

Implementing Trustworthy AI in VA High Reliability Health Care Organizations

David B. Isaacks, FACHE^a; Andrew A. Borkowski, MD^{a,b,c}

Background: Artificial intelligence (AI) has great potential to improve health care quality, safety, efficiency, and access. However, the widespread adoption of health care AI needs to catch up to other sectors. Challenges, including data limitations, misaligned incentives, and organizational obstacles, have hindered implementation. Strategic demonstrations, partnerships, aligned incentives, and continued investment are needed to enable responsible adoption of AI. High reliability health care organizations offer insights into safely implementing major initiatives through frameworks like the Patient Safety Adoption Framework, which provides practical guidance on leadership, culture, process, measurement, and person-centeredness to successfully adopt safety practices. High reliability health care organizations ensure consistently safe and high quality care through a culture focused on reliability, accountability, and learning from errors and near misses.

Observations: The Veterans Health Administration applied a high reliability health care model to instill safety principles and improve outcomes. As the use of AI becomes more widespread, ensuring its ethical development is crucial to avoiding new risks and harm. The US Department of Veterans Affairs National AI Institute proposed a Trustworthy AI Framework tailored for federal health care with 6 principles: purposeful, effective and safe, secure and private, fair and equitable, transparent and explainable, and accountable and monitored. This aims to manage risks and build trust.

Conclusions: Combining these AI principles with high reliability safety principles can enable successful, trustworthy AI that improves health care quality, safety, efficiency, and access. Overcoming AI adoption barriers will require strategic efforts, partnerships, and investment to implement AI responsibly, safely, and equitably based on the health care context.

Author affiliations can be found at the end of this article.

Correspondence:

Andrew Borkowski
(andrew.borkowski@va.gov)

Fed Pract. 2024;41(2).
Published online February 15.
doi:10.12788/fp.0454

Artificial intelligence (AI) has lagged in health care but has considerable potential to improve quality, safety, clinician experience, and access to care. It is being tested in areas like billing, hospital operations, and preventing adverse events (eg, sepsis mortality) with some early success. However, there are still many barriers preventing the widespread use of AI, such as data problems, mismatched rewards, and workplace obstacles. Innovative projects, partnerships, better rewards, and more investment could remove barriers. Implemented reliably and safely, AI can add to what clinicians know, help them work faster, cut costs, and, most importantly, improve patient care.¹

AI can potentially bring several clinical benefits, such as reducing the administrative strain on clinicians and granting them more time for direct patient care. It can also improve diagnostic accuracy by analyzing patient data and diagnostic images, providing differential diagnoses, and increasing access to care by providing medical information and essential online services to patients.²

HIGH RELIABILITY ORGANIZATIONS

High reliability health care organizations have considerable experience safely launching new programs. For example, the Patient

Safety Adoption Framework gives practical tips for smoothly rolling out safety initiatives (Table 1). Developed with experts and diverse views, this framework has 5 key areas: leadership, culture and context, process, measurement, and person-centeredness. These address adoption problems, guide leaders step-by-step, and focus on leadership buy-in, safety culture, cooperation, and local customization. Checklists and tools make it systematic to go from ideas to action on patient safety.³

Leadership involves establishing organizational commitment behind new safety programs. This visible commitment signals importance and priorities to others. Leaders model desired behaviors and language around safety, allocate resources, remove obstacles, and keep initiatives energized over time through consistent messaging.⁴ Culture and context recognizes that safety culture differs across units and facilities. Local input tailors programs to fit and examines strengths to build on, like psychological safety. Surveys gauge the existing culture and its need for change. Process details how to plan, design, test, implement, and improve new safety practices and provides a phased roadmap from idea to results. Measurement collects data to drive improvement and show impact. Metrics

TABLE 1 Patient Safety Adoption Framework³

Domain	Subdomain	Explanation
Leadership	Governance	Establishes organizational and system-level structures and networks, including decision making, communication, and information flow
	Accountability	Leaders at every level are accountable for achieving results and acting in ways that reflect organizational values
	Prioritization	Requires stakeholders to collaborate toward a shared patient safety goal by evaluating the current state and assessing data
Culture and context	—	Profoundly affects other domains; strong safety culture promotes transparency and reduces adverse events
Process	Cocreation	Partnering with patients, families, and staff throughout the improvement process
	High reliability	Integrating principles into processes to reduce system failures and respond effectively when failures occur
	Engagement	Engaging stakeholders throughout implementation by building consensus, sharing stories, and clear communication
Measurement	—	Data focused on crucial aspects of care, centered around patient and clinician, and incorporate patient-reported outcomes
Person-centered	—	Individual preferences and values are central to their health care

track progress and allow benchmarking. Person-centeredness puts patients first in safety efforts through participation, education, and transparency.

The Veterans Health Administration piloted a comprehensive high reliability hospital (HRH) model. Over 3 years, the Veterans Health Administration focused on leadership, culture, and process improvement at a hospital. After initiating the model, the pilot hospital improved its safety culture, reported more minor safety issues, and reduced deaths and complications better than other hospitals. The high-reliability approach successfully instilled principles and improved culture and outcomes. The HRH model is set to be expanded to 18 more US Department of Veterans Affairs (VA) sites for further evaluation across diverse settings.⁵

TRUSTWORTHY AI FRAMEWORK

AI systems are growing more powerful and widespread, including in health care. Unfortunately, irresponsible AI can introduce new harm. ChatGPT and other large language models, for example, sometimes are known to provide erroneous information in a compelling way. Clinicians and patients who use such programs can act on such in-

formation, which would lead to unforeseen negative consequences. Several frameworks on ethical AI have come from governmental groups.⁶⁻⁹ In 2023, the VA National AI Institute suggested a Trustworthy AI Framework based on core principles tailored for federal health care. The framework has 6 key principles: purposeful, effective and safe, secure and private, fair and equitable, transparent and explainable, and accountable and monitored (Table 2).¹⁰

First, AI must clearly help veterans while minimizing risks. To ensure purpose, the VA will assess patient and clinician needs and design AI that targets meaningful problems to avoid scope creep or feature bloat. For example, adding new features to the AI software after release can clutter and complicate the interface, making it difficult to use. Rigorous testing will confirm that AI meets intent prior to deployment. Second, AI is designed and checked for effectiveness, safety, and reliability. The VA pledges to monitor AI's impact to ensure it performs as expected without unintended consequences. Algorithms will be stress tested across representative datasets and approval processes will screen for safety issues. Third, AI models are secured from vulnerabilities and misuse. Technical

TABLE 2 Trustworthy AI Principles⁹

Principle	Explanation
Purposeful	AI provides clear benefits to patients with minimal risks
Effective and safe	Systems are created and supervised with the utmost attention to accuracy, reliability, and robustness; any possible risks are identified and proactively managed to guarantee the safety and well-being of patients
Secure and private	AI models are resilient against vulnerabilities and malicious exploitation; patient data is maintained in accordance with laws and federal data ethics principles to preserve privacy
Fair and equitable	It is important to oversee and keep track of AI systems to ensure their algorithms are not biased or discriminatory
Transparent and explainable	Patients need to be informed about the use of AI systems in health care and the data used by those systems; the federal government should provide clear and concise information on the workings of AI systems and how they are utilized in making health care decisions
Accountable and monitored	Designating the accountable parties, proactively monitoring and evaluating inputs and outcomes, and addressing concerns with the appropriate parties to ensure continued improvement

Abbreviation: AI, artificial intelligence.

controls will prevent unauthorized access or changes to AI systems. Audits will check for appropriate internal usage per policies. Continual patches and upgrades will maintain security. Fourth, the VA manages AI for fairness, avoiding bias. They will proactively assess datasets and algorithms for potential biases based on protected attributes like race, gender, or age. Biased outputs will be addressed through techniques such as data augmentation, reweighting, and algorithm tweaks. Fifth, transparency explains AI's role in care. Documentation will detail an AI system's data sources, methodology, testing, limitations, and integration with clinical workflows. Clinicians and patients will receive education on interpreting AI outputs. Finally, the VA pledges to closely monitor AI systems to sustain trust. The VA will establish oversight processes to quickly identify any declines in reliability or unfair impacts on subgroups. AI models will be retrained as needed based on incoming data patterns.

Each Trustworthy AI Framework principle connects to others in existing frameworks. The purpose principle aligns with human-centric AI focused on benefits. Effectiveness and safety link to technical robustness and risk management principles. Security maps to privacy protection principles. Fairness connects to principles of

avoiding bias and discrimination. Transparency corresponds with accountable and explainable AI. Monitoring and accountability tie back to governance principles. Overall, the VA framework aims to guide ethical AI based on context. It offers a model for managing risks and building trust in health care AI.

Combining VA principles with high-reliability safety principles can ensure that AI benefits veterans. The leadership and culture aspects will drive commitment to trustworthy AI practices. Leaders will communicate the importance of responsible AI through words and actions. Culture surveys can assess baseline awareness of AI ethics issues to target education. AI security and fairness will be emphasized as safety critical. The process aspect will institute policies and procedures to uphold AI principles through the project lifecycle. For example, structured testing processes will validate safety. Measurement will collect data on principles like transparency and fairness. Dashboards can track metrics like explainability and biases. A patient-centered approach will incorporate veteran perspectives on AI through participatory design and advisory councils. They can give input on AI explainability and potential biases based on their diverse backgrounds.

CONCLUSIONS

Joint principles will lead to successful AI that improves care while proactively managing risks. Involve leaders to stress the necessity of eliminating biases. Build security into the AI development process. Co-design AI transparency features with end users. Closely monitor the impact of AI across safety, fairness, and other principles. Adhering to both Trustworthy AI and high reliability organizations principles will earn veterans' confidence. Health care organizations like the VA can integrate ethical AI safely via established frameworks. With responsible design and implementation, AI's potential to enhance care quality, safety, and access can be realized.

Acknowledgments

We would like to acknowledge Joshua Mueller, Theo Tiffney, John Zachary, and Gil Alterovitz for their excellent work creating the VA Trustworthy Principles. This material is the result of work supported by resources and the use of facilities at the James A. Haley Veterans' Hospital.

Author affiliations

^aVeterans Affairs Sunshine Healthcare Network, Tampa, Florida

^bUniversity of South Florida Morsani College of Medicine, Tampa

^cVeterans Affairs National Artificial Intelligence Institute

Author disclosures

The authors report no actual or potential conflicts of interest or outside sources of funding with regard to this article.

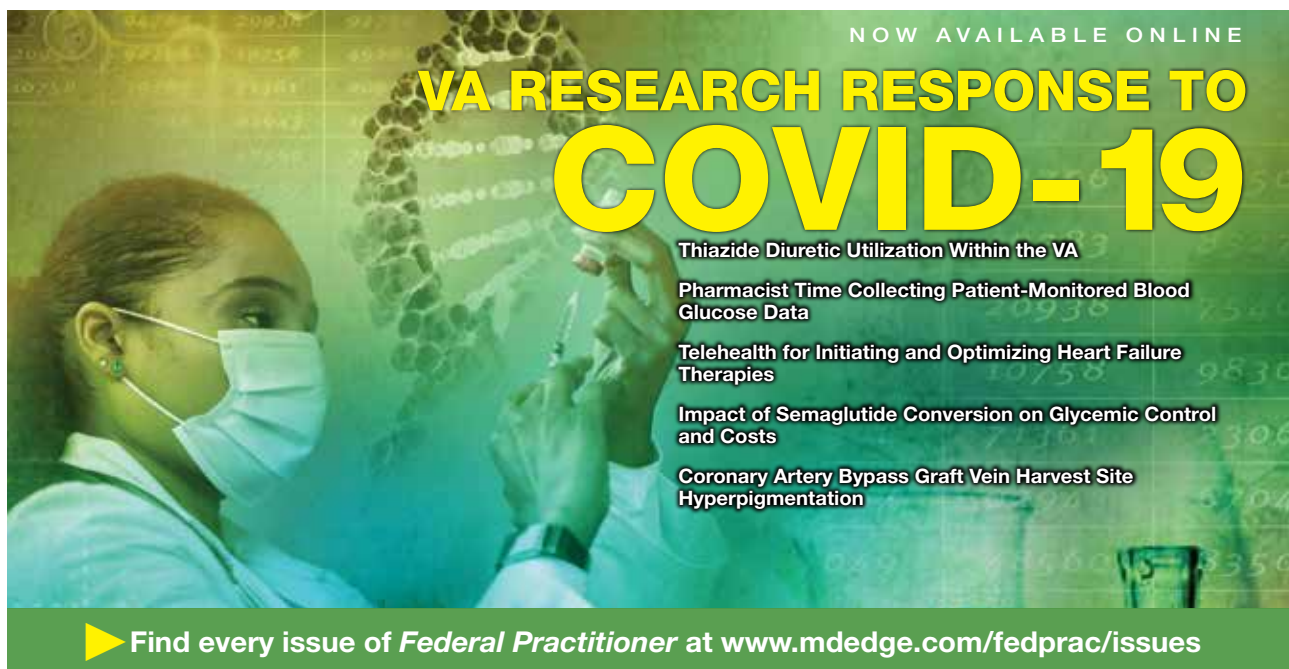
Disclaimer

The opinions expressed herein are those of the authors and do not necessarily reflect those of *Federal Practitioner*, Frontline Medical Communications Inc., the US Government, or any of its agencies.

tioner, Frontline Medical Communications Inc., the US Government, or any of its agencies.

References

1. Sahni NR, Carrus B. Artificial intelligence in U.S. health care delivery. *N Engl J Med*. 2023;389(4):348-358. doi:10.1056/NEJMra2204673
2. Borkowski AA, Jakey CE, Mastorides SM, et al. Applications of ChatGPT and large language models in medicine and health care: benefits and pitfalls. *Fed Pract*. 2023;40(6):170-173. doi:10.12788/fp.0386
3. Moyal-Smith R, Margo J, Maloney FL, et al. The patient safety adoption framework: a practical framework to bridge the know-do gap. *J Patient Saf*. 2023;19(4):243-248. doi:10.1097/PTS.0000000000001118
4. Isaacks DB, Anderson TM, Moore SC, Patterson W, Govindan S. High reliability organization principles improve VA workplace burnout: the Truman THRIVE2 model. *Am J Med Qual*. 2021;36(6):422-428. doi:10.1097/01.JMQ.0000735516.35323.97
5. Sculli GL, Pendley-Louis R, Neily J, et al. A high-reliability organization framework for health care: a multiyear implementation strategy and associated outcomes. *J Patient Saf*. 2022;18(1):64-70. doi:10.1097/PTS.0000000000000788
6. National Institute of Standards and Technology. AI risk management framework. Accessed January 2, 2024. <https://www.nist.gov/itl/ai-risk-management-framework>
7. Executive Office of the President, Office of Science and Technology Policy. Blueprint for an AI Bill of Rights. Accessed January 11, 2024. <https://www.whitehouse.gov/ostp/ai-bill-of-rights>
8. Executive Office of the President. Executive Order 13960: promoting the use of trustworthy artificial intelligence in the federal government. *Fed Regist*. 2020;89(236):78939-78943.
9. Biden JR. Executive Order on the safe, secure, and trustworthy development and use of artificial intelligence. Published October 30, 2023. Accessed January 11, 2024. <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/>
10. US Department of Veterans Affairs. Trustworthy AI. Accessed January 11, 2024. <https://department.va.gov/ai/trustworthy/>



NOW AVAILABLE ONLINE

VA RESEARCH RESPONSE TO COVID-19

- Thiazide Diuretic Utilization Within the VA
- Pharmacist Time Collecting Patient-Monitored Blood Glucose Data
- Telehealth for Initiating and Optimizing Heart Failure Therapies
- Impact of Semaglutide Conversion on Glycemic Control and Costs
- Coronary Artery Bypass Graft Vein Harvest Site Hyperpigmentation

▶ Find every issue of *Federal Practitioner* at www.mdedge.com/fedprac/issues