Navigating the Black Box: Transparency Requirements for Medical AI

Rasit Dinc

Rasit Dinc Digital Health & AI Research

Published: December 7, 2022 | AI Diagnostics

DOI: 10.5281/zenodo.17997688

Abstract

The integration of Artificial Intelligence AI into healthcare is rapidly transforming diagnostics and treatment. As AI systems take on critical roles in clin...

The integration of Artificial Intelligence (AI) into healthcare is rapidly transforming diagnostics and treatment. As AI systems take on critical roles in clinical decision-making, the demand for **transparency**—the ability to understand how an AI arrives at its output—has become a non-negotiable requirement for regulators, clinicians, and the public. This article examines the emerging global regulatory landscape and the core requirements shaping the future of transparent medical AI.

The Regulatory Mandate for Explainability

Transparency in medical AI is moving from an ethical ideal to a formal regulatory mandate across major global jurisdictions. These frameworks are designed to ensure patient safety, build trust, and enable effective human oversight of AI-driven tools.

1. The U.S. FDA: Guiding Principles for MLMDs

The U.S. Food and Drug Administration (FDA), in collaboration with Health Canada and the UK’s MHRA, has established guiding principles for Machine Learning-Enabled Medical Devices (MLMDs) that emphasize transparency throughout the total product lifecycle [1]. The FDA defines transparency as the clear communication of appropriate information about an MLMD, including its intended use, development, performance, and underlying logic.

Key requirements for developers include: ***Intended Use and Performance:*** *Clear documentation of the device's medical purpose, target population, and integration into the clinical workflow.* **Data Characterization:** Disclosure of the training and testing data, including known gaps or biases, particularly concerning underrepresented patient populations. ***Logic and Explainability:*** *Communicating the basis for a device's output, which is essential for clinicians to critically assess the AI's recommendation before making a final decision.*

2. The European Union AI Act: A High-Risk Classification

The European Union's AI Act classifies AI systems used as medical devices as "high-risk," subjecting them to stringent transparency requirements [2]. Article 13 mandates that high-risk AI systems must be designed to ensure their operation is sufficiently transparent for deployers (e.g., hospitals, clinicians) to interpret the system's output and use it appropriately.

The instructions for use must contain comprehensive information, including: **Performance Metrics:** The expected level of accuracy, robustness, and cybersecurity, and any known circumstances that may impact these levels. ***Limitations and Risks:*** *Any known or foreseeable circumstances that may lead to risks to health and safety, including potential misuse.* **Human Oversight:** Measures put in place to facilitate the interpretation of the AI's outputs by human deployers.

3. The ONC's HTI-1 Final Rule: Transparency in Certified Health IT

In the U.S., the Office of the National Coordinator for Health Information Technology (ONC) introduced the Health Data, Technology, and Interoperability (HTI-1) Final Rule, which focuses on algorithmic transparency within certified health IT systems, such as Electronic Health Records (EHRs) [3]. This rule establishes transparency requirements for **Predictive Decision Support Interventions (DSIs)**.

The ONC rule mandates that developers of certified health IT must provide specific information about their predictive DSIs, including: ***Source Data:*** *Information about the data used to train, test, and validate the DSI.* **Intervention Logic:** A description of the logic, methods, and underlying assumptions of the DSI. ***Bias Mitigation:*** *Documentation of the steps taken to identify and mitigate bias in the DSI.*

The Role of Explainable AI (XAI) in Clinical Practice

Beyond compliance, the academic and ethical discourse centers on ***Explainable Artificial Intelligence (XAI)****—the set of techniques that allows human users to understand, trust, and manage the outputs of AI models. In medicine, XAI is vital for clinical acceptance, accountability, and continuous improvement.*

A ***clinician*** *requires a* local explanation*—why the model made a specific diagnosis for a single patient—to validate the AI's suggestion against their expertise. Conversely, a* ***regulator*** *or* ***hospital administrator*** *is more concerned with a* global explanation*—the model's overall performance characteristics and general operating principles.*

The future of medical AI hinges on the successful translation of these regulatory and academic principles into practical, user-friendly tools. The goal is to move from simply providing data about *the model to offering meaningful, context-specific explanations* from *the model. For more in-depth analysis on this topic, the resources at www.rasitdinc.com provide expert commentary on the intersection of digital health, AI, and regulatory strategy, offering valuable insights for professionals navigating this*

complex landscape.

Conclusion: Building Trust Through Disclosure

Transparency is the foundation for building trust in medical AI. The global regulatory landscape is rapidly evolving to mandate clear, comprehensive disclosure regarding the design, performance, and limitations of these critical systems. The ongoing work in Explainable AI (XAI) will continue to bridge the gap between technical complexity and clinical usability, ensuring that AI remains a powerful, yet accountable, partner in patient care.

References

[1] U.S. Food and Drug Administration. Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles*. https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles [2] EU Artificial Intelligence Act.* Article 13: Transparency and Provision of Information to Deployers*. [https://artificialintelligenceact.eu/article/13/] (https://artificialintelligenceact.eu/article/13/) [3] U.S. Department of Health and Human Services.* Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Blocking*. Federal Register. [https://www.federalregister.gov/documents/2024/01/09/2023-28857/health-data-technology-and-interoperability-certification-program-updates-algorithm-transparency-and] (https://www.federalregister.gov/documents/2024/01/09/2023-28857/health-data-technology-and-interoperability-certification-program-updates-algorithm-transparency-and)
