uses of AI approved for use, benefiting patients and providers. It could also help reduce regulatory burden and cost.

Clarify Boundaries of Device and Non-device Definitions

The FDA also should clarify that measurement and evaluation standards would not apply to clinical decision support tools and administrative AI tools. The 21st Century Cures Act excluded certain AI tools from the definition of “medical devices,” so long as they support clinical decision-making, with the clinician independently reviewing recommendations. Clinical decision support tools are often developed to address specific clinical challenges over limited periods. Additional requirements can add barriers to delivering timely solutions and may be impractical if the tool is only designed to support clinicians for a short timeframe (like six months). To maximize scarce resources, the FDA should focus monitoring and evaluation activities on higher-risk areas related to the diagnosis of disease or other conditions, or in the cure, mitigation, treatment or prevention of disease.

Place Primary Post-market Monitoring Responsibility on Developers

While hospitals and health systems support evaluation and monitoring activities, vendors and developers must be responsible for the ongoing performance, maintenance and integrity of the tools they market. Hospitals and health systems have broadly developed multi-disciplinary governance teams to review AI use cases, technologies and value over time. According to the AHA’s Health IT supplemental survey, 74% of hospitals have multiple teams responsible for evaluating predictive AI, including (but not limited to) senior leaders, department leaders and IT staff, with many having specific committees for machine learning and/or clinical decision support.⁷

⁷ <https://healthit.gov/data/data-briefs/hospital-trends-use-evaluation-and-governance-predictive-ai-2023-2024/>

Hospitals and health systems continually assess the strengths and limitations of AI models they use, but the “black box” nature of many AI systems can make it more challenging for hospitals and health systems to identify flaws in models that may affect the accuracy and validity of an AI tool’s analyses and recommendations. As such, post-market measurement and evaluation standards should be developed for vendors. Standards should include metrics for performance, thresholds for further evaluation, and communication requirements to end users related to ongoing performance.

Additionally, some hospitals (particularly rural, critical access and other safety net hospitals) may not have the staff or resources to support governance structures and ongoing measurement activities. As such, many of these healthcare providers have not deployed AI tools. Resource and infrastructure barriers can exacerbate the “digital divide” in certain geographies, where rural and other underserved areas have less access to digital services, including AI tools for clinicians and patients. While addressing incentives and broader infrastructure barriers to AI is not the sole responsibility of the FDA, we encourage cross-agency collaboration to develop training, technical assistance and potential grant funding opportunities to support educational efforts and governance activities for AI health tools. This includes education and resources to support evaluation and monitoring activities. Cross-agency coordination could include collaboration among Health and Human Services, the FDA, the Federal Communications Commission, the Department of Commerce, the Department of Agriculture and the Department of Education.

OTHER TOPICS

The agency also requests feedback on a variety of other related topics like Model Device Master Files and voluntary submission of AI model cards to the FDA. We note that transparency of AI model cards has been a part of the Office of the National Coordinator for Health IT’s (ONC) certified health IT program. We have heard from hospitals that this criterion is one of the few examples of transparency standards providing information on AI to help inform tool procurement and implementation. One of the primary barriers to AI adoption has been the “black box” nature of algorithms. This criterion provides information on how a predictive or GenAI application was designed, developed, tested, evaluated and should be used. The AHA opposed the proposed removal of these criteria in ONC’s HTI-5 proposed rule, and continues to advocate that transparency of model card data is imperative to foster trust in AI tools and ensure patient safety.⁸

The AHA looks forward to working with the FDA on a risk-based approach to GenAI-enabled device regulation that protects patients, provides clear expectations for developers, aligns with operational realities of care delivery and supports responsible

⁸ <https://www.aha.org/lettercomment/2026-02-27-aha-comments-astponc-health-care-technology-interoperability-proposed-rule>

Director Michelle Tarver

October 19, 2026

Page 8 of 8

innovation. Please contact me if you have questions, or feel free to have a member of your team contact Jennifer Holloman, AHA director of health IT policy, at jholloman@aha.org.

Sincerely,

/s/

Ashley Thompson

Senior Vice President

Public Policy Analysis and Development