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

This paper was developed under the auspices of the Steering Committee of the National Academy of Medicine (NAM)'s project on Artificial Intelligence in Health, Health Care, and Biomedical Science, including **Andrew Bindman,** Kaiser Permanente; **Grace Cordovano,** Enlightening Results; **Jodi Daniel,** Crowell & Moring; **Wyatt Decker,** UnitedHealth Group; **Peter Embí,** Vanderbilt University Medical Center; **Gianrico Farrugia,** Mayo Clinic; **Kadija Ferryman,** Johns Hopkins University; **Sanjay Gupta,** Emory University; **Eric Horvitz,** Microsoft; **Roy Jakobs,** Royal Philips; **Kevin Johnson,** University of Pennsylvania; **Kedar Mate,** Institute for Healthcare Improvement; **Deven McGraw,** Ciitizen; **Bakul Patel,** Google; **Philip R. O. Payne,** Washington University School of Medicine; **Vardit Ravitsky,** The Hastings Center; **Suchi Saria,** Johns Hopkins University and Bayesian Health; **Eric Topol,** Scripps Research Translational Institute; and **Selwyn M. Vickers,** Memorial Sloan Kettering Cancer Center.

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.

Background

As an emergent constellation of technologies, AI presents both unprecedented opportunities and potential risks for human health and well-being. At the October 2016 launch of the Leverhulme Centre for the Future of Intelligence, Stephen Hawking observed, "Every aspect of our lives will be transformed. In short, success in creating AI could be the biggest event in the history of our civilisation" (Hern, 2016).

Since the early 1970s, the increasing deployment of digital tools in health care and biomedical science has led to an explosive generation of health data (National Academy of Medicine, 2022). Leveraging these data to transform health outcomes is the aim of a continuously learning health system (LHS), "one in which knowledge generation is so embedded into the core of the practice of medicine that it is a natural outgrowth and product of the healthcare delivery process and leads to continual improvement in care" (IOM, 2007, page 6). This may, for some time, be a vision in progress, but developments in science, technology, and practice are rapidly setting the stage for its actualization. Until recently, progress in meaningful data use has occurred incrementally through the use of expert systems, clinical decision support algorithms, predictive modeling, big data analytics, and machine learning. Additionally, to date, there has been limited translation of exciting AI prototypes and models into practice. In 2022, the National Academy of Medicine (NAM) published *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril,* which highlighted the potential for AI to disrupt and transform health care, presenting a new range of possibilities in which it might augment human capacity and improve health (Matheny et al., 2022). In that same publication, the authors acknowledged the potential for AI to introduce significant risks to equity, safety, and privacy, and called for strategies to balance these risks with anticipated benefits.

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 practice. AI methods are being employed in a variety of applications including screening for and detecting disease, predicting real-time clinical outcomes, personalizing treatment, improving patient engagement, streamlining administrative tasks, easing the burden of clinical documentation, and shortening the timeline for therapeutic drug discovery (Ardila et al., 2019; Cai et al., 2015; Glover et al., 2022; Gulshan et al., 2016; Rajpurkar et al., 2022). Some AI applications can also be harnessed to assist with human-like tasks and provide input for human decision making.

However, despite the appearance of data-driven objectivity, AI outputs are built on datasets and models created and influenced by humans and may also result in harms including implicit and/or explicit bias, notably for individuals from underrepresented groups (Obermeyer et al., 2019). Without adequate recognition and redress of these risks, health AI has the potential to exacerbate existing inequities and create new ones (Christensen et al., 2021). As noted in *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril,* additional risks requiring attention include misdiagnosis, overuse of resources, privacy breaches, and workforce displacement or inattention based on over-reliance on AI (Matheny et al., 2022). In addition, by design, some AI models consume new data, learn, and evolve over time, which can contribute to user confusion in understanding their underlying logic.

To realize the benefits and mitigate the risks associated with AI, numerous entities—including an assortment of intergovernmental agencies (The White House, 2022), private enterprises (Google AI, n.d.), and academic communities (MITRE, 2023)—have issued frameworks to guide responsible use. For example, in 2017, the Asilomar AI Principles advanced a set of 23 guidelines covering priorities for research, considerations for ongoing safety, and the importance of value alignment in the development and use of AI (Future of Life Institute, 2017). While these frameworks address critical components of responsible AI, a need remains for convergence or alignment to achieve a socio-technical approach to governance and interoperability (Dorr et al., 2023). A more detailed gap analysis appears in the section of this commentary titled "Landscape Review Gaps and Opportunities." Additionally, the granularity of many of these frameworks could make adoption and adaptation more difficult over time. For example, many previously published AI frameworks identify the criticality of explainability and reproducibility as essential elements for responsible AI—however, the very nature of LLMs and other advanced AI methods makes the practicality of applying these principles challenging.

To promote trustworthy AI in health, health care, and biomedical science, and in the spirit of co-creation and consensus building, this commentary offers for comment and discussion a draft framework for an AI Code of Conduct. The draft AI Code of Conduct framework is comprised of a harmonized set of principles (the Code Principles) grounded in antecedent work and a distilled set of simple rules (the Code Commitments) for broad adoption by the key stakeholders of the AI life cycle: developers, researchers, health systems, payers, patients, and federal agencies. The Code Principles provide touchstones around which health AI governance—facilitative and precautionary—can be shaped, tested, validated, and continually improved as technology, governance capability, and insights advance. Along with these principles, a small set of simple rules emerging from Steering Committee discussions reflect a process consistent with complex adaptive systems (CAS) theory. CAS explores how health systems with many individual interacting components adapt and organize over time and states that a small set of agreed-upon rules guiding individual behavior can create desired outcomes at the system level. See *Box 1* and Appendix B of *Crossing the Quality Chasm: A New Health System for the 21st Century* for more detail on CAS (IOM, 2001).

By applying rules that are widely acceptable and broadly incontrovertible, stakeholders in the AI life cycle will be equipped with the awareness and guidance required to make responsible decisions in a changing environment in real time. These simple rules—titled the Code Commitments—build from the core principles of a LHS and are intended to broadly direct the application and evaluation of the Code Principles in practice. When applied by various stakeholder groups, the Commitments will serve as the foundation for translation into tailored implementation plans, serving to accelerate collective progress toward the safe, effective, ethical, equitable, reliable, and responsible use of AI in health, health care, and biomedical science.

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.

BOX 2 | Code Principles
Applying the Trust Framework of the Learning Health System Core Principles

Engaged: Understanding, expressing, and prioritizing the needs, preferences, goals of people, and the related implications throughout the AI life cycle.
Safe: Attendance to and continuous vigilance for potentially harmful consequences from the application of AI in health and medicine for individuals and population groups.
Effective: Application proven to achieve the intended improvement in personal health and the human condition, in the context of established ethical principles.
Equitable: Application accompanied by proof of appropriate steps to ensure fair and unbiased development and access to AI-associated benefits and risk mitigation measures.
Efficient: Development and use of AI associated with reduced costs for health gained, in addition to a reduction, or at least neutral state, of adverse impacts on the natural environment.
Accessible: Ensuring that seamless stakeholder access and engagement is a core feature of each phase of the AI life cycle and governance.
Transparent: Provision of open, accessible, and understandable information on component AI elements, performance, and their associated outcomes.
Accountable: Identifiable and measurable actions taken in the development and use of AI, with clear documentation of benefits, and clear accountability for potentially adverse consequences.
Secure: Validated procedures to ensure privacy and security, as health data sources are better positioned as a fully protected core utility for the common good, including use of AI for continuous learning and improvement.
Adaptive: Assurance that the accountability framework will deliver ongoing information on the results of AI application, for use as required for continuous learning and improvement in health, health care, biomedical science, and, ultimately, the human condition.

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 are of particular importance as they highlight the need for clear, intentional action between and among various stakeholders comprising the interstitium, or connective tissue that unify a system in pursuit of a shared vision.

First, *inclusive collaboration*. Multistakeholder engagement across the life cycle of problem identification, AI model development and deployment, post-implementation vigilance, and ongoing governance is essential. The perspectives of individuals across organizations, sectors, and roles in the process, as well as across socioeconomic groups, should be included at different points in the AI life cycle. Broad involvement of impacted parties will ensure that the right problem is being solved for the right beneficiary, appropriate data is used and properly stewarded, the model is achieving its stated goals without introducing harmful bias, tools are incorporated into the workflow effectively and transparently, AI users and subjects are educated, models are monitored after implementation, and accountabilities are clear to all involved. The perspectives of patients, providers, developers, and regulators are just a sample of the inputs required to ensure that AI performs as expected, rather than exacerbates existing or creates new inequities in health, health care, and biomedical science. For example, unchecked and unintentional implicit developer bias can lead to discriminatory algorithm results. Though the importance of fair and unbiased AI receives adequate mention in the surveyed publications, the editors of this publication observed limited acknowledgement of the linkages between broad collaboration, inclusive design, and substantively less discriminatory outputs.

Second, *ongoing safety assessment*. The trajectory of AI development in health care, particularly that of LLMs, has outpaced the existing regulatory safety infrastructure (Meskó and Topol, 2023). Unlike other physical medical devices or some software as a medical device, which are regulated by the Food and Drug Administration, some emerging forms of AI are being designed to learn and adapt over time, meaning that a tool approved after testing in one environment could achieve different results at a different time or in a different environment. Considering the implications of and planning for adaptive AI, before it is more widely deployed, seems prudent. Additionally, regardless of AI model type, population, behavior, or technology, changes over time could result in model drift or less accurate outputs. Left unchecked, biomedical AI implementations could not only further entrench existing medical inequities, but inadvertently give rise to new macro-level social problems—e.g., the monopolization of health-related industries as a function of diminishing market competition and reductions in health care workers' collective bargaining power (Allianz Research, 2023; California Nurses

BOX 3 | Proposed Code Commitments

1. Focus: Protect and advance human health and human connection as the primary aims.
2. Benefits: Ensure equitable distribution of benefit and risk for all.
3. Involvement: Engage people as partners with agency in every stage of the life cycle.
4. Workforce well-being: Renew the moral well-being and sense of shared purpose to the health care workforce.
5. Monitoring: Monitor and openly and comprehensibly share methods and evidence of AI's performance and impact on health and safety.
6. Innovation: Innovate, adopt, collaboratively learn, continuously improve, and advance the standard of clinical practice.

The goal is that all decisions associated with, and actions taken, to develop and deploy AI in the health sector will be consistent with these Commitments to develop and foster trust.

Association/National Nurses United, 2023; Qiu and Zhanhong, 2023). The federal government is highly engaged in addressing risks associated with AI, including a recent executive order that calls for federal agencies to identify a chief artificial intelligence officer to ensure safe, secure, and trustworthy AI use within their agency, as well as requiring vendors to share safety test results (The White House, 2023). However, substantially less attention has been given to the need for a "safety culture" for the development and deployment of AI, which would address "individual and group values, attitudes, perceptions, competencies and patterns of behavior that determine the commitment to, and the style and proficiency of, an organization's health and safety management" (ACSNI, 1993, p.23). While regulation enshrines best practice requirements and establishes consequences for malfeasance, a culture of safety lays a foundation of ideas and principles upon which to develop forward-looking initiatives (Manheim, 2023).

Third, *efficiency or environmental protection*. Using excessive resources (minerals, water, electricity, etc.) to power AI development presents potential risks to human health, making efficiency and environmental protection an important consideration for responsible AI. AI computing and storage requirements are growing and creating significant energy demands for data centers. According to a 2018 analysis, the information and communication technology sector is projected to exceed 14% of global emissions by 2040, the bulk of which will come from data centers and communication network infrastructure (Belkhir and Elmeligi, 2018; Nordgren, 2022). While some large technology companies are projecting that their data centers will be carbon-free by 2030 (Bangalore, et al, 2023), global emissions will need to be transparently measured to assess progress toward national and international decarbonization goals (International Energy Agency, n.d.). Beyond emissions, the associated environmental impact of the demand for rare elements used in electronic components and other resources such as water, used for cooling data centers, must also be considered. Despite these facts, *none* of the 60 publications included in this paper's literature review substantively addressed the environmental implications of AI development. The imperative to correct this omission is reflected in the Code Principles below.

A universal Code of Conduct, suitable for current needs and adaptable for future risks and opportunities, should address these three gaps at the system and policy levels, thereby safeguarding the ongoing advantages of AI use and fostering innovation.

Draft Code of Conduct Framework: Code Principles and Code Commitments

Mapped onto the NAM's Learning Health System principles, the Code Principles (see *Box 2*), address gaps identified in the literature review conducted by the authors and include internationally harmonized guiding principles that attend to both present and future contexts.

The Code Principles embed the essential values underpinning responsible behavior in AI development, use, and ongoing monitoring. The Code Commitments (see *Box 3*) meanwhile, are intended to support the application of these Principles in practice. Thus, the Commitments are intentionally broad but action- and decision-oriented. The Commitments call on CAS theory as a small set of simple rules for applying the Code Principles to guide behaviors in a complex system. The Commitments promote governance at every level—individual, organizational, community, state, national, and transnational—with an understanding that collaborative governance includes regulation *and* local governance (Price II et al., 2023).

Given AI's constant evolution and the rapidly changing global contexts in which it operates, attention to the AI Code of Conduct Code Principles and Commitments is especially important as work on the AI Code of Conduct framework moves to its next phase. Consultative outreach, public discussion, and feedback on this commentary, the Code Principles, and the Code Commitments are strongly encouraged. Input will be reflected in a subsequent AICC publication.

Next Steps

Over the coming months, the NAM will continue its work, including the following:

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,* May 3. Available at: https://www.allianz.com/en/economic_research/publications/specials_fmo/eurozone-labor-markets.html (accessed October 18, 2023).
3. Ardila, D., A. P. Kiraly, S. Bharadwaj, B. Choi, J. J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, D. P. Naidich, and S. Shetty. 2019. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine* 25(6):954–961. https://doi.org/10.1038/s41591-019-0447-x.
4. Bangalore, S., A. Bhan, A. Del Miglio, P. Sachdeva, V. Sarma, R. Sharma, and B. Srivathsan. 2023. Investing in the rising data center economy. *McKinsey,* January 17. Available at: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/investing-in-the-rising-data-center-economy (accessed March 5, 2024).
5. Belkhir, L., and A. Elmeligi. 2018. Assessing ICT global emissions footprint: Trends to 2040 & recommendations. *Journal of Cleaner Production* 177:448–463. https://doi.org/10.1016/j.jclepro.2017.12.239.
6. California Nurses Association/National Nurses United. 2023. Nurses stand with striking writers, actors in fight against AI. *National Nurses United,* August 28. Available at: https://www.nationalnursesunited.org/press/nurses-stand-striking-writers-and-actors-in-fight-against-ai (accessed February 27, 2024).
7. Cai, X., O. Perez-Concha, E. Coiera, F. Martin-Sanchez, R. Day, D. Roffe, and B. Gallego. 2015. Real-time prediction of mortality, readmission, and length of stay using electronic health record data. *Journal of the American Medical Informatics Association* 23(3):553–561. https://doi.org/10.1093/jamia/ocv110.
8. Christensen, D. M., J. Manley, and J. Resendez. 2021. Medical algorithms are failing communities of color. *Health Affairs Forefront.* https://doi.org/10.1377/forefront.20210903.976632.
9. Dorr, D. A., L. Adams, and P. Embí. 2023. Harnessing the promise of artificial intelligence responsibly. *JAMA* 329(16):1347–1348. https://doi.org/10.1001/jama.2023.2771.
10. Future of Life Institute. 2017. Asilomar AI Principles. *Future of Life Institute Open Letters,* August 11. Available at: https://futureoflife.org/open-letter/ai-principles/ (accessed May 18, 2023).

11. Glover, W. J., Z. Li, and D. Pachamanova. 2022. The AI-enhanced future of health care administrative task management. *NEJM Catalyst.* https://catalyst.nejm.org/doi/full/10.1056/CAT.21.0355
12. Google AI. n.d. *Responsibility: Our Principles.* Available at: https://ai.google/responsibility/principles/ (accessed May 31, 2023).
13. Gulshan, V., L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster. 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316(22):2402–2410. https://doi.org/10.1001/jama.2016.17216.
14. Halligan, M., and A. Zecevic. 2011. Safety culture in healthcare: A review of concepts, dimensions, measures, and progress. *BMJ Quality & Safety* 20:338–343. https://qualitysafety.bmj.com/content/20/4/338.
15. Hern, A. 2016. Stephen Hawking: AI will be 'either best or worst thing' for humanity. *The Guardian,* October 19. Available at: https://www.theguardian.com/science/2016/oct/19/stephen-hawking-ai-best-or-worst-thing-for-humanity-cambridge (accessed February 27, 2024).
16. High-Level Expert Group on AI. 2019. *Ethics guidelines for trustworthy AI: High-level expert group on artificial intelligence.* European Commission: Brussels, Belgium. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed February 27, 2024).
17. Hutson, M. 2022. Taught to the test: AI software clears high hurdles on IQ tests but still makes dumb mistakes. Can better benchmarks help? *Science,* May 5. https://doi.org/10.1126/science.abq7853.
18. Institute of Medicine (IOM). 2000. *To Err Is Human: Building a Safer Health System.* Washington, DC: The National Academies Press. https://doi.org/10.17226/9728.
19. IOM. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, DC: The National Academies Press. https://doi.org/10.17226/10027.
20. IOM. 2007. *The Learning Healthcare System: Workshop Summary.* Washington, DC: The National Academies Press. https://doi.org/10.17226/11903.
21. International Energy Agency. n.d. *Data Centres and Data Transmission Networks.* Available at: https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks (accessed March 5, 2024).
22. Manheim, D. 2023. Building a culture of safety for AI: Perspectives and challenges. *SSRN.* https://doi.org/10.2139/ssrn.4491421.
23. Matheny, M., S. Thadaney Israni, M. Ahmed, and D. Whicher, editors. 2022. *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril.* Washington, DC: National Academy of Medicine.
24. Meskó, B., and E. Topol. 2023. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ digital medicine* 6(1):120. https://doi.org/10.1038/s41746-023-00873-0.
25. MITRE. 2023. CHAI releases recommendations for trustworthy AI in health. *MITRE,* April 4. Available at: https://www.mitre.org/news-insights/news-release/chai-releases-recommendations-trustworthy-ai-health (accessed February 27, 2024).
26. National Academy of Medicine. 2022. *Transforming Human Health: Celebrating 50 Years of Discovery and Progress.* Washington, DC: The National Academies Press. https://doi.org/10.17226/26722.
27. National Institute of Standards and Technology. 2023. *Artificial intelligence risk management framework (AI RMF 1.0).* https://doi.org/10.6028/NIST.AI.100-1.
28. Nordgren, A. 2022. Artificial intelligence and climate change: Ethical issues. *Journal of Information, Communication and Ethics in Society* 21(1):1–15. https://doi.org/10.1108/JICES-11-2021-0106.
29. Obermeyer, Z., B. Powers, C. Vogeli, and S. Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366(6464):447–453. https://doi.org/10.1126/science.aax2342.
30. Organisation for Economic Co-operation and Development. 2023. *Recommendation of the council on artificial intelligence, OCED/LEGAL/0449.* Available at: https://legalinstruments.oecd.org/api/print?ids=648&lang=en (accessed February 27, 2024).
31. Price II, W. N., M. Sendak, S. Balu, and K. Singh. 2023. Enabling collaborative governance of medical AI. *Nature Machine Intelligence* 5(8):821–823. https://doi.org/10.1038/s42256-023-00699-1.
32. Qiu, R., and L. Zhanhong. 2023. AI widens the gap between the rich and the poor. *SHS Web of Conferences* 152:05004. https://doi.org/10.1051/shsconf/202315205004.
33. Rajpurkar, P., E. Chen, O. Banerjee, and E. J. Topol. 2022. AI in health and medicine. *Nature Medicine* 28:31–38. https://doi.org/10.1038/s41591-021-01614-0.
34. Siala, H., and Y. Wang. 2022. SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. *Social Science & Medicine* 296:114782. https://doi.org/10.1016/j.socscimed.2022.114782.
35. The White House. 2022. *Blueprint for an AI bill of rights.* Available at: https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (accessed May 31, 2023).
36. The White House. 2023. Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. *The White House,* October 30. Available at:

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 Chinese University of Hong Kong. **Andrew A. Gonzalez, MD, JD, MPH,** is Associate Director for Data Science at the Regenstrief Institute Center for Health Services Research and Co-Director of the Indiana University School of Medicine's Center for Surgical Outcomes & Quality Improvement Center.

Acknowledgments

Peter Lee, Microsoft; **Kenneth Mandl,** Harvard University; **Brian Anderson,** MITRE/CHAI; **Timothy Hsu,** AFDO/RAPS Healthcare Products Collaborative; **Mark Sendak** and **Suresh Balu,** Duke University; **Greg Singleton,** HHS; **Signe Braafladt,** Indiana University School of Medicine; **Tyler Philip Robinson,** Indiana University School of Medicine; and **Sunita Krishnan** and **Stephanie Stan,** National Academy of Medicine all provided valuable feedback and support for this paper.

Conflict-of-Interest Disclosures

None to disclose.

Correspondence

The Steering Committee invites and encourages widespread comment and suggestions on the draft Code of Conduct framework using the following link: https://survey.alchemer.com/s3/7767528/NAM-Leadership-Consortium-AICC-Commentary-Paper-Public-Comment

Questions should be directed to Laura Adams at ladams@nas.edu.

Sponsors

This work was conducted with the support of the California Healthcare Foundation, Epic, Gordon and Betty Moore Foundation, National Institutes of Health, and Patrick J. McGovern Foundation.

Disclaimer

The views expressed in this paper are those of the editors and not necessarily of the editors' organizations, the National Academy of Medicine (NAM), or the National Academies of Sciences, Engineering, and Medicine (the National Academies). The paper is intended to help inform and stimulate discussion. It is not a report of the NAM or the National Academies.