

Response Framework for GenAI-Enabled Medical Devices

Considerations for the Regulation of Generative AI-Enabled Medical Devices
Discussion Paper and Request for Feedback

October 2026

ClaraWell Inc.

Prepared By:

Dr. Emanuel Tkach

Eric Blum

FDA's Considerations for the Regulation of Generative AI-Enabled Medical Devices:

Discussion Paper and Request for Feedback presents 26 discussion questions.

ClaraWell organizes these questions into four underlying regulatory areas: risk, premarket evidence, postmarket control, and autonomy or external model dependency.

- 1. What makes a GenAI-enabled medical device function risky?**
- 2. How should safety and effectiveness be demonstrated before deployment?**
- 3. How should safety be maintained through deployment and change?**
- 4. How should responsibility and control be assigned when intelligence is external or autonomous?**

Each question is considered through three perspectives:

Patient | Clinical / Institutional | Technological

The three perspectives reflect where risk is experienced and managed:

- Patient: understanding, reliance, consequence, escalation, and recourse.
- Clinical / Institutional: clinical judgment, standard of care, supervision, governance, and accountability.
- Technological: variability, reproducibility, provenance, monitoring, change control, and autonomy.

Together, these perspectives provide a consistent structure for examining the four regulatory questions.

The central challenge is to distinguish acceptable variability from safety-critical behavior that must remain controlled across the total product lifecycle.

1. What makes a GenAI-enabled medical device function risky?

Primary FDA discussion questions: 1-6

Patient	Risk rises when a patient cannot reasonably evaluate an output or when an incorrect output leads to harmful or delayed action. Patient-facing use should not by itself determine risk. The key factors are reliance, consequence, reversibility, and access to escalation or recourse. Effective escalation depends not only on patient-facing instructions, but also on reliable integration with a clinical workflow that preserves context and routes the patient to appropriate follow-up.
Clinical / Institutional	Clinical and institutional risk depends on the device's autonomy, the consequence of error, the deployment context, and the safeguards available. Assessment should consider the deployed model and product configuration, intended user, and clinical setting. Workflow integration can increase risk when it encourages automation bias, loses clinical context during handoffs, or fails to route outputs and escalations to the appropriate care pathway. Clinician oversight is meaningful only when the reviewer has the expertise, time, and authority to question or override the output. Retrospective review supports learning but does not protect the individual patient in real time. Before deployment, institutions should assess clinical and workflow fit and establish standard operating procedures (SOPs) that incorporate the technology into existing care processes. Training should be tailored to the intended user and setting, address the device's limitations and escalation pathways, and explicitly address automation bias and inappropriate overreliance.
Technological	Risk should be characterized by consequence, autonomy, directiveness, variability, observability, reversibility, and control. Ordinary generative variation may be acceptable. Safety-critical behaviors such as escalation, refusal, or diagnostic conclusions require tighter bounds and reproducible testing.

2. How should the safety and effectiveness of a GenAI-enabled medical device function be demonstrated before deployment?

Primary FDA discussion questions: 7-17

Patient	Premarket evaluation should reflect how patients will actually use and understand the device. Testing should account for realistic variation in literacy, phrasing, incomplete information, and repeated interaction, with emphasis on whether users can understand the output and its limitations. FDA specifically raises communication quality, user comprehension, subgroup performance, and clinically representative use. Testing should also assess whether the user interface helps patients understand the device’s output, limitations, and appropriate next steps, including when to seek further care.
Clinical / Institutional	Premarket evidence should combine structured benchmarking with clinical confirmation appropriate to the device’s intended use and risk profile. Benchmarking can demonstrate defined capabilities and known failure modes, but it cannot fully predict performance in clinical practice. When feasible, sponsors should confirm performance in real-world or clinically representative settings that reflect the intended users and workflows. Shadow deployment can be a useful pre-deployment stress test: the device processes real clinical inputs in a live workflow, but its outputs do not affect patient care. When this is not feasible, an alternative clinically representative evaluation should be used.
Technological	Safety and effectiveness should be demonstrated on the deployed device configuration, not the foundation model alone. Evaluation should be prespecified, reproducible, risk-proportionate, and broad enough to test safety behavior, clinical task performance, robustness, and generalizability before clinical confirmation. FDA’s proposed framework explicitly combines benchmarking with clinical confirmation and expects methods, scoring, and acceptance criteria to be defined in advance.

3. How should the safety of a GenAI-enabled medical device function be maintained through deployment and change?

Primary FDA discussion questions: 18-24

Patient	Postmarket safety should account for whether changes in the device alter what patients experience or how they may act on its outputs. Material changes in behavior, escalation, or reliability should be detectable before they create avoidable harm, and patients should retain clear paths to escalation or recourse when performance degrades.
Clinical / Institutional	The manufacturer remains responsible for device-level postmarket monitoring, investigation, corrective action, and re-evaluation of changes. Clinical and institutional safety should be supported through a risk-proportionate quality process that tracks clinically meaningful events, performance concerns, and workflow effects; identifies recurring or safety-relevant signals; investigates and addresses them through established quality-system processes when warranted; and evaluates whether resulting actions are effective. Clinicians and healthcare institutions should share relevant real-world feedback when available and applicable through existing reporting and quality processes. Planned, evolving, or externally initiated changes should be assessed for their potential effect on clinical use, workflow, and the applicable standard of care, with the scope of re-evaluation proportionate to the change and the device’s risk profile.
Technological	Postmarket control should maintain a traceable baseline for the deployed configuration and require re-evaluation after changes that could affect safety or effectiveness. Monitoring should address drift, model or system updates, third-party dependencies, and other material modifications through versioning, re-benchmarking, auditability, and rollback where appropriate.

4. How should responsibility and control be assigned when a GenAI-enabled medical device function relies on external or autonomous intelligence?

Primary FDA discussion questions: 25-26

Patient	When a GenAI-enabled medical device function depends on external or autonomous intelligence, patients should still have a clear understanding of who is responsible for the device's actions and how concerns can be escalated. Increased autonomy should not reduce transparency, recourse, or the ability to identify accountable parties when harm occurs.
Clinical / Institutional	Responsibility for a GenAI-enabled medical device function should remain identifiable and aligned with the ability to control the relevant clinical and technical risks. The manufacturer should remain accountable for the safety and effectiveness of the deployed device function, including its use of external models, tools, and updates. Clinicians retain responsibility for clinical judgment and decisions within their professional role; use of an AI-enabled device does not transfer that responsibility to the device or its manufacturer. Healthcare institutions remain responsible for implementation of governance, workflow design, user training, and local oversight appropriate to the function's risk and degree of autonomy. As autonomy increases, the system should operate within a clearly defined clinical scope, with explicit authority limits, appropriate human-oversight points, and escalation pathways for situations that exceed its intended use or safety boundaries. Institutions should be able to identify the device's intended role in the workflow, the actions it may take, the circumstances requiring clinician review, and the process to disable or suspend the device locally if safety concerns arise. External-model dependencies should not create ambiguity about accountability or reduce the ability of clinicians, institutions, patients, or regulators to identify and escalate concerns.
Technological	Responsibility should remain tied to control of the deployed device function, including its models, tools, permissions, updates, and safety boundaries. External foundation-model dependencies require sufficient visibility into model versioning and change control. While autonomous systems should operate within defined permissions, human oversight points, auditable action histories, and safe failure states. FDA specifically raises Foundation Model MAFs and additional oversight for agentic systems with autonomous multi-step action and reduced opportunity for human review.

Concluding Position

ClaraWell views the regulation of GenAI-enabled medical devices as a lifecycle problem rather than a model-specific problem.

A device may use probabilistic or generative components while still operating within defined clinical, institutional and technical boundaries. The regulatory objective should therefore be to identify which behaviors may vary, which must remain reliably constrained, and what evidence is required to demonstrate that distinction.

Across the patient, clinical / institutional, and technological perspectives, four principles remain consistent:

1. Risk should reflect the consequence and controllability of the device function, not the use of GenAI alone.
2. Premarket evidence should be proportionate to the function's intended use, autonomy, and potential harm.
3. Safety-critical behavior should remain observable, testable, traceable, and subject to appropriate re-evaluation after material change.
4. Accountability should remain identifiable even when a device depends on third-party models or increasingly autonomous systems.

The central question is therefore not whether GenAI-enabled medical devices can vary. It is whether that variability can be understood, bounded, evaluated, and managed throughout the total product lifecycle.

ClaraWell believes a regulatory framework built around those principles can preserve meaningful innovation while maintaining clear expectations for safety, evidence, and accountability.