



## Viewpoint

# Patient Safety and Artificial Intelligence in Clinical Care

Raj M. Ratwani, PhD; David W. Bates, MD; David C. Classen, MD

Artificial intelligence (AI), the next health technology disruptor, is upon us and could greatly improve patient safety. Examples include detection and prediction of sepsis, pressure ulcers, postpartum hemorrhage, adverse drug events, and patient decompensation, to name a few.<sup>1,2</sup> Yet if not designed, developed, implemented, and used appropriately, AI in clinical settings may contribute to patient harm. For example, a widely used AI system intended to detect patients with sepsis only picked up 7% of 2552 patients with sepsis, resulting in delayed antibiotic administration and failure to identify 1709 patients with sepsis that the hospital identified through other means.<sup>3</sup>

To realize the potential benefits of AI and reduce risk, patient safety safeguards must be developed and put in place.<sup>4</sup> Historically, health care has been slow to respond to patient safety challenges presented by new technologies. For example, the 2009 Health Information Technology for Economic and Clinical Health Act that spurred widespread electronic health record (EHR) adoption was closely followed by calls to address important safety concerns.<sup>5</sup> These safety concerns included gag clauses in EHR vendor contracts preventing health care professionals from sharing patient safety issues, the need for improved usability and safety certification criteria to promote better vendor design and development, and the need for frameworks to inform safe EHR implementation and use. It has taken almost a decade to begin to address these issues, and serious safety challenges persist.<sup>6</sup>

President Biden's October 30, 2023, executive order calls for the Secretary of Health and Human Services to establish an AI safety program within 365 days that (1) establishes a framework for approaches to identify clinical errors resulting from AI and creates specifications for a central tracking repository, (2) analyzes data and generates guidelines for improvement, and (3) disseminates these guidelines to health care organizations (eg, hospitals, clinics, and other organizations providing patient care) and other stakeholders.<sup>7</sup> The executive order lays the foundation for the development of broader patient safety safeguards to protect against patient safety risks that may emerge from the use of AI in clinical settings. We provide 3 recommendations for health care organizations and other stakeholders to consider as part of the AI safety program.

Author affiliations and article information are listed at the end of this article.

## Develop and Encourage Use of Guidelines for Safe AI Implementation in Clinical Settings

Safe and effective AI implementation in clinical settings requires consideration of the health care workforce, existing technologies, policies, and processes. High-risk domains such as aviation and defense have developed a theoretical framework, called human-systems integration, to effectively and safely integrate machines into the work environment. Leveraging principles from these other domains, specific guidelines to inform safe implementation of AI algorithms for clinical use should be developed. Patient safety AI guidelines must address issues such as the tendency for people to overly trust AI, algorithm drift that results in variation in AI performance in different settings, and health care professional complacency. These guidelines can serve as a proactive approach for safely implementing AI technology by considering critical safety issues before widespread use.

The Office of the National Coordinator for Health Information Technology (ONC) funded the development of EHR implementation guides for health care organization use, called the Safety Assurance Factors for EHR Resilience (SAFER) guides. These guides can serve as a model for what will

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.

need to be developed for the safe use of AI embedded in EHRs. The US Centers for Medicare and Medicaid Services recently required hospitals to attest to annual use of the SAFER guides, and a similar requirement for AI-focused guidelines would encourage use. Similarly, the US Food and Drug Administration (FDA) should consider guidelines for health care organizations to support safe use of AI-enabled medical devices.

---

### Frequently Monitor AI for Patient Safety Risks

Health care organizations should have processes in place to frequently monitor for AI-related patient safety issues. The Biden executive order describes AI red-teaming, which is “a structured testing effort to find flaws and vulnerabilities in an AI system, often in a controlled environment and in collaboration with developers of AI.”<sup>7</sup> Health care organizations should conduct patient safety-focused testing of AI on a frequent basis, given the dynamic nature of AI—it is fundamentally different than rule-based decision support. One way to monitor AI is through the application of test cases that represent the health care organization's patients to assess the AI system for safety issues. For AI systems that are commercially developed, the developer should provide a standardized set of test cases that all health care organizations can apply to their implemented AI system with the opportunity for local customization to represent patient population differences. Some AI systems may be developed by the health care organization, and the development of test cases should be the responsibility of that organization. This type of safety assessment already exists for EHR-based decision support and is deployed across the US in approximately 3000 hospitals.<sup>8</sup> It serves to identify patient safety issues at hospitals that may emerge from local customization in EHR products. This decision support assessment is currently being customized to work for AI applications and could serve as one method for AI patient safety monitoring.

Health care organizations will need to create appropriate governance structures to address adding new AI tools, frequency of monitoring to identify patient safety issues, and development of solutions. The Joint Commission should consider basic accreditation standards to promote creation of these governance structures and clear processes for monitoring for AI-related patient safety issues.

---

### Develop Processes to Enable Traceability of AI System Contributions to Patient Safety Events

When it is recognized that an AI system may have contributed to patient harm, a thorough event review will be required to examine how AI may have been involved. AI systems should capture metadata to support detailed review and analysis of how the AI system was operating at the time of the patient safety incident. A core set of metadata elements should be established and required of all AI systems in clinical use to support traceability so that consistent and thorough review can be conducted when needed. Traceability is also important for addressing AI bias that can lead to inequitable patient outcomes, and our recommendation is aligned with recommendations from the Gravity Project, which is focused on standardization of social determinants of health data.<sup>9</sup> This type of effort is likely best achieved through a public-private effort that brings together federal agencies such as the FDA, the ONC, and the National Institute of Standards and Technology (NIST) with industry, health care organizations, and standards-setting stakeholders. Determining what metadata elements should be captured and standardizing their capture across AI systems will take time, and in the interim, health care organizations should seek out AI systems that have metadata to support safety incident review.

## Conclusions

AI safety program efforts should be aligned with existing AI frameworks and guidelines proposed by the FDA, NIST, and other federal agencies.<sup>10</sup> Recognizing that many health care organizations may not have the resources to implement AI safeguards, federal and state programs may be required to support these efforts. Some aspects of our recommendations can be implemented within 365 days, while others, such as new policies and standards, will take more time. With the speed at which AI is being developed and is evolving, the executive order should serve as the impetus for the rapid development of patient-focused safeguards in a way that can also enable adaptation to the rapidly changing landscape.

---

### ARTICLE INFORMATION

**Published:** February 23, 2024. doi:[10.1001/jamahealthforum.2023.5514](https://doi.org/10.1001/jamahealthforum.2023.5514)

**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](https://creativecommons.org/licenses/by/4.0/). © 2024 Ratwani RM et al. *JAMA Health Forum*.

**Corresponding Author:** Raj M. Ratwani, PhD, MedStar Health National Center for Human Factors in Healthcare, 3007 Tilden St, NW, Ste 6N, Washington, DC 20008 ([raj.m.ratwani@medstar.net](mailto:raj.m.ratwani@medstar.net)).

**Author Affiliations:** MedStar Health National Center for Human Factors in Healthcare, Washington, DC (Ratwani); Georgetown University School of Medicine, Washington, DC (Ratwani); Brigham and Women's Hospital, Boston, Massachusetts (Bates); University of Utah, Salt Lake City (Classen).

**Conflict of Interest Disclosures:** Dr Ratwani reported grants from the Agency for Healthcare Research and Quality during the submitted work. Dr Bates reported grants and personal fees from the Agency for Healthcare Research and Quality during the submitted work and grants and personal fees from EarlySense; personal fees from CDI Negev; equity from Valera Health, CLEW, and MDClone; and personal fees and equity from AESOP, FeelBetter, and Guided Clinical Solutions outside the submitted work; in addition, Dr Bates had a patent for PHC-028654 issued, Massachusetts General Hospital. Dr Classen reported part-time employment with Pascal Metrics, a federally certified patient safety organization, outside the submitted work.

**Funding/Support:** This work was supported by grant R01HS026481 from the Agency for Healthcare Research and Quality to the MedStar Health Research Institute and Dr Ratwani.

**Role of the Funder/Sponsor:** The Agency for Healthcare Research and Quality had no role in the preparation, review, or approval of the manuscript and decision to submit the manuscript for publication.

### REFERENCES

1. Bates DW, Levine D, Syrowatka A, et al. The potential of artificial intelligence to improve patient safety: a scoping review. *NPJ Digit Med*. 2021;4(1):54. doi:[10.1038/s41746-021-00423-6](https://doi.org/10.1038/s41746-021-00423-6)
2. Alsentzer E, Rasmussen MJ, Fontoura R, et al. Zero-shot interpretable phenotyping of postpartum hemorrhage using large language models. *NPJ Digit Med*. 2023;6(1):212. doi:[10.1038/s41746-023-00957-x](https://doi.org/10.1038/s41746-023-00957-x)
3. Classen DC, Longhurst C, Thomas EJ. Bending the patient safety curve: how much can AI help? *NPJ Digit Med*. 2023;6(1):2. doi:[10.1038/s41746-022-00731-5](https://doi.org/10.1038/s41746-022-00731-5)
4. Gottlieb S, Silvis L. Regulators face novel challenges as artificial intelligence tools enter medical practice. *JAMA Health Forum*. 2023;4(6):e232300. doi:[10.1001/jamahealthforum.2023.2300](https://doi.org/10.1001/jamahealthforum.2023.2300)
5. Committee on Patient Safety and Health Information Technology; Institute of Medicine. *Health IT and Patient Safety: Building Safer Systems for Better Care*. National Academies Press; 2011.
6. Ratwani RM, Reider J, Singh H. A decade of health information technology usability challenges and the path forward. *JAMA*. 2019;321(8):743-744. doi:[10.1001/jama.2019.0161](https://doi.org/10.1001/jama.2019.0161)
7. Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. Accessed November 10, 2023. <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>
8. Classen DC, Holmgren AJ, Co Z, et al. National trends in the safety performance of electronic health record systems from 2009 to 2018. *JAMA Netw Open*. 2020;3(5):e205547. doi:[10.1001/jamanetworkopen.2020.5547](https://doi.org/10.1001/jamanetworkopen.2020.5547)
9. The Gravity Project. About us. Accessed December 24, 2023. <https://thegravityproject.net/overview/>
10. Gottlieb S, Silvis L. How to safely integrate large language models into health care. *JAMA Health Forum*. 2023; 4(9):e233909. doi:[10.1001/jamahealthforum.2023.3909](https://doi.org/10.1001/jamahealthforum.2023.3909)