COMMENTARY

Artificial Intelligence in Health, Health Care, and Biomedical Science: An AI Code of Conduct Principles and Commitments Discussion Draft

Editors: Laura Adams, MS, National Academy of Medicine; **Elaine Fontaine, BS,** National Academy of Medicine; **Steven Lin, MD,** Stanford University School of Medicine; **Trevor Crowell, BA,** Stanford University School of Medicine; **Vincent C. H. Chung, MSc, PhD,** Faculty of Medicine, The Chinese University of Hong Kong; and **Andrew A. Gonzalez, MD, JD, MPH,** Regenstrief Institute Center for Health Services Research and Indiana University School of Medicine

April 8, 2024

Introduction

This commentary presents initial concepts and content that the Steering Committee feel may be important to a draft Code of Conduct framework for use in the development and application of artificial intelligence (AI) in health, health care, and biomedical science. It introduces suggested principles and commitments, discusses the aim of a continuously learning h

In just the year prior to this commentary's publication, the landscape has changed. Advanced predictive and generative AI and language models have appeared across multiple application domains, including the rapid evolution and diffusion of large language models (LLMs), such as ChatGPT by Open AI which was made publicly available in 2022. Just as AI technologies are rapidly advancing, it is essential that health system stakeholders—individually and collectively—rapidly learn, adapt, and align on necessary guardrails responsible use of AI in health, health care, and biomedical science (Hutson, 2022). This imperative is consistent with the LHS, with core principles building upon the landmark publications, To Err is Human (IOM, 2000) and the Crossing the Quality Chasm Series (IOM, 2001), which identified quality health care as that which is: safe, effective, patient-centered, timely, efficient, and equitable. These principles have been expanded over a dozen years to embrace both health and health care, and add the critical care elements of transparency, accountability, and security. In addition to establishing common ground in a fragmented ecosystem, the core LHS principles also serve as a framework for increasing system trust, including in health AI.

The rapidly expanding use of AI in health, health care, and biomedical science amplifies existing risks and creates new ones across the health and medicine sectors from research to clinical

BOX 1 | Description of Complex Adaptive Systems Theory for Health Care

In the complex adaptive health care system, interdependent elements (e.g., patients, clinicians, policies, and organizations—including hospitals, clinics, payers, pharmacies, and regulators) act independently, making decentralized decisions.

These decisions may be impacted by external factors and create feedback loops or result in nonlinear impacts (e.g., small changes lead to disproportionate effects), resulting in emergent system behaviors. That is, the system experiences outcomes or emergent behaviors that are not solely attributable to the actions of single actor but rather to the interaction of system elements.

However, simple rules implemented locally may amplify outcomes at the system level due to feedback loops and non-linear interactions. Small changes made by individual elements can cascade through the system, resulting in significant changes in overall behavior or system state.

Technology companies; health-focused coalitions; researchers; and local, national, and international governmental agencies have published guidance on responsible AI, but these efforts have not yet been harmonized or compared for overlap and completeness. With momentum building around the use of AI and demand for guardrails in the health sector, the value and critical nature of stakeholder alignment is clear (Dorr et al., 2023). This moment presents a unique opportunity for the health care community, within the context of a competitive marketplace, to act collectively and with intention to design the future of health, health care, and biomedical science in the era of AI. Alignment and transparent rapid cycle learning is necessary to realize the promise and avoid the peril associated with AI in the health sector. This collective effort is aligned with and complementary to related efforts across the field of health, including NAM convenings to address LLMs in health care, and will serve as the foundation for ongoing work to provide more detailed guidance on accountability and priorities for centralized infrastructure needed to support responsible AI.

Overview of the Literature and Published Guiding Principles

In recognition of the importance of building on previous efforts to define key principles to ensure trustworthy use of AI in the health ecosystem, the editors of this publication conducted a landscape review of existing health care AI guidelines, frameworks, and principles. A 2022 systematic literature review by Siala and Wang (2022) identified five key characteristics of socially responsible AI: human-centeredness, inclusiveness, fairness, transparency, and sustainability. This 2022 framework was then compared with 56 documents drawn from 3 core domains to identify similarities and gaps: scientific literature published between 2018–2023 that focused on responsible AI principles; guidance developed by medical specialty societies for physicians using AI; and frameworks, policies, and guidance issued by the federal government through May 2023, including the foundational National Institute of Standards and Technology AI Risk Management Framework (National Institute of Standards and Technology, 2023).

As the editors synthesized content extracted from the 56 publications, 2 consistent elements emerged: fairness and transparency were well-represented across the reviewed documents, but inclusiveness, sustainability, and human-centricity were not. Importantly, this review revealed that while the 2022 functional framework established a necessary baseline, it omitted or provided inadequate attention to themes that are essential to a forward-looking evaluation of guiding principles for the LHS and ethical AI, including accountability, data protection, ongoing assessment, and safety. This review therefore identified the following Code Principles based on the core LHS principles: engaged, safe, effective, equitable, efficient, accessible, transparent, accountable, secure, and adaptive. These core LHS principles define the agreed upon values and norms required to demonstrate trustworthiness between and among the participants in the health system; the trust, in turn, is foundational and embedded in the LHS.

One relevant additional feature was identified for inclusion by this review—international guidance and regulation—given that AI built by global companies will be used inside and outside the United States, and so four additional documents were also reviewed: international guidance on responsible AI from the World Health Organization, United Nations, European Union, and the Organisation for Economic Co-operation and Development (High-Level Expert Group on AI, 2019; Organisation for Economic Co-operation and Development, 2023; United Nations System, 2022; World Health Organization, 2021). The principles presented in these documents were also compared to the 2022 framework. The principles in the international publications did align with the Code Principles, but also included environmental protection or efficiency, which is not present in the 56 U.S.-focused publications but is clearly an important consideration moving forward.

Landscape Review Gaps and Opportunities

Among the 60 publications reviewed, 3 areas of inconsistency were identified: inclusive collaboration, ongoing safety assessment, and efficiency or environmental protection. These issues

1. Solicit key stakeholder feedback and public comment on the draft Code of Conduct Framework's Code Principles and Code Commitments for incorporation into a final publication.
2. Convene working groups representing critical contributors to ensuring responsible AI in health, health care, and biomedical science. Each group will define the expected behaviors (conduct), accountabilities, and relationships to other key stakeholders throughout each stage of the AI life cycle. Upon completing this group work, cross-cutting reviews from experts in equity and ethics; workforce and clinician well-being; quality and safety; and individuals, patients, and clinicians will be solicited, and their feedback will be incorporated. The working groups will consider how to address the required overall health system changes to realize the Code Commitments.
3. The draft Code of Conduct Framework's Code Principles and Code Commitments will be tested by case studies beginning with individuals and patient advocates, as well as health system and product development partners.
4. Key stakeholders involved in AI governance, including federal agencies with relevant responsibilities, professional societies, and related technology associations will be consulted.
5. An NAM Special Publication will be released, containing 1) the final AI Code of Conduct framework, modeled on the LHS core principles, informed by public input, and vetted and co-created with the working groups and external consultations, and 2) recommended options for implementation, monitoring, and continuous improvement of the Code of Conduct framework.

Conclusion

After decades of progress toward a data-driven health system, advanced AI methods and systems present a new and important opportunity to achieve the vision of a learning health system. These adaptive technologies also present risks, particularly when applied in a complex system, and therefore must be carefully and collectively managed. Based on a bounded review of the literature and guidance on responsible AI in health and health care, informed by ongoing dialogue with national thought leaders, and mapped to the principles of the continuously learning health system, this paper proposes a harmonized draft AI Code of Conduct framework. The Code Principles and the proposed Code Commitments reflect simple guideposts to guide and gauge behavior in a complex system and provide a starting point for real-time decision making and detailed implementation plans to promote the responsible use of AI. Engagement of all key stakeholders in the co-creation of this Code of Conduct framework is essential to ensure the intentional design of the future of AI-enabled health, health care, and biomedical science that advances the vision of health and well-being for all.

References

1. Advisory Committee on the Safety of Nuclear Installations. 1994. *Study group on human factors. Third report: Organising for safety.* H.M.S.O., London, United Kingdom.
2. Allianz Research. 2023. No quick wins: More jobs but little productivity in the Eurozone. *Allianz Research*, tle0(e3)20(.)(H.M1(:

11. Glover, W. J., Z. Li, and D. Pachamanova. 2022. The AI-enhanced future of health care administrative task management. *NEJM Catalyst.*

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/ (accessed November 15, 2023).
37. United Nations System. 2022. *Principles for the ethical use of artificial intelligence in the United Nations system.* Available at: https://unsceb.org/principles-ethical-use-artificial-intelligence-united-nations-system (accessed February 27, 2024).
38. World Health Organization. 2021. *Ethics and governance of artificial intelligence for health: WHO guidance.* Geneva, Switzerland. Available at: https://www.who.int/publications/i/item/9789240029200 (accessed February 27, 2024).

DOI

https://doi.org/10.31478/202403a

Suggested Citation

Adams, L., E. Fontaine, S. Lin, T. Crowell, V. C. H. Chung, and A. A. Gonzalez, editors. 2024. Artificial intelligence in health, health care and biomedical science: An AI code of conduct framework principles and commitments discussion draft. *NAM Perspectives.* Commentary, National Academy of Medicine, Washington, DC. https://doi.org/10.31478/202403a.

Editor Information

Laura Adams, MS, is a Senior Advisor at the National Academy of Medicine. **Elaine Fontaine, BS,** is a Consultant at the National Academy of Medicine. **Steven Lin, MD,** is a Clinical Professor of Medicine, the Section Chief of General Primary Care, and the Director of the Stanford Healthcare AI Applied Research Team Division of Primary Care and Population Health at Stanford University School of Medicine. **Trevor Crowell, BA,** is a Research Associate with the Stanford Healthcare AI Applied Research Team (HEA3RT) at the Stanford University School of Medicine. **Vincent C. H. Chung, PhD,** is an Associate Professor at the JC School of Public Health and Primary Care, The fessor cuector forDat a scienct at the
RigestrhieflestituateCceneor
icalOut coced& Qualsit