Cooley

Regulating AI Like a Doctor: FDA Floats Competency-Based Path for Generative AI-Enabled Devices

September 28, 2026

On August 18, 2026, the US Food and Drug Administration (FDA) Center for Devices and Radiological Health (CDRH) released a discussion paper proposing new approaches to regulating medical devices enabled by generative artificial intelligence (genAI), and requesting stakeholder feedback, which must be submitted by October 19, 2026. The paper is not guidance, but it offers insight into how the agency is thinking about the unique risks and evidentiary challenges posed by genAI-enabled devices. The paper presents two main ideas: a two-axis framework for assessing risk and a "competency-based" model for premarket evaluation inspired by how human clinicians are evaluated and credentialed.

Background

CDRH has been concerned about genAI-enabled medical devices for some time. Following a November 2024 public meeting of its Digital Health Advisory Committee, CDRH identified two categories of regulatory challenges posed by these devices:

1. Challenges associated with applying a risk-based approach to classifying and determining regulatory requirements for genAI-enabled devices.
2. Challenges associated with determining the types of valid scientific evidence for FDA's evaluation of the safety and effectiveness of such devices across the total product life cycle.

This discussion paper offers potential ideas to address these challenges and requests stakeholder feedback.

The two-axis framework

CDRH proposes organizing risk assessment for genAI-enabled software functions in a chart with the horizontal axis representing the degree and independence of the device's activity and the vertical axis representing the severity of consequences if a user relies on an incorrect output. Risk increases moving from the lower left toward the upper right of the line chart. (See Figure 1)

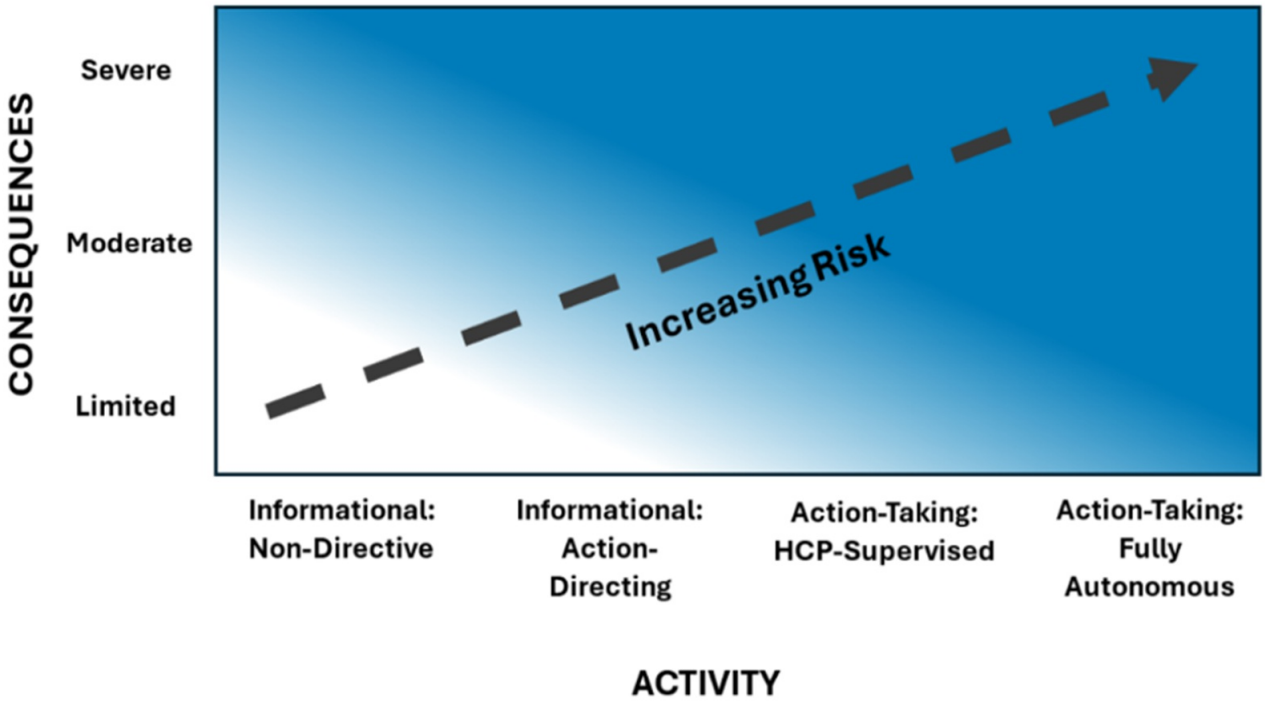

Figure 1: FDA's two-axis framework for risk

Within that framework, CDRH puts forth several specific factors that it suggests should shift a function's risk classification, including, for example:

- **Measurement and signal processing functions.** FDA claims that genAI-enabled in vitro diagnostic, measurement and signal processing functions raise increased concern as their outputs typically cannot be independently assessed by the user. Even though such functions do not direct users to perform an action (a function CDRH considers higher risk than simply providing nondirective information), CDRH proposes to treat them as higher risk because a user cannot identify and avoid relying on an incorrect result.
- **Patient-facing versus HCP-facing informational functions.** CDRH also proposes to weigh whether informational functions delivered directly to patients should be treated as higher risk than the same functions delivered to healthcare professionals (HCPs), since FDA believes that HCPs are generally better positioned to interpret an output, recognize limitations, spot an error and avoid over-relying on it.

Interestingly, both of these considerations appear to come directly from concepts that are structural to FDA's clinical decision support (CDS) software guidance. The CDS guidance's nondevice carve-out excludes software that acquires, processes or analyzes signals from an in vitro diagnostic device or a signal acquisition system and requires devices meeting the carve-out to enable an HCP (explicitly not a patient or caregiver) to independently review the basis for a recommendation. CDRH draws that connection itself in the discussion paper, in a footnote where it notes that its concern about measurement and signal processing functions "seems consistent with the independent- review criterion of the clinical decision support exclusion."[1] Despite criticism that the CDS guidance significantly narrows the statutory carve-out from the 21st Century Cures Act, it appears CDRH proposes to import some of the CDS guidance's risk logic into a broader framework for genAI-enabled devices generally.[2]

CDRH also flags other factors it is considering, including how to treat "action-directing" versus "action-taking" functions – terms CDRH uses descriptively as part of its proposed risk framework rather than as terms with fixed regulatory definitions. CDRH has not defined these categories with precision and is specifically asking stakeholders to help refine the distinctions between them. Other factors under consideration include generalist HCP-facing functions versus specialist-facing ones, multi-turn conversations that migrate from informational to action-directing over time and care escalation functions. CDRH specifically requests feedback on this proposed risk framework.

Competency-based model for premarket evaluation

CDRH also previews a possible shift in how genAI-enabled devices could be evaluated by FDA before they are permitted to be marketed. Traditional software device evaluation has relied on testing across a representative sample of defined inputs and outputs. However, CDRH notes that this approach likely will not work for genAI-enabled devices as the range of possible inputs and outputs may be too large for such testing to be practical. Instead, citing recent academic papers that propose evaluating genAI more like a clinician than an unchanging product, CDRH is considering a two-part "competency-based" evaluation model, which includes device benchmarking and clinical confirmation. CDRH anticipates that the scope of benchmarking and the rigor of clinical confirmation would scale with the device's position on the two-axis risk framework.

- **Device benchmarking**. CDRH envisions device benchmarking as a scalable, nonclinical evaluation to assess whether the device, in its deployed configuration, demonstrates the clinical knowledge, analytic capabilities, safety behavior, communication and generalizability to support reasonable assurance of safety and effectiveness of the device for its intended use. CDRH proposes evaluating devices across categories of competencies, including safety (e.g., recognizing safety-critical situations and communicating uncertainty), clinical proficiency (e.g., clinical knowledge and analysis), generalizability (e.g., robustness, reliability and reproducibility) and additional agentic-specific competencies for agentic AI functions.
- **Clinical confirmation**. CDRH also believes that device benchmarking alone may not fully establish how a genAI-enabled device will perform in actual clinical use, since such devices interact with users, workflows and patient populations in ways that testing may not capture. CDRH therefore contemplates an additional clinical confirmation step, noting that this would not necessarily require a prospective clinical study in every case. Instead, the type and rigor of evidence would be consistent with the device's risk profile and could range from retrospective evaluation on real patient data to prospective clinical studies or randomized controlled trials.

This competency-based evaluation model would represent a significant departure from traditional premarket review, which has generally relied on testing a device against a defined set of inputs and comparing its outputs to a known standard, such as the paradigm underlying the 510(k) substantial equivalence framework. Under the competency-based approach, by contrast, evaluation would center on whether a device demonstrates proficiency across prespecified categories of clinical knowledge, safety behavior and real-world generalizability – an assessment framework modeled less on product testing and more on how human clinicians are credentialed.

The shift is consequential because it would move FDA toward a qualitative, multidimensional evaluation of device capability. If adopted, this model could fundamentally reshape the FDA's marketing authorization policy for genAI-enabled devices, replacing conventional input-output testing with a credentialing-style assessment of device competency, paired with clinical evidence scaled to risk. Whether and how these proposals take shape can still be influenced by industry engagement during the comment period and beyond.

Postmarket monitoring

Through the introduction of the competency-based model, CDRH recognizes that the novel capabilities of genAI-enabled devices can make it difficult for premarket testing alone to fully capture device performance. CDRH states that it is therefore considering whether it may be appropriate to accept greater premarket uncertainty regarding a genAI-enabled device's benefit-risk profile, and instead place greater reliance on postmarket monitoring. The discussion paper identifies a handful of potential approaches to postmarket monitoring that are intended to be proportionate to a device's risk profile:

Periodic device benchmarking

The manufacturer would reassess the device against the same prespecified benchmarking thresholds used in the premarket evaluation, on a defined cadence and after defined triggering events, including changes to the underlying model or other components of the deployment architecture. This approach is conceptually similar to FDA's existing Predetermined Change Control Plan (PCCP) framework, under which preauthorized modifications can be implemented without new premarket submissions.

- **Periodic sample-based clinician review**. Qualified, independent clinician adjudicators would review samples of real-world

inputs and outputs against prospectively defined criteria, sampled to reflect device encounters in actual use. Although the discussion paper contemplates the use of qualified independent third parties as clinical adjudicators and describes a general methodology for comparing device outputs against independent clinician assessments, it does not address the institutional infrastructure for this process – for example, how adjudicators would be credentialed, whether they would operate under FDA oversight or institutional review board supervision, or what contractual or governance arrangements would apply. These implementation details remain open questions, consistent with the paper's discussion-only status.

- **Performance degradation monitoring**. The manufacturer would monitor for drift resulting from changes in the input population, data environment or underlying model components.

CDRH also asks for feedback on whether machine-based supervisory agents could help facilitate some aspects of postmarket monitoring, effectively raising the concept of AI agents evaluating AI device functions. This is a notable step for the agency: There does not appear to be precedent for FDA endorsing machine-based oversight of regulated devices, even at the discussion stage, and the proposal underscores the extent to which genAI-enabled devices may require fundamentally different regulatory approaches. Additionally, CDRH proposes a "shared ecosystem responsibility" for postmarket oversight, noting that clinicians, healthcare institutions, payors, professional societies and other stakeholders each may have roles to play in the deployment, monitoring, reporting and ongoing evaluation of genAI-enabled devices.

This framing reflects a broader pattern in FDA's current thinking: Rather than concentrating all oversight within the agency, FDA is increasingly exploring models that leverage external institutions and stakeholders. For example, FDA's Expedited IND Pilot, announced on September 15, 2026, under the Department of Health and Human Services' Operation TrialBlazer initiative, pairs drug sponsors with qualified research institutions to accelerate first-in-human clinical trial timelines through rolling investigational new drug (IND) review – a model that, while structurally different from CDRH's postmarket vision, similarly relies on distributing regulatory functions across a network of qualified participants within the clinical trial ecosystem rather than handling them entirely within FDA. However, it is unclear under what authority FDA could compel such parties to engage with FDA on these issues.

Foundation model master files

Because many genAI-enabled devices are built on third-party foundation models, CDRH is also exploring whether the existing Device Master File program could be leveraged to improve the agency's visibility into those underlying models. Specifically, CDRH seeks feedback on the feasibility of voluntary Foundation Model Master Files (MAFs), under which foundation model developers and platform providers could submit information on their models to FDA. These files would be held confidentially by FDA and could be referenced by device sponsors, with the file holder's authorization, in support of individual premarket submissions. Importantly, CDRH emphasizes that submission of a Foundation Model MAF would not constitute authorization of the underlying model for any device intended use, and device sponsors would remain independently responsible for demonstrating the safety and effectiveness of their own devices.

The voluntary nature of this proposal raises an obvious question: Foundation model developers may have limited commercial incentive to disclose safety-relevant information to a regulator, particularly where doing so could expose limitations or failure modes. CDRH itself acknowledges this tension and asks commenters whether the program could be made sufficiently useful for premarket review given that participation would be optional. If adopted, however, Foundation Model MAFs could streamline premarket review by giving FDA reviewers a baseline understanding of the models underlying multiple device submissions, potentially reducing duplicative information requests and promoting more consistent review across devices built on the same foundation model.

Themes from public comments

Stakeholders filing comments to date have raised a few recurring points. Many commenters, including AI developers, health systems and individual physicians, argue that the two-axis framework should be supplemented with additional risk considerations – such as the detectability of an incorrect output, its reversibility and traceability to the underlying data or specific model – rather than relying on device activity and consequences alone. A similar group cautioned that postmarket monitoring is not an adequate substitute for premarket evidence where harm may be severe, fast-acting or irreversible, and several commenters urge CDRH to require sponsors to affirmatively state what failure modes a monitoring program can and cannot detect before any such trade-off is accepted. Some commenters questioned the feasibility and reliability of the Foundation

Model MAFs, given that some models are continuously changing and updated hundreds of times a day. These comments, along with others addressing agentic AI oversight, human-in-the-loop design and foundation model change management, suggest that industry feedback may be helpful and ultimately shape CDRH's eventual approach.

Practical considerations

As this is a discussion paper, not guidance, CDRH does not propose any new regulatory policy concerning the marketing authorization process or evidentiary requirements at this stage. However, the paper offers a meaningful preview of the agency's thinking, and stakeholders should not underestimate its significance, particularly given the novelty of its proposals. CDRH discussion papers have historically foreshadowed the agency's regulatory direction, and stakeholders who engage early in this process will be better positioned to shape the framework before it solidifies into formal policy, guidance for industry or regulation. Companies developing, deploying or investing in genAI-enabled medical devices should consider several near-term actions:

- **Submit comments by the October 19, 2026, deadline**. The open feedback period is one of the most direct ways for stakeholders to influence CDRH's regulatory approach to genAI-enabled products. Comments submitted through regulations.gov during this window carry particular weight, as CDRH has specifically requested input on its proposed risk framework, competency-based evaluation model and postmarket monitoring approaches. Companies should consider engaging individually or through industry associations.
- **Map existing and pipeline products against the two-axis framework**. Even though the framework is not yet final, companies should begin assessing where their genAI-enabled devices fall along CDRH's proposed risk axes – degree of device activity and severity of consequences from incorrect outputs. This exercise can identify products that may face heightened scrutiny and help prioritize regulatory strategy and resource allocation accordingly.
- **Evaluate evidence generation strategies**. The competency-based evaluation model, if adopted, would provide guidance for companies to demonstrate device proficiency across defined competency categories rather than relying solely on traditional input-output testing. Companies should assess whether their current clinical evidence and testing infrastructure can support device benchmarking and clinical confirmation at the rigor CDRH envisions.
- **Review foundation model supply chain arrangements**. CDRH's proposed Foundation Model Master File program, coupled with its emphasis on deployment architecture and model change management, underscores the importance of contractual and operational visibility into third-party foundation models. Companies that rely on third-party models should evaluate whether their existing agreements provide sufficient access to the safety, performance and change-management information that CDRH may expect device sponsors to possess or reference.
- **Prepare for expanded postmarket obligations**. CDRH's discussion of periodic benchmarking, sample-based clinician review and performance degradation monitoring signals a potentially significant expansion of ongoing compliance obligations for marketed genAI-enabled devices. Companies should consider how these proposals would affect their quality systems, postmarket surveillance capabilities and operational costs.

Cooley's life sciences and healthcare regulatory team will continue to monitor CDRH's approach to genAI-enabled devices and is available to discuss how this discussion paper may affect product roadmaps, regulatory strategies or premarket submission planning. We also can assist in preparing a comment letter to FDA in response to the proposals outlined in this paper.

Notes

1. Discussion paper at 11 n.17 (noting that genAI-enabled measurement and signal processing functions raise concerns "seem[ing] consistent with the independent-review criterion of the clinical decision support exclusion").
2. For more information regarding FDA's CDS guidance, read Cooley's January 2026 client alert.

not guarantee a similar outcome. Do not send any confidential information to Cooley, as we do not have any duty to keep any information you provide to us confidential. When advising companies, our attorney-client relationship is with the company, not with any individual. This content may have been generated with the assistance of artificial intelligence (AI) in accordance with our AI Principles, may be considered Attorney Advertising and is subject to our legal notices.

Key Contacts

Sonia Nath **Washington, DC**	**snath@cooley.com** **+1 202 776 2120**
Son Nguyen Washington, DC	snguyen@cooley.com +1 202 728 7100
Wyatt Kernell Colorado	wkernell@cooley.com +1 720 566 4490
Laura Wenzel Washington, DC	lwenzel@cooley.com +1 202 776 2396