PUBLICATION

Responsible AI in Health Care: What Providers and AI Vendors Must Do Now

Authors: Julie A. Kilgore
November 13, 2025

Despite the rapid invention and widespread use of artificial intelligence (AI), the federal government is just beginning to shape the regulatory landscape. In the absence of comprehensive federal AI regulations, private health care organizations are creating voluntary "responsible AI" frameworks. While these frameworks are not based on specific statutory authority, they are reshaping industry expectations and establishing baseline measures for transparency. These early frameworks are likely to be used as the template for future regulations, making early adoption a competitive advantage. Early operational adoption can also reduce procurement friction, speed implementation, and mitigate safety and compliance risk. This alert reviews health AI frameworks created for developers and deployers.

Regulatory Activity Snapshot

July 23, 2025: The White House's Office of Science and Technology Policy issued America's AI Action Plan. While the plan did not specifically address AI uses in health care, it outlines the Trump administration's goals for a federal regulatory scheme, emphasizing the need to establish a common-sense regulatory environment to promote innovation with respect to AI applications.

September 30, 2025: The U.S. Food & Drug Administration (FDA) issued a Request for Public Comment on methods for assessing the real-world performance of AI-enabled medical devices. The request asks for input on performance metrics, post-deployment monitoring methods, data management practices, signs of performance degradation, clinical usage patterns, and implementation barriers, including approaches to maintaining patient privacy and data protections. The deadline for comments is December 1, 2025.

November 6, 2025: FDA's Digital Health Advisory Committee met to discuss generative AI-enabled digital mental health medical devices. Acknowledging generative AI can improve access to mental health treatment in cases where traditional barriers would have prevented care, the Committee also found that it introduces novel risks because models evolve over time. Committee members discussed the need for structured taxonomy for mental health uses cases, post-market monitoring, and clearer regulatory pathways for developers of AI-enabled mental health therapeutic devices.

Emerging Frameworks

Category 1: Developed Primarily for AI Developers

The Consumer Technology Association (CTA) – Performance Verification and Validation for Predictive Health AI Solutions

- **Type of Framework:** Technical standard for predictive health AI solutions.
- **Intent:** Provide a consistent way for developers to address safety and effectiveness of predictive AI solutions before deployment and ensure solutions meet prescribed purposes and user expectations in real-world settings, which helps foster trust among clinicians, patients, and policymakers.
- **Broader Challenge Targeted:** The lack of standardized, risk-based methods to assess predictive health AI before deployment. Without such measures, inconsistencies in quality and reliability can perpetuate health disparities and erode trust among clinicians and patients.

- **Data Management Approach:** Supports rigorous data quality and transparency requirements in model development, noting the importance of complete data verification, mitigation of potential biases in different data processes, and transparency of input and output data elements. References the Health Insurance Portability and Accountability Act (HIPAA), the European Union's General Data Protection Regulation (GDPR), and frameworks such as the National Institute of Standards and Technology (NIST) AI Risk Management Framework (AI RMF) to encourage privacy and security throughout validation.

Stanford University Human-Centered Artificial Intelligence (HAI) - MedAgentBench

- **Type of Framework**: Academic testing benchmark / research initiative.
- **Intent**: Help developers improve model performance and identify error patterns before real-world clinical deployment. MedAgentBench offers a reproducible benchmark system for evaluating how well large language models (LLMs) can function as AI agents performing real-world physician tasks. The framework tests whether AI agents can handle operational clinical workflows through 300 physician-developed scenarios utilizing Fast Healthcare Interoperability Resources (FHIR) application programming interface (API) endpoints to navigate electronic health record (EHR) systems.
- **Broader Challenge Targeted:** The gap in current testing focused on AI agents' medical knowledge and complexities of real-world, interactive clinician tasks. While advanced LLMs have demonstrated strong performance on medical licensing exams and clinical knowledge tests, no dataset exists to measure their ability to function as autonomous agents navigating actual EHR environments.
- **Data Management Approach**: Uses a virtual EHR environment containing profiles of more than 100 patients pulled from deidentified records in Stanford's STARR clinical data warehouse.

Category 2: Developed Primarily for AI Developers and Deployers

Paragon Health Institute - Targeted Post-market Surveillance: The Way Toward Responsible AI Innovation in Health Care

- **Type of Framework**: Policy proposal / concept paper.
- **Intent**: Create a scalable, cost-efficient, post-market surveillance framework for AI medical devices that balances innovation with patient safety. The framework proposes a risk-based system for AI that is unpredictable (i.e., relies on adaptive algorithms, open training datasets, or architectures that can produce variable outputs for the same inputs) and presents medium-to-high risks to patients. The framework includes periodic manufacturer-led revalidation using existing test data and performance monitoring through aggregated outcome data registries.
- **Broader Challenge Targeted:** The gap between the FDA's premarket validation, designed for devices with consistent outputs, and AI systems whose performance variability may emerge after deployment. According to Paragon Health Institute, a substantial regulatory gray area exists for AI-enabled software with unclear oversight of quality and performance for technologies such as internally developed health system tools, EHR vendor algorithms, and clinical-operational hybrid technologies.
- **Data Management Approach**: Proposes an aggregated outcome data registry where participating health systems contribute anonymized summary data extracted from local deployment environments. The registry would enable manufacturers, regulators, and participating providers to detect adverse events or performance trends without exposing patient-level information or proprietary code. The authors note that, because meaningful AI performance depends on local context, a central government agency cannot unilaterally create an effective post-marketing monitoring system.

URAC – Health Care AI Accreditation Program

- **Type of Framework**: Formal third-party accreditation program.
- **Intent**: Provide third-party validation for responsible AI innovation and deployment.
  - The developer track evaluates regulatory compliance, contracting practices, data governance, risk analyses, AI system training, and transparency.
  - The deployer track assesses safe implementation, clinical oversight, workforce training, responsible use, and impact disclosure.
- **Broader Challenge Targeted:** The lack of standardized, independent validation mechanisms for AI systems in health care. Without third-party validation, health care organizations face overlapping or inconsistent measures for governance, transparency, risk management, and performance monitoring.
- **Data Management Approach**: Requires accredited entities to document and maintain procedures for AI data governance, validation, and performance monitoring, as well as disclosures about system use and impact. The program is evidence-based and auditable, relying on documentation, records, and internal analyses.

Joint Commission / Coalition for Health AI (CHAI) - The Responsible Use of AI in Healthcare (RUAH)

- **Type of Framework**: High-level implementation guidance.
- **Intent**: Promote a shared understanding of responsible AI deployment and use across health care organizations. The guidance defines seven operational elements for managing AI tools throughout their lifecycle: (1) AI policies and governance, (2) patient privacy and transparency, (3) data security and data-use protections, (4) ongoing quality monitoring, (5) voluntary, blinded safety event reporting, (6) risk and bias assessment, and (7) education and training.
- **Broader Challenge Targeted:** Health care organizations face increasing pressure to adopt AI tools without uniform safety standards or adequate implementation resources, including education and training protocols.
- **Data Management Approach**: Directs organizations to define obligations to protect data and establish shared requirements within data use agreements to limit permissible uses of exported data, including third-party vendor compliance. For AI processing protected health information (PHI), the guidance encourages organizations to execute Business Associate Agreements (BAAs), apply the "minimum necessary" standard, and maintain appropriate guardrails under HIPAA's Privacy, Security, and Breach Notification Rules. Organizations should also monitor re-identification risks because AI tool development can involve de-identified data in training, testing, or fine-tuning processes.

The Digital Medicine Society (DiMe)- The Playbook - Implementing AI in Health Care

- **Type of Framework:** Implementation framework / educational toolkit.
- **Intent**: Help health care organizations scale innovations by using a roadmap to align AI selection with organization needs, capabilities, and implementation readiness. The playbook guides organizations through three phases: (1) problem identification and readiness assessment, (2) tool selection, and (3) implementation.
- **Broader Challenge Targeted:** The high failure rate of AI initiatives in health care, which is often related to organizational, operational, or workflow barriers rather than limitations of technology. Organizations struggle to align AI selection with real-world needs and operational readiness, which limits the value of AI for clinicians and patients.
- **Data Management Approach**: Emphasizes foundational IT and data science responsibilities to ensure data quality before validation or deployment. The playbook encourages organizations to establish robust data infrastructure and proactively identify data gaps or quality issues that could undermine model performance. Generally, data sharing occurs within organizations or established

partnerships for validation, monitoring, supported by privacy protections, deidentification, and documentation for auditability.

Call to Action / Practical Implications

The health care AI ecosystem is moving from general principles to operational frameworks. Industry organizations are developing structured approaches to AI evaluation, validation, deployment, and continuous monitoring. Health care organizations are increasingly incorporating these voluntary frameworks to support baseline expectations for transparency and risk management protocols in the absence of comprehensive federal mandates. The following are steps both health care provider organizations (HCPs) and AI vendors should consider taking now to align with industry standards, gain a competitive advantage, and better prepare for future regulations:

- For HCPs:
  - Establish an AI governance operating plan of action:
    - Approve an AI use case register and designate an executive sponsor and clinical safety lead.
    - Stand up a change-control process for any AI model update, with pre- and post-performance documentation.
  - Update procurement checklists to align with recognized frameworks:
    - Require model cards, performance by subpopulation, bias mitigations, data flow diagrams, secondary data use terms, and monitoring telemetry disclosures.
    - Include these in RFPs and contract templates, including BAAs and data use and processing agreements, and confidentiality provisions.
  - Implement real-world performance monitoring:
    - Define KPIs per AI use case, thresholds, and alert routing; log outcomes to support audits and adverse event detection.
  - Train clinical and operations staff:
    - Provide role-specific training on intended AI use, limitations, override paths, and incident reporting.
- For AI vendors:
  - Package transparency materials:
    - Deliver a model card; validation summaries with subpopulation performance; known limitations; and bias mitigation steps.
    - Provide a data stewardship appendix: PHI handling, de-identification, retention, secondary data use, and security controls aligned with HIPAA and state privacy laws.
  - Build for monitoring and auditability:
    - Expose versioning, telemetry, and performance dashboards that HCPs can consume; document drift detection and retraining triggers.
  - Align with recognized frameworks:
    - Map your practices to URAC criteria and CTA validation elements; consider scenario-based testing akin to MedAgentBench to demonstrate workflow fitness.
  - Contracting readiness:
    - Prepare standard contract clauses covering change notifications, customer approval for material model updates, and incident reporting; be BAA-ready where PHI is processed.

For more information or assistance on this topic, please contact Julie A. Kilgore or another member of Baker Donelson's Data Protection, Privacy and Cybersecurity Team.

Paige Kobza, Director, Health Policy at Maverick Health Policy and Ashley Progebin, Senior Policy Analyst at Maverick Health Policy, contributed to this article.