An open interoperability standard for GenAI medical devices

Comment on FDA-2026-N-7874

19 October 2026

IM Healthcare research project

Responses to Questions 16 and 20 through 26

Recommendation

FDA should encourage manufacturers, model suppliers, clinical organizations and independent evaluators to co-develop a voluntary, open interoperability standard for modular healthcare, initially focused on GenAI-enabled medical-device workflows. Its first specification should define how devices, models and medical software describe their capabilities, exchange clinically meaningful data, request authorized actions, confirm completion and communicate safety-relevant changes. A portable evidence record, action controls and an auditable record would be required parts of this common framework.

The goal is to make a manufacturer's capability and integration work reusable across competing implementations, with the conditions of each deployment evaluated explicitly. Manufacturers should help shape both the specification and its conformance tests. FDA's questions on independent assessment, deployed-device evaluation, agentic actions and change control provide a practical basis for this effort. [1] We propose an initial draft and a bounded technical exercise to establish whether this approach reduces integration work while exposing clinically significant incompatibilities.

A failure between functioning components

Consider a hypothetical home-monitoring workflow. A sensor produces an accurate measurement. After an update, an integration service treats its upload time as its measurement time. A model interprets an old reading as current and requests clinical review. The messaging system acknowledges that the message was queued; the agent records that a clinician was notified. No clinician has acknowledged responsibility for the case.

A useful interoperability standard must make these differences explicit and testable. Correct interfaces alone cannot establish that the complete workflow reached its required clinical state.

1 Define a common core and profiles for specific uses

Questions 16, 25 and 26

The specification should distinguish a common core from profiles for particular uses. The core would identify the supplier and product capability, supported operations, interface and profile versions, inputs and outputs, operating limits, relevant dependencies and responsible contacts. Each use profile would identify the intended workflow and population, mandatory requirements and applicable tests. Implementations should build on suitable existing data-exchange and device-interface standards, documenting how their meanings and limitations are preserved. FDA's interoperability guidance provides an established basis for describing interface purposes and boundaries. [2]

A profile should address clinically material data semantics, including patient matching, units, source, measurement time separately from transmission time, freshness, quality and missing or corrected values. Invocation rules should define permitted actions, required human review, responses, timeouts, cancellation, retry and duplicate prevention. A declared capability would also

Website copy revised 30 October 2026 - Human Care. Submission has not been confirmed.

carry scoped evidence and change information. This gives manufacturers a common description that multiple integrators can implement and evaluate.

Conformance must refer to a named profile and version, the tested configuration, methods, conditions and limitations. Interface and behavioral conformance, evidence applicability and evaluation of the deployed clinical workflow need distinct conclusions. A conformance record would not establish the safety or effectiveness of every combination or confer FDA authorization.

2 Test whether evidence can be reused

Questions 16 and 25

Within each use profile, a portable evidence record would link each bounded product claim to its intended use, tested configuration, input requirements, supporting tests, limitations, dependencies and evidence owner. It should identify the method, comparator and evaluation setting, when applicability was last verified, justified review or expiry conditions, and how receiving parties learn that a claim is no longer supported. Scores from unrelated benchmarks should not be treated as interchangeable.

The receiving integrator should record which assumptions remain satisfied, what differs and what additional evidence is needed. The decision should state that identified evidence is applicable within its scope, that specified differences require additional testing, or that applicability remains unresolved. Device-specific safety and effectiveness remain to be established for the deployed workflow.

The correctness of these reuse decisions should itself be tested. An apparently compatible integration that violates a clinically important assumption should be included deliberately. Independent reviewers using prespecified adjudication criteria should assess both inappropriate acceptance of evidence and unnecessary rejection of applicable evidence. Held-out configurations and cases should test performance beyond the examples used to develop the method.

Public capability descriptions and evidence records can identify interfaces and the scope of testing while detailed reports remain under appropriate access controls. Underlying model evidence could use the voluntary Foundation Model Device Master File approach discussed by FDA, with the file holder's authorization. [1] Public records should not require patient records, model weights or proprietary implementation details.

3 Keep evidence aligned with changes

Questions 21, 22, 23 and 24

For each safety-relevant change, the parties should record the changed dependency, affected capabilities and workflows, evidence whose applicability is uncertain, temporary controls, required retesting and responsibility for restoration. Dependencies include relevant prompts, retrieval sources, tools, permissions and orchestration logic as well as model and device versions. Unknown versions and missing supplier information should have defined handling rules.

Suppliers should provide agreed version information, change notices and supporting evidence. Integrators should maintain dependency maps and assess interactions. The device sponsor should retain applicable regulatory responsibilities and release decisions; deploying organizations should manage the local conditions and clinical response processes they control. These assignments should close gaps in monitoring without transferring away manufacturer accountability.

Affected reuse decisions should become unresolved until continuing applicability or revalidation is documented. Evidence for unaffected claims may remain usable where justified; unexpected changes may require broader reassessment. A stable commercial model name or unchanged API is insufficient assurance. Detection, notification and containment times should be justified against the time to clinically significant harm. Hazard-appropriate controls and tested fallbacks should

preserve necessary care, including evaluated standalone device functions where appropriate. Changes in operational evidence status would not cancel regulatory authorization. FDA could provide examples distinguishing changes managed within an authorized PCCP, other applicable controls and changes requiring additional FDA review. [3]

4 Verify clinical handoffs and agent authority

Questions 20 and 26

For consequential actions, the standard should distinguish what was proposed, authorized, attempted, acknowledged and actually completed, including refusal, cancellation and an unknown outcome. Completion criteria must match the clinical workflow. The record should identify the assigned recipient, required human review, escalation deadline and observed response, with unresolved handoffs remaining visible.

An action record should link configuration and evidence scope to the necessary patient context, permission and observable result. Evaluation should include denied actions, retries, omissions, failed escalation and safe deferrals. Prespecified denominators should match the endpoint, such as patient episodes, eligible action opportunities or execution attempts, accounting for repeated and clustered events. Clinically significant boundary failures should be reported separately from average task success. Records should minimize sensitive data, remain exportable and permit authorized investigation without requiring hidden model reasoning.

Permissions should be enforced by separately testable controls outside the generative model's instructions. Tests should cover stale data, patient or unit mismatches, duplicate execution, revoked authority, unapproved delegation, prompt injection and loss of supervision. False blocks and delayed care also require measurement.

Controlled research on autonomous replication in simplified IoT-like environments provides an additional reason to test unauthorized persistence, software installation and expansion of access. It does not demonstrate such compromise of deployed medical devices. [4] Tests should establish whether spawning or delegating to another agent bypasses authorization or revocation in the evaluated configuration, and whether containment preserves clinically necessary functions. These tests would extend existing cybersecurity practices to the relevant workflow. [5]

5 Keep participation open and assessment independent

Questions 16 and 21

The specification should have published participation and change procedures, with manufacturers, clinicians, patient representatives and independent evaluators involved in its development. Multiple implementations and qualified assessors should be able to use it. Methods, assessor qualifications, conflicts of interest, fees and appeals should be transparent; protected challenge cases can discourage optimization solely for a public test.

Assessment decisions should be protected from sales incentives where an organization also supplies integration or distribution. Manufacturers should retain their intellectual property, customer relationships, exportable evidence and non-exclusive channels. A profile should remain usable without membership in a particular marketplace. Monitoring duties and required resources should be explicit, including the burden on small manufacturers and resource-constrained hospitals.

6 Develop and test the first specification together

We propose a 90-day voluntary technical exercise beginning after participants agree its scope. Its target cohort would include three unrelated device or medical-software manufacturers, two model suppliers, two independently implemented receiving systems, clinical representatives and an independent evaluator. These are participants to recruit, not announced partnerships. The

exercise would develop an initial common core and one use profile, with a prototype test harness. It would not complete a universal healthcare standard.

Each manufacturer would contribute one bounded capability, an available interface or simulator and its integration requirements. Participants would co-develop the draft specification and test the capability in both receiving systems, using existing interface adapters where needed. Deliverables would include the draft, conformance tests, scoped conformance records, evidence-applicability decisions, unsupported combinations and tested change and recovery procedures. Withheld cases should include deliberate semantic mismatches and failed handoffs.

Measures should include engineering hours, elapsed integration time, adapter and evidence reuse, duplicated evidence requests, acceptance of incompatible configurations or inapplicable evidence, unnecessary rejection or retesting, missed handoffs, false blocks, containment time and completeness of incident reconstruction. Acceptance criteria should be specified before results are reviewed. Comparisons with usual integration should use matched tasks and equivalent evidence access, reporting learning effects, scope differences and safety-relevant failures alongside any gains. Synthetic tests could establish technical feasibility; clinical benefit or real-world safety would require subsequent appropriate evaluation.

FDA could convene or observe the exercise and use its findings to develop examples for Questions 16 and 20-26. Stakeholders could begin voluntarily; participation would not confer authorization, preferred review or a waiver of applicable requirements.

An invitation to manufacturers

We invite manufacturers to become technical co-authors of the first draft and its conformance tests. One product capability would be evaluated in two independently implemented integration settings. Participants would help define the shared connection rules and receive documented findings on integration effort, reusable work and evidence still needed by a clinical buyer. This is a practical opportunity to shape the specification around real product constraints while retaining control of the product and freedom to work with competing platforms.

These recommendations define the proposed interoperability foundation of Human Care, an independent private-sector concept developed by our team for connected healthcare with measurable outcomes, which we are prepared to present for stakeholder discussion in November 2026. We welcome manufacturers and clinical organizations interested in shaping its initial specification and pilot. Human Care is a working title for our independent proposal and has no government affiliation or endorsement.

Partnership enquiries: timemap88@gmail.com

References

[1] FDA. Considerations for the Regulation of Generative AI-Enabled Medical Devices. Discussion paper, August 2026. Sections V-VII and Appendix A. https://www.fda.gov/media/194242/download

[2] FDA. Design Considerations and Pre-market Submission Recommendations for Interoperable Medical Devices. September 2017. https://www.fda.gov/media/95636/download

[3] FDA. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions. August 2025. https://www.fda.gov/media/166704/download

[4] Anthropic. AI models may be capable of self-replicating without access to a datacenter. October 12, 2026. https://www.anthropic.com/research/self-replicating-ai

[5] FDA. Cybersecurity in Medical Devices: Quality Management System Considerations and Content of Premarket Submissions. February 2026. https://www.fda.gov/media/119933/download