CAHIR SOLUTIONS | INDUSTRY INSIGHTS

The Human-in-the-Loop Illusion:

Why Checking the Box on AI Safety Is Not Enough

A Roadmap for Validating Cognitive Engagement in AI/ML Medical Devices

April 2026 | CAHIR Solutions

Key Insight

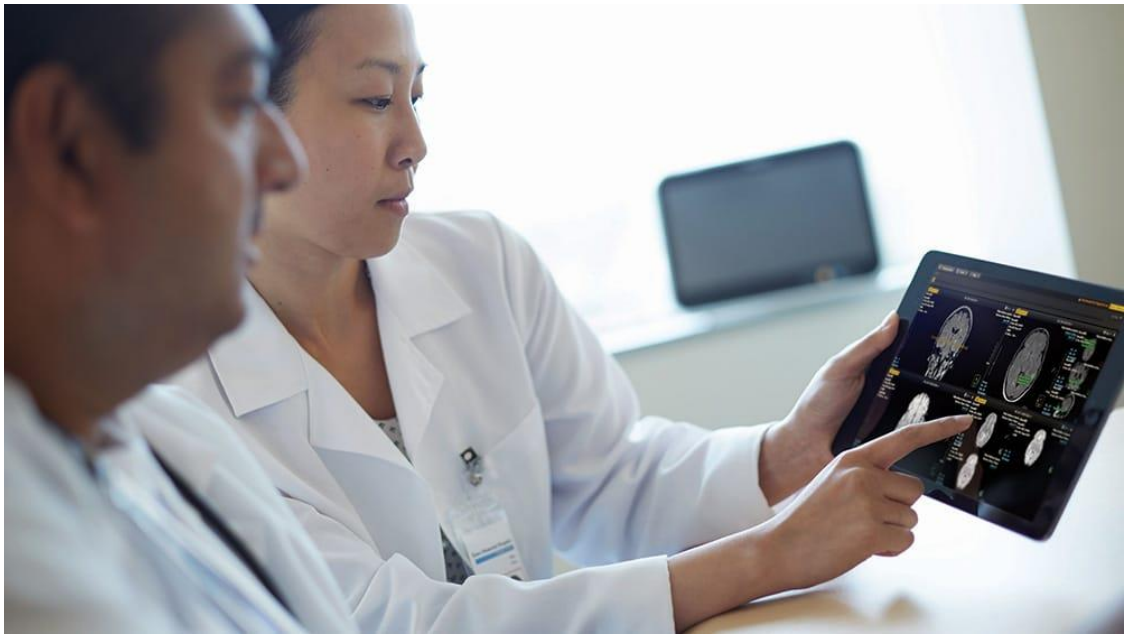
The FDA has cleared over 1,000 AI/ML-enabled medical devices, yet the industry still lacks a validated framework for proving that the human in the loop is cognitively engaged—not just technically present.

For the vast majority of AI/ML-enabled medical devices not approved for fully autonomous use, "Human-in-the-Loop" (HITL) serves as the default safety control. The underlying assumption is straightforward: if a clinician can intervene, they will intervene effectively. But mounting evidence from cognitive science, clinical workflow research, and real-world deployment data reveals this assumption to be **fundamentally flawed.**

As the FDA and global regulators tighten expectations around human-AI collaboration, MedTech leaders face an urgent question: Are your HITL systems actually safe, or are they creating a dangerous illusion of oversight?

The Cognitive Vulnerabilities Hiding in Plain Sight

Traditional human factors evaluations for medical devices focus on physical interaction errors—can the user press the right button, read the screen, follow the workflow? But AI-enabled devices introduce an entirely different category of risk: **cognitive risk.**



Three cognitive phenomena pose the greatest threat to effective HITL oversight:

Automation Bias. Clinicians accept AI-generated recommendations without independent verification—particularly when systems are perceived as generally reliable. Research shows this leads to both errors of commission (following incorrect AI recommendations) and errors of omission (failing to act without AI guidance). Risk homeostasis theory suggests that the perceived accuracy of AI may paradoxically make clinicians less careful in their independent evaluation.

Alert Fatigue and Cognitive Overload. In high-volume clinical environments, AI systems can generate hundreds of alerts per shift. Studies in intensive care and remote patient monitoring show that threshold-based alert systems overwhelm clinical teams, making it nearly impossible to distinguish critical signals from noise. Cleveland Clinic's 2025 deployment of an AI sepsis detection system demonstrated the stakes: by reducing false alerts by 90% and improving true-positive sensitivity by 46%, the system showed that smarter alerting—not more alerting—is the path forward.

The "Check-the-Box" Phenomenon. Perhaps the most insidious vulnerability: clinicians rubber-stamp AI outputs due to cognitive fatigue, time pressure, or institutional incentives. The human is technically "in the loop" but is no longer performing the independent cognitive evaluation that the safety framework assumes. In diagnostic radiology and pathology, where AI tools are most prevalent, workload pressures create conditions where meaningful oversight degrades to perfunctory acknowledgment.

Regulators Are Closing the Gap

The FDA and global regulators are rapidly evolving their expectations around HITL validation. The January 2025 draft guidance on AI-Enabled Device Software Functions introduced a paradigm shift: from traditional usability validation to **cognitive interaction validation.**

How the FDA's Human-AI Team Model Reframes Validation

Focus Area	Traditional Devices	AI-Enabled Devices
Usability	Task execution accuracy	Interpretation and cognitive understanding
Output Type	Static	Variable / probabilistic
Risk Consideration	Physical interaction errors	Cognitive misinterpretation and automation bias
Validation	Task success	Human-AI team performance validation

Key Regulatory Developments in 2025–2026

- **FDA Human-AI Team Model:** Manufacturers must now demonstrate that clinicians correctly interpret AI outputs, understand model limitations, recognize uncertainty, and apply outputs appropriately within clinical workflows.
- **Predetermined Change Control Plans (PCCP):** The August 2025 final guidance establishes a framework for iterative AI improvement, including validation of continuous human-AI engagement as algorithms evolve post-deployment.
- **EU AI Act (August 2026):** AI-enabled SaMD classified under MDR will automatically qualify as high-risk AI systems, triggering new obligations around data governance, transparency, human oversight, and record-keeping.
- **AHA Recommendations (December 2025):** The American Hospital Association urged the FDA to update adverse event reporting mechanisms to capture AI-specific risks, including model drift, bias, and hallucination—recognizing that existing frameworks are inadequate for AI device oversight.
- **QMSR Harmonization (February 2026):** The FDA's Quality Management System Regulation now aligns U.S. requirements with ISO 13485 for the first time, integrating human factors engineering as an AI lifecycle discipline rather than a standalone usability activity.

Regulatory Reality

The legal system still holds humans—not algorithms—accountable for medical decisions. A clinician who blindly follows an AI recommendation may be found negligent, even when the AI itself was cleared by the FDA. This creates a dual imperative: the device must support engagement, and the clinician must demonstrate independent judgment.

Engineering SaMD for Continuous Human-AI Engagement

Proving that a human is cognitively engaged requires more than good user interface design. It demands purpose-built software infrastructure that actively monitors, measures, and supports the quality of human-AI interaction throughout the device lifecycle.

A Technical Roadmap for MDMs

1. **Implement Cognitive Engagement Metrics.** Move beyond click-through rates. Track interaction dwell time, override frequency and rationale, query behavior (did the clinician investigate before confirming?), and confidence-score comprehension through periodic validation checks.
2. **Design for Intelligent Alert Prioritization.** Replace threshold-based alerting with context-aware triage. AI systems should learn each patient's baseline, evaluate trends across multiple data streams, and surface only actionable signals—as demonstrated by next-generation RPM platforms that cut false positives by up to 90%.
3. **Build Transparency and Explainability Into the Interface.** Display confidence levels, model limitations, and reasoning pathways. The FDA now expects manufacturers to assess whether users correctly interpret AI outputs and understand when to override them. Model Cards are increasingly expected in 2026 submissions.
4. **Engineer Feedback Loops for Continuous Learning.** Capture structured clinician feedback (confirmations, overrides, corrections) and route it back into model retraining pipelines through FDA-recognized PCCP frameworks. This creates a virtuous cycle where clinical interaction data improves both AI performance and engagement quality.
5. **Conduct Team-Based Validation Testing.** Research from ICU simulation studies with 180 physicians and nurses shows that human-AI team dynamics differ fundamentally from human-to-human collaboration. Validation must include multi-user workflow scenarios that reflect real-world uncertainty—not just individual task success.

6. **Establish Post-Market Cognitive Surveillance.** Extend monitoring beyond device performance metrics to include longitudinal tracking of human engagement quality, automation dependency trends, and override accuracy rates. This aligns with the AHA's December 2025 call for AI-specific adverse event reporting.

What MedTech Leaders Should Be Doing Now

The convergence of FDA expectations, EU AI Act deadlines, and clinical reality demands immediate action from Medical Device Manufacturers:

- **Audit your current HITL assumptions.** Identify where "human can intervene" has been treated as synonymous with "human will intervene effectively."
- **Invest in cognitive engagement infrastructure.** Budget for software capabilities that go beyond UI design to actively measure and support clinician cognitive engagement.
- **Prepare for dual regulatory compliance.** With the EU AI Act's high-risk AI obligations taking effect August 2026, manufacturers operating globally need unified compliance strategies that address both FDA TPLC requirements and EU human oversight obligations.
- **Bridge the clinical-engineering divide.** Bring cognitive scientists and clinical workflow experts into the SaMD design process alongside software engineers. Human-AI team performance cannot be optimized by either discipline alone.
- **Advocate for industry standards.** Engage with FDA, standards bodies, and professional organizations to shape the emerging frameworks for HITL validation before they are finalized.

The Bottom Line

"Human-in-the-loop" must evolve from a regulatory checkbox to a validated, continuously monitored safety system. The MedTech organizations that invest in proving cognitive engagement—not just

technical presence—will be the ones that earn regulatory confidence, reduce clinical liability, and ultimately deliver safer AI-augmented care.

About CAHIR Solutions

CAHIR Solutions provides strategic advisory services at the intersection of digital health innovation, regulatory compliance, and organizational governance. We help MedTech organizations navigate the complex landscape of AI-enabled medical device commercialization—from FDA strategy and clinical validation to post-market surveillance and health equity outcomes.

For more insights on MedTech regulatory strategy and digital health innovation, visit cahir.ai

Sources

1. FDA, "Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations" (Draft Guidance, January 2025) – https://www.fda.gov/regulatory-information/search-fda-guidance-documents
2. Maven Regulatory Solutions, "FDA AI Device Human Factors 2026 Guide" (February 2026) – https://www.mavenrs.com/blog/fda-ai-enabled-medical-device-human-factors-2026
3. American Hospital Association, "AHA Letter to FDA on AI-enabled Medical Devices" (December 2025) – https://www.aha.org/lettercomment/2025-12-01-aha-letter-fda-ai-enabled-medical-devices
4. RegDesk, "Navigating Global SaMD Regulations: FDA, EU MDR, TGA, PMDA" (April 2026) – https://www.regdesk.co/blog/navigating-global-regulations-for-software-as-a-medical-device-samd/
5. Censinet, "Clinical AI Bias Testing: How to Assess and Mitigate Algorithmic Risks in Healthcare" – https://censinet.com/perspectives/clinical-ai-bias-testing-how-to-assess-and-mitigate-algorithmic-risks-in-healthcare
6. MedPro Group, "Artificial Intelligence Risks: Automation Bias" – https://resource.medpro.com/artificial-intelligence-risks-automation-bias
7. Frontiers in Psychology, "Human-AI teaming: leveraging transactive memory and speaking up for patient safety" (2023) – https://pmc.ncbi.nlm.nih.gov/articles/PMC10436524/
8. npj Artificial Intelligence, "Human-AI teaming in healthcare: 1 + 1 > 2?" (December 2025) – https://www.nature.com/articles/s44387-025-00052-4
9. Prevounce Health, "From Alert Fatigue to Smart Triage: AI-Driven Escalation Workflows" (September 2025) – https://blog.prevounce.com/ai-powered-rpm-smart-triage
10. IntuitionLabs, "FDA Digital Health Guidance: 2026 Requirements Overview" (February 2026) – https://intuitionlabs.ai/articles/fda-digital-health-technology-guidance-requirements