Journal Pre-proof

Physician-Centered Monitoring of AI-Enabled Medical Devices

George CM SiontisMD , Yijun RenMD , Enrico G CaianiMD , Alan G FraserMD

PII: S0002-9343(26)00167-1
DOI: https://doi.org/10.1016/j.amjmed.2026.02.024
Reference: AJM 18395

To appear in: *The American Journal of Medicine*

Received date: 22 February 2026
Accepted date: 23 February 2026

Please cite this article as: George CM SiontisMD , Yijun RenMD , Enrico G CaianiMD , Alan G FraserMD , Physician-Centered Monitoring of AI-Enabled Medical Devices, *The American Journal of Medicine* (2026), doi: https://doi.org/10.1016/j.amjmed.2026.02.024

Physician-Centered Monitoring of AI-Enabled Medical Devices

George CM Siontis MD[1,*], Yijun Ren MD[2], Enrico G Caiani MD[2,3], Alan G Fraser MD[4]

[1]Department of Cardiology, Bern University Hospital, Inselspital, University of Bern, Bern, Switzerland.

[2]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy.

[3]IRCCS Istituto Auxologico Italiano, Ospedale S. Luca, Milan, Italy

[4]Department of Cardiology, University Hospital of Wales, Heath Park Way, Cardiff CF14 4XW, the United Kingdom

*Correspondance to: George CM Siontis, MD, PhD, Department of Cardiology, Bern University Hospital, University of Bern, Freiburgstrasse 18, CH-3010 Bern, Switzerland, Tel: +41 031 632 21 11, Email: georgios.siontis@insel.ch

Funding sources: No funding was obtained.

Conflicts of interest: The authors declare no relevant conflicts of interest related to this manuscript.

Contributions: All authors had access to the data and a role in writing the manuscript.

Article type: Commentary

Key words: artificial intelligence; monitoring; physicians; AI-enabled medical devices.

Running head: Physician Oversight of AI Medical Devices

Efforts to accelerate the development and adoption of a new technology should not compromise its reliability or reproducibility, or undermine the need for continuous evaluation of its performance. These principles apply as much as ever to the rapidly evolving field of artificial intelligence (AI) in medicine. Some AI algorithms appear to have been approved for clinical use based on limited clinical evidence[1,2], making robust post-market monitoring with systematic feedback from end-users essential. Continuous monitoring and iterative refinement are indispensable for optimizing the safety, validity, and clinical utility of most AI systems, throughout their lifecycle. Physicians will not be replaced as the primary end-users with responsibility for the care of their patients, so they must play a central role in monitoring the performance of AI-enabled medical devices.

Importance of post-market monitoring of AI-enabled medical devices

In real-world clinical settings, the performance of AI-enabled medical systems may diverge substantially from that observed in controlled environments or clinical trials.[1] After deployment, these systems are susceptible to degradation of accuracy, due to temporal changes in patients' characteristics, clinical workflows, distribution of data, or technical infrastructures.[3] This can lead to unreliable clinical decisions and potential non-compliance with the regulatory requirements based on which the AI-enabled medical device was initially approved for clinical use. Algorithmic bias may be critical when software is applied to patients with features or geographic backgrounds that were under-represented in the original training data, resulting in models with suboptimal performance and pronounced risk of exacerbating disparities in care.[2]

To mitigate these risks, periodic updating or retraining of a model is essential to ensure that its internal parameters remain calibrated to evolving real-world conditions and applicable to each specific clinical environment in which it is deployed.[4] Post-market monitoring helps to identify issues affecting usability, track adverse events due to erroneous outputs, and support timely interventions (such as system adaptation or recall) that can prevent continuing problems.

Key components of effective AI monitoring

Effective monitoring of AI-enabled medical systems requires a comprehensive, multi-layered framework that keeps physicians at its centre (**Figure**). Core components include continuous assessment of model performance, supported by rigorous supervision of input and validation of output, to detect data drift, anomalous patterns, or unexpected behaviors. Because validity of model outputs depends heavily on completeness, consistency, and reliability of the input data streams, monitoring of data quality is essential to prevent performance degradation and bias. In parallel, robust oversight of infrastructure and resource usage helps ensure that computational environments remain stable, accessible, and compliant with security requirements. Modern monitoring frameworks also provide real-time visibility and scalability, allowing institutions to identify issues rapidly while model usage grows, which supports the safe expansion of AI tools across different clinical contexts and deployment scales.

Beyond internal analytics, real-world evidence collected by techniques such as web scraping of publicly available information or feedback posted in social media, can help to identify external signals of system failures or users' concerns.[5] Crucially, clinicians must remain an integral part of this ecosystem, as their structured feedback to

developers enables continuous refinement of the models and the early identification of clinically relevant errors (**Figure**).[3] Complementing this, explainability, interpretability, and transparency tools support clinicians in validating outputs, while human-in-the-loop oversight ensures that medical expertise remains central to decision-making, thereby preserving safety and trust in clinical AI applications.

Implementation of effective continuous AI monitoring

Implementing an effective continuous, systematic, and data-driven AI monitoring framework requires a structured approach built around well-defined technical and clinical priorities.[3] It is essential to define the structured performance metrics, including model accuracy, calibration, and latency, and to combine these with performance monitoring that detects distributional drift and clinically relevant error patterns. Data quality is maintained through rigorous data integrity checks that assess the completeness and consistency of input data, and through data harmonization to ensure stable input over time. To ensure consistent oversight, monitoring systems must automate acquisition of data from clinical workflows and operational sources, thereby providing timely and unbiased information on the model's real-world performance.

Computational robustness is ensured through system performance tracking and resource optimization to maintain operational continuity. The use of interactive dashboards and automated anomaly detection tools supports real-time visibility, enabling both technical teams and clinical stakeholders to interpret trends, investigate anomalies, and intervene promptly when the performance of the model becomes suboptimal. Finally, explainability and transparency should be operationalized through

model explanation tools and transparent reporting, to support clinical understanding and regulatory accountability.

Critical role of physicians for monitoring

Physicians play a critical and irreplaceable role in the development, deployment, and post-market monitoring of AI-enabled medical devices. Human oversight should be preserved through continuous feedback loops, which can identify edge cases, usability barriers, and context-specific failures. While regulatory frameworks and technical performance metrics provide essential safeguards, the real-world effectiveness and safety of AI systems ultimately depend on the clinical judgment and continuous vigilance of the clinicians who use them, as clinical decision-making is inherently contextual.

Physicians are uniquely positioned to detect subtle performance issues, context-specific failures, or data-drift effects that may not be captured by automated monitoring alone. Their direct feedback on unexpected outputs, challenges in usability, and patient-specific discrepancies is critical for guiding iterative refinement of the models, mitigating emerging biases, and preventing potential harm. Establishing robust mechanisms for clinicians to provide feedback (integrated into routine workflow) is therefore essential to ensure that AI tools evolve in alignment with clinical needs, maintain high reliability across diverse populations, and continually improve their performance over time. In this sense, keeping physicians “in the loop” is not only a regulatory or operational necessity, but a cornerstone of safe, effective, and trustworthy application of AI in healthcare.

Conclusions

A multilayered framework for post-market monitoring of AI-enabled medical devices must include structured and accessible mechanisms for systematic feedback to physicians, as they are uniquely positioned to detect unexpected behaviors, subtle errors, and context-specific limitations that automated systems may fail to capture. Physicians can maintain the integrity and clinical trustworthiness of medical AI systems, ensuring that innovation does not outpace patient safety and scientific rigor.

Acknowledgments

The authors have no relevant disclosures related to the content of this manuscript.

References

1. Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. *Lancet Digit Health*. 2024;6(5):e367-e373. doi:10.1016/S2589-7500(24)00047-5

2. Windecker D, Baj G, Shiri I, et al. Generalizability of FDA-Approved AI-Enabled Medical Devices for Clinical Use. *JAMA Netw Open*. 2025;8(4):e258052. doi:10.1001/jamanetworkopen.2025.8052

3. Lekadir K, Frangi AF, Porras AR, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ*. 2025;388:e081554. doi:10.1136/bmj-2024-081554

4. Davis SE, Greevy RA, Lasko TA, Walsh CG, Matheny ME. Detection of Calibration Drift in Clinical Prediction Models to Inform Model Updating. *J Biomed Inform*. 2020;112:103611. doi:10.1016/j.jbi.2020.103611

5. Ren Y, Caiani EG. Leveraging natural language processing to aggregate field safety notices of medical devices across the EU. *NPJ Digit Med*. 2024;7(1):352. doi: 10.1038/s41746-024-01337-9.

Figure

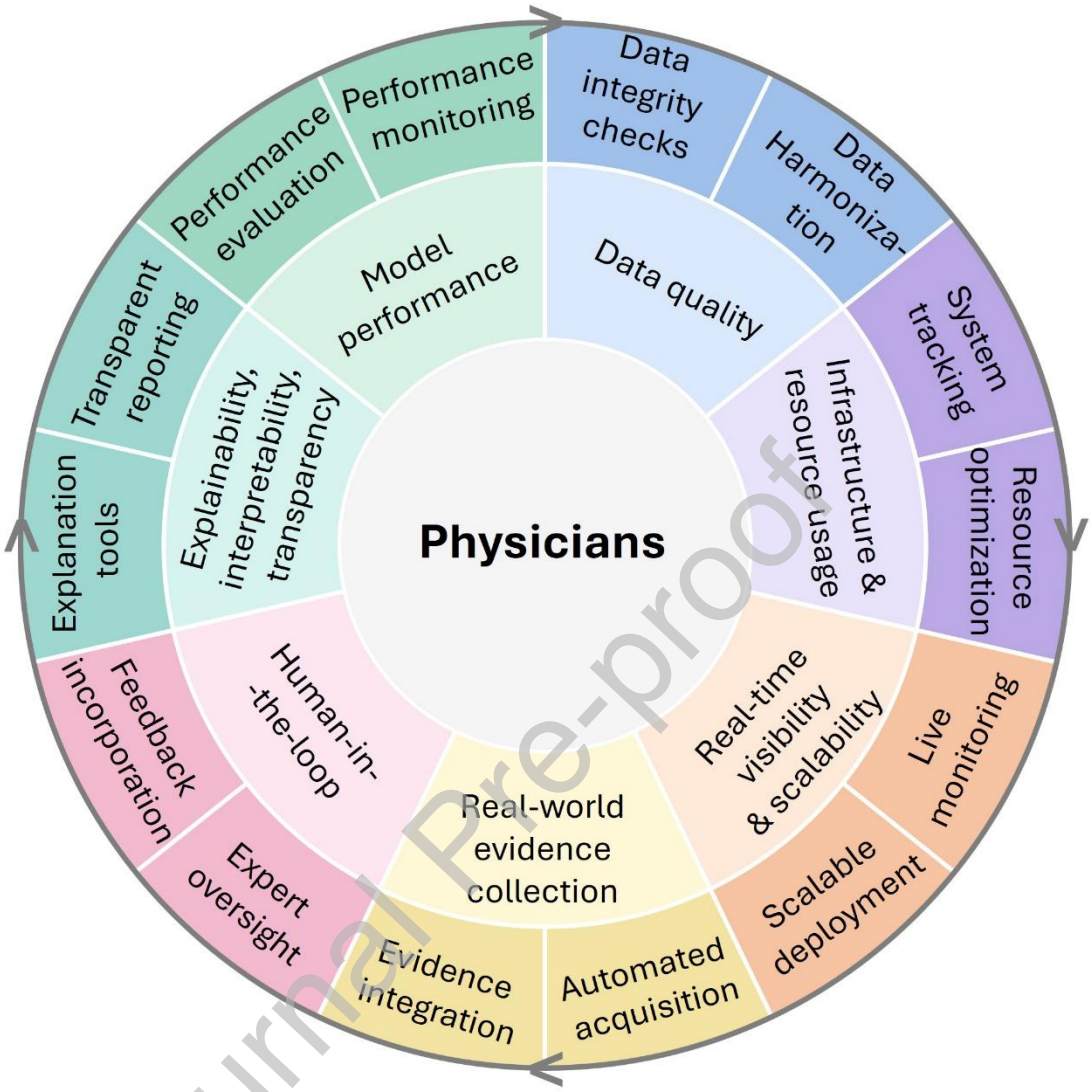

Figure: Factors to consider for monitoring of AI-enabled devices.

Overview of key considerations along the continuous monitoring pathway of artificial intelligence enabled medical devices following approval for clinical use. Physicians should retain a central role in the "living" process of continuously evaluating model performance and providing feedback.

Declaration of interest statement:

The authors declare no relevant conflicts of interest related to this manuscript.