Letter to the Editor

DOI: 10.4274/jarem.galenos.2026.27879

J Acad Res Med

Responsible Healthcare AI Requires Responsible Evidence: Methodological Governance for Clinical Implementation

Ahmet Aykut

University of Health Sciences Türkiye, Van Education and Research Hospital, Department of Emergency Medicine, Van, Türkiye

Cite this article as: Aykut A. Responsible healthcare AI requires responsible evidence: methodological governance for clinical implementation. J Acad Res Med. [Epub Ahead of Print]

Keywords: Artificial intelligence, clinical validation, methodological transparency

Dear Editor,

The commentary by Öney Doğanyiğit and Yılmaz Altuntaş (1) provides a timely reminder that artificial intelligence (AI) adoption in healthcare should be governed by transparency, privacy, clinical validation, accountability, and lifecycle monitoring. This position is highly relevant, as generative systems and large language models (LLMs) are increasingly framed not only as administrative tools but also as interactive clinical partners.

However, the central challenge is the responsible transformation of not only healthcare but also evidence. In the absence of structured reporting, external validation, calibration, and post-deployment surveillance, the phrase "clinical validation" may remain too broad to protect patients or guide clinicians. This is particularly important when AI systems are evaluated using convenience datasets, static benchmark questions, single-center records, or unblinded expert panels without sufficient detail on model versioning, prompt design, input preprocessing, missing data handling, reference standards, or error adjudication.

Recent reporting frameworks provide a practical route from principles to implementable evidence. TRIPOD+AI clarifies reporting for diagnostic and prognostic prediction models using regression or machine learning methods (2). DECIDE-AI addresses early-stage clinical evaluation of AI-driven decision-support systems, including workflow interaction and human-system communication (3). For generative models and LLM-based applications, TRIPOD-LLM and MI-CLEAR-LLM highlight aspects that are often invisible in conventional biomedical reporting: model identity, access route, prompt and execution details, stochasticity management, adaptation strategy, independence of test data, and task-specific performance reporting (4,5). These elements are essential because LLM outputs may change with vendor updates, interface modifications, retrieval layers, and prompt variations, even when the nominal clinical task remains the same.

For this reason, healthcare AI governance should include a minimum evidentiary chain before clinical implementation: precise description of the AI system, transparent reporting of the clinical task and intended users, representative test data, clinically meaningful reference standards, external validation when possible, calibration and error analysis, assessment of human-AI interaction, and lifecycle monitoring after deployment. Such a chain would make regulatory and ethical principles auditable rather than merely aspirational.

The commentary rightly emphasizes that AI should augment, not replace, clinicians (1). Meaningful human oversight, however, requires access to interpretable and reproducible evidence. Clinicians cannot responsibly supervise systems whose development, evaluation, and updating are insufficiently reported. Therefore, responsible healthcare AI depends not only on trust, empathy, and regulation but also on methodological transparency. Without robust evidence, responsible transformation may remain incomplete.

Acknowledgements

The author acknowledges the use of artificial intelligence-based language assistance during the preparation of the manuscript. The final content, interpretation, and conclusions are the sole responsibility of the author.

ORCID IDs of the authors: A.A. 0009-0001-3173-8994

Corresponding Author: Ahmet Aykut, MD;
E-mail: ahmet.aykut@gmail.com

Received Date: 07.06.2026 **Accepted Date:** 25.06.2026
Epub: 26.06.2026

Footnotes

Financial Disclosure: The author declared no financial support.

REFERENCES

1. Öney Doğanyiğit S, Yılmaz Altuntaş E. Artificial intelligence in healthcare: a critical moment for responsible transformation. J Acad Res Med. 2026; 16: 1-3.
2. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024; 385: e078378. Erratum in: BMJ. 2024; 385: q902.
3. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al.; DECIDE-AI Expert Group. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ. 2022; 377: e070904.
4. Gallifant J, Afshar M, Ameen S, Aphinyanaphongs Y, Chen S, Cacciamani G, et al. The TRIPOD-LLM reporting guideline for studies using large language models. Nat Med. 2025; 31: 60-9.
5. Park SH, Suh CH, Lee JH, Tejani AS, You SC, Kahn CE, et al. Minimum reporting items for clear evaluation of accuracy reports of large language models in healthcare (MI-CLEAR-LLM): 2025 updates. Korean J Radiol. 2025; 26: 1123-32.