Advancing Diagnostic Excellence: From Artificial Intelligence to Patient Impact

Grace Cordovano, PhD

Philip R.O. Payne, PhD

NAM Leadership Consortium
Collaboration for a Learning Health System

Today’s Discussion

1. **AI and healthcare**
   - Opportunities and challenges associated with the expanded use of AI in healthcare
2. **Towards responsible healthcare AI development and use**
   - The need for shared values and behaviors in the context of healthcare AI and the role of the NAM Healthcare AI Code of Conduct
3. **Connecting “the dots”**
   - The intersection of AI and diagnostic excellence
4. **Focusing on humans**
   - Anchoring on the humanistic needs and impact of healthcare AI

AI and Healthcare

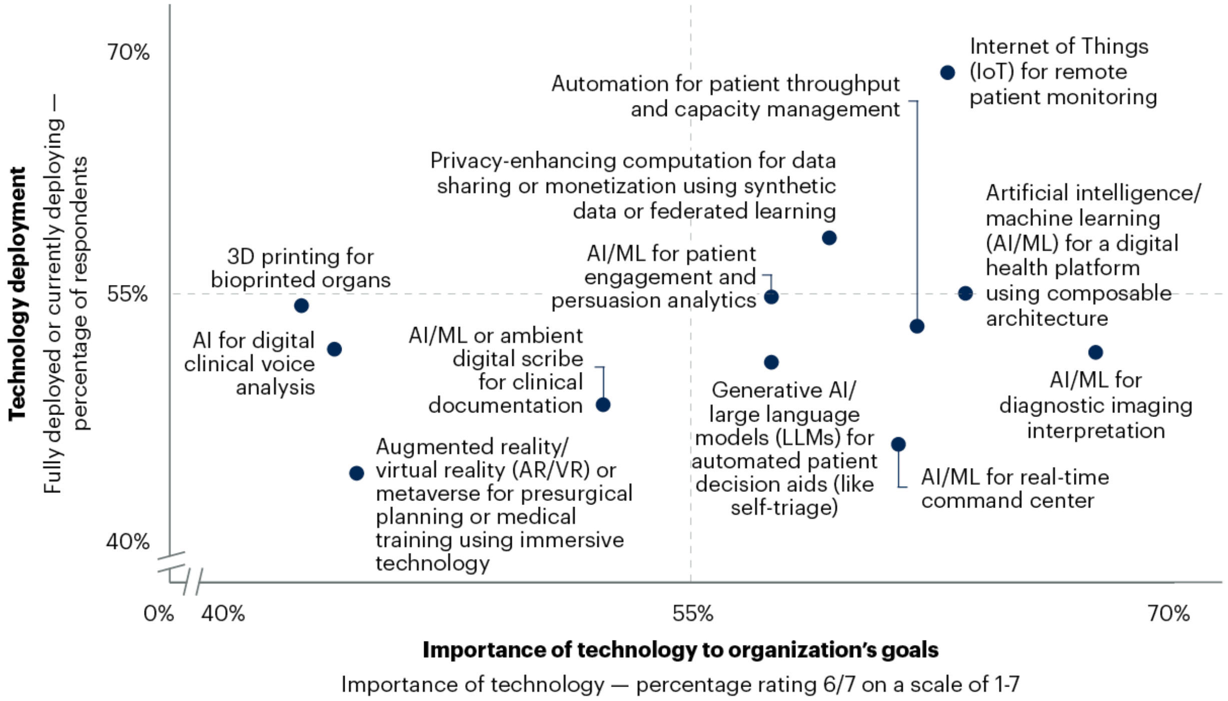
Opportunities and challenges associated with the expanded use of AI in healthcare

AI is Dominating the Healthcare Landscape

- **Patient Monitoring and Care:** AI-driven wearables and predictive analytics enable early detection and personalized treatment.
- **Clinical Decision Support:** AI enhances diagnostics, predicts disease progression, and recommends treatments using medical images and EHRs.
- **Precision Medicine and Drug Discovery:** AI identifies biomarkers, personalizes treatments, and accelerates drug discovery.
- **Medical Education and Innovation:** AI powers virtual simulations, automates data analysis, and accelerates literature reviews.
- **Operational Efficiency:** AI automates medical coding, scheduling, and documentation, reducing clinician workload and optimizing revenue cycle relationships.

Importance vs. Deployment of Technology Use Cases

Percent of healthcare provider respondents



n = 66-69 technology and business leaders at healthcare providers (importance); 63-68 technology and business leaders at healthcare providers (deployment)

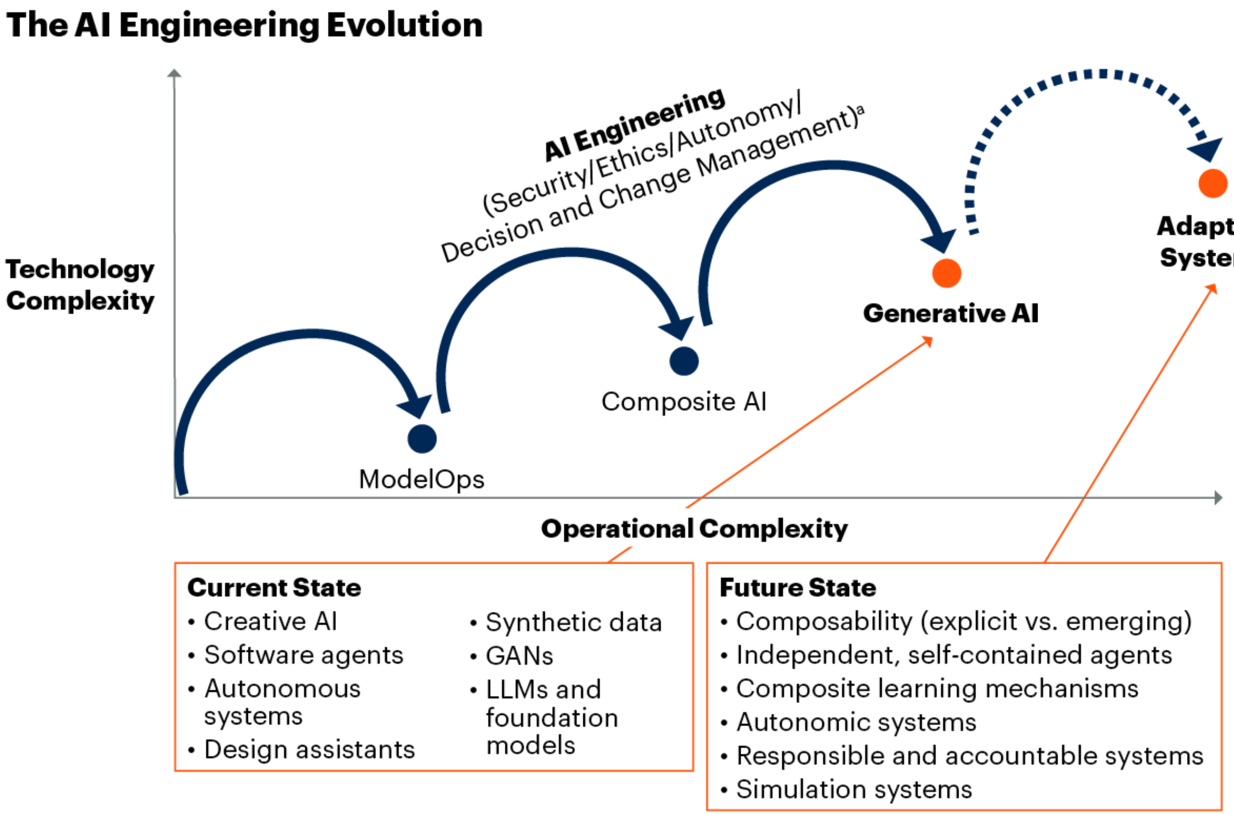
Q: How important is each of these to meeting your enterprise's overall goals and objectives?

Q: Please choose whether you have deployed, plan to deploy or have no interest in deploying these technologies coupled with the use cases.

Source: 2023 Gartner Business Outcomes of Technology Survey

803181_C

Enabled by Step-Changes in Technology



Source: Gartner

[a] Continued list of examples: transparency, decision engineering, skills and organization, design patterns, infrastructure and smart processing foundation.

793932_C

***Open-AI*: Care team and AI collaborating to treat a patient**

NEJM AI 2024; 1 (3)
DOI: 10.1056/AIp2400036

PERSPECTIVE

To Do No Harm — and the Most Good — with AI in Health Care

Carey Beth Goldberg,[1] Laura Adams, R.N., M.S.,[2] David Blumenthal, M.D., M.P.P.,[3] Patricia Flatley Brennan, R.N., Ph.D.,[4,5] Noah Brown,[6] Atul J. Butte, M.D., Ph.D.,[7] Morgan Cheatham,[8] Dave deBronkart,[9,10] Jennifer Dixon, M.D., Ph.D.,[11] Jeff Drazen, M.D.,[12] Barbara J. Evans, Ph.D., J.D., L.L.M.,[13] Sara M. Hoffman,[6] Chris Holmes, Ph.D.,[14,15] Peter Lee, Ph.D.,[16] Arjun Kumar Manrai, Ph.D.,[6,12] Gilbert S. Omenn, M.D., Ph.D.,[17] Jonathan B. Perlin, M.D., Ph.D.,[18] Rachel Ramoni, D.M.D., Sc.D.,[19] Guillermo Sapiro, M.S., Ph.D.,[20,21] Rupa Sarkar, Ph.D.,[22] Harpreet Sood, M.B.B.S., M.P.H., M.R.C.G.P., F.F.C.I.,[23,24] Effy Vayena, Ph.D.,[25] Isaac S. Kohane, M.D., Ph.D.,[6] for the RAISE Consortium*

Received: January 10, 2024; Accepted: January 11, 2024; Published: February 22, 2024

NEJM AI

PERSPECTIVE

A Call for Artificial Intelligence Implementation Science Centers to Evaluate Clinical Effectiveness

Authors: Christopher A. Longhurst, M.D., M.S., Karandeep Singh, M.D., M.M.Sc., Aneesh Chopra, M.P.P., Ashish Atreja, M.D., M.P.H., and John S. Brownstein, Ph.D. Author Info & Affiliations

Published July 10, 2024 | NEJM AI 2024;1(8) | DOI: 10.1056/AIp2400223 | VOL. 1 NO. 8

Annals of Internal Medicine

IDEAS AND OPINIONS

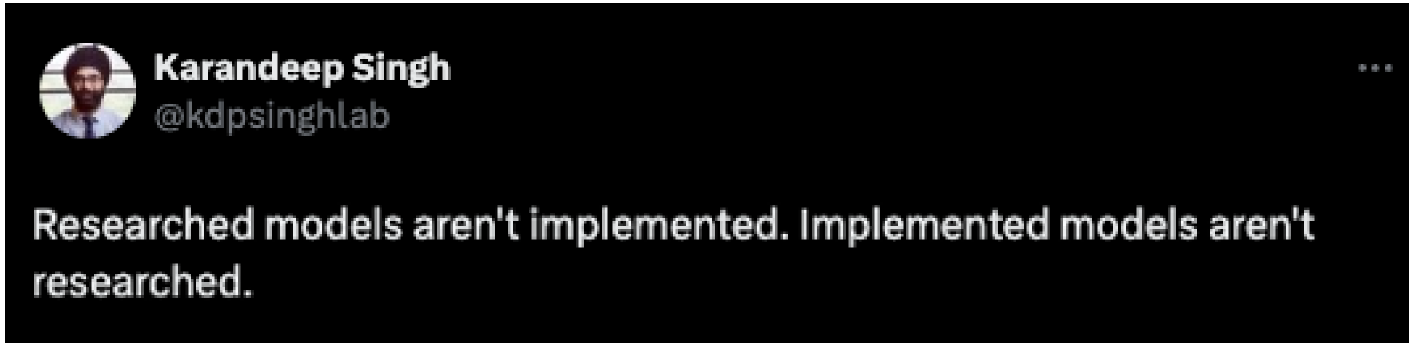
Meeting the Artificial Intelligence Needs of U.S. Health Systems

Patrick G. Lyons, MD, MSc; David A. Dorr, MD, MS; Genevieve B. Melton, MD, PhD; Karandeep Singh, MD, MMSc; and Philip R.O. Payne, PhD

Artificial intelligence (AI) is being ad
many ubiquitous problems facin
health systems (1). Concurrent with im
capabilities largely demonstrated outsi
(for example, prediction, transcription,
empirical evidence for AI improving h
comes is growing (2). For these rese
translate into better care quality, heal
to have in place the right structural ele
esses to evaluate, implement, and mo
gies, as well as mechanisms to meas
related health outcomes (**Table**) (3). Cli
systems must therefore prepare for A
clinical care and operations with some
Structural requirements for effect
AI into clinical care include organization
and technical infrastructure, and individuals with clinical
AI expertise. First, organizations should recognize

local, regional, and national impacts on equity and
care continuity.

- **Strategic AI Integration** – AI adoption is accelerating in healthcare; success requires strong organizational strategies and infrastructure.
- **Governance and Infrastructure** – Establish governance for safety, fairness, and accountability; modernize data systems for AI compatibility.
- **Workforce and Leadership** – Train AI-literate clinicians and engineers; appoint AI leaders like the *Chief Health AI Officer* to guide implementation.
- **Evaluation and Outcomes** – Implement structured assessments to measure AI's impact on care quality and mitigate risks.

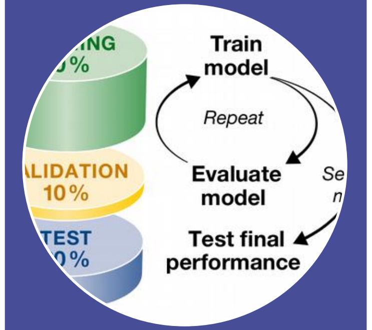


“Researched models are *rarely* implemented; Implemented models are *rarely* researched.”

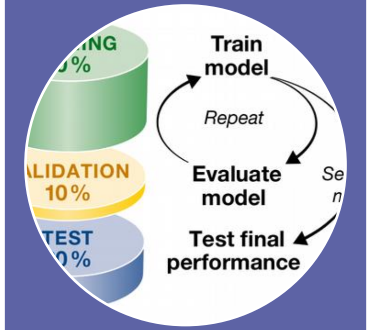
The Complex (and Lengthy) Healthcare AI “Lifecycle”

Needs assessment

Model development



Model validation (internal)



Model validation (external)

Model implementation in the “real world” systems

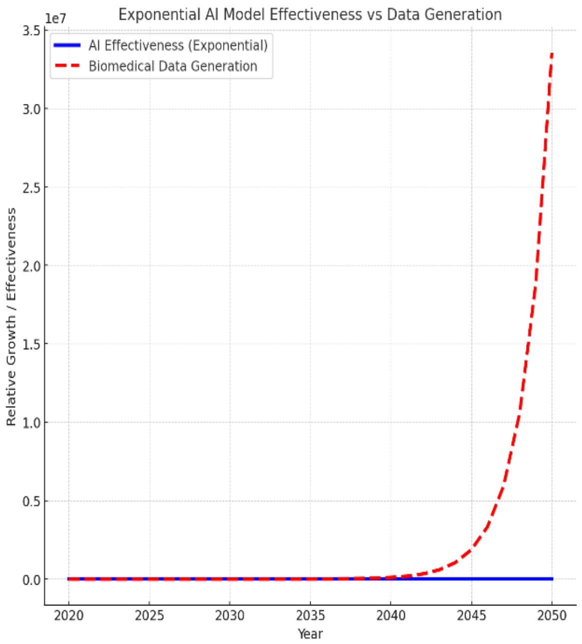
The “last mile” problem (workflow, sociotechnical considerations)

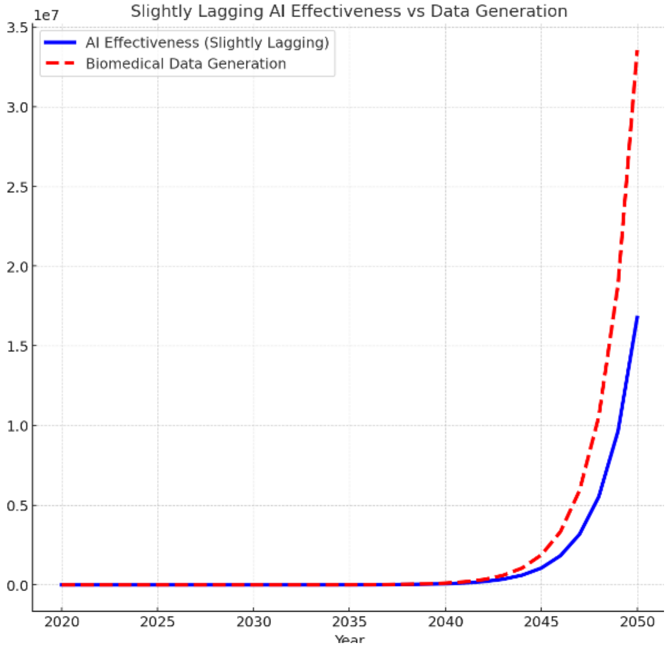
Iterative Cycles

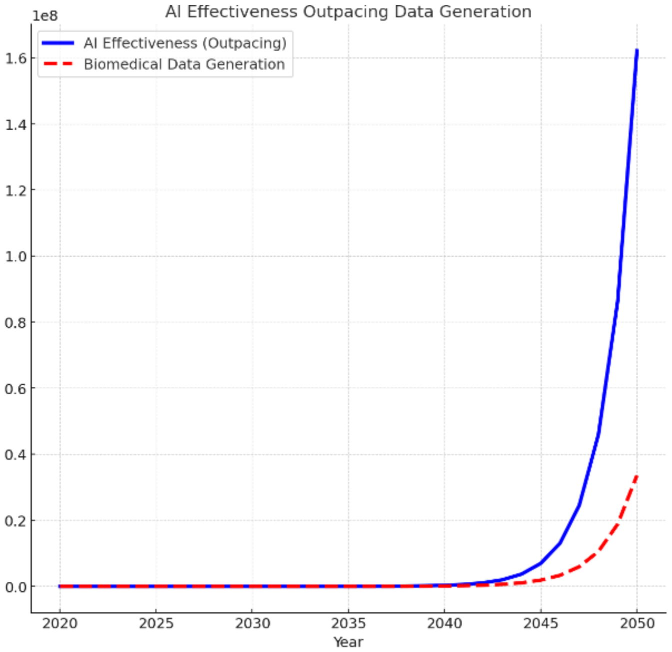
Biomedicine is a complex system, and therefore, our approach to AI must be aligned with this complexity

- **Interdependent Components:** Biomedicine involves a vast network of stakeholders, including researchers, clinicians, patients, payers, regulatory bodies, and technology vendors, whose behaviors and decisions influence outcomes.
- **Non-Linear Interactions:** Small changes in one area (e.g., a discovery or clinical guideline) can have unpredictable, widespread effects on ensuing hypotheses, patient outcomes, provider behaviors, and operational workflows.
- **Adaptive Behaviors:** Biomedical systems must constantly adapt to new knowledge, emerging diseases, evolving patient needs, and technological advancements.
- **Unpredictable Outcomes:** Even with knowledge-based systems, experimental results and health outcomes can vary due to biological, human, and environmental factors.
- **Diverse Expertise:** Effective research and care delivery require collaboration across specialties and disciplines.

Effective Altruism vs. Accelerationism in AI Adoption







Scenario 1:

- Data growth outpaces model effectiveness
- Growth in model effectiveness plateaus as a function of computation and/or human constraints (potentially by design)

Scenario 2:

- Data growth and model effectiveness grow at comparable paces
- Growth in model effectiveness is less constrained by computation and/or human constraints

Scenario 3:

- Model effectiveness grows at a more rapid rate than data
- This scenario would require the identification of new data sources for training

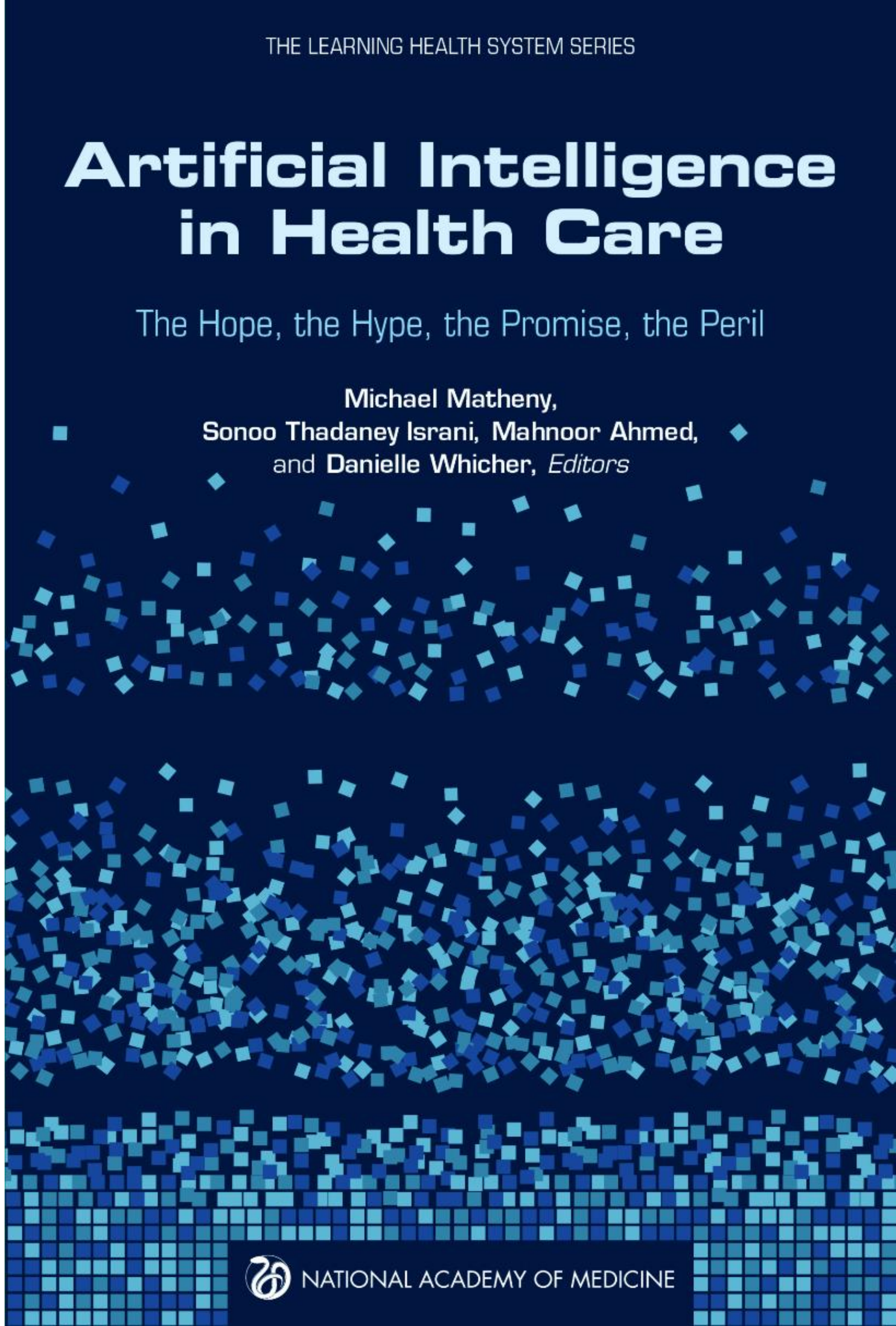
Towards Responsible Healthcare AI Development and Use

The need for shared values and behaviors in the context of healthcare AI and the role of the NAM Healthcare AI Code of Conduct

The NAM Healthcare AI Code of Conduct Initiative

- The goal of the initiative is to advance health care AI “governance interoperability” via a broadly supported Code of Conduct comprised of a harmonized set of principles and commitments
- This will require a description of the relationships and accountabilities of key stakeholders to each other:
  - Translating the Code of Conduct into clearly defined and observable behaviors, and
  - Advancing a national connected “interstitium” that promotes a systems—not siloed—approach
- The NAM endeavors to create a national systems view of the elements required to support responsible AI in health care and biomedical science, and assure that the benefits are equitably distributed

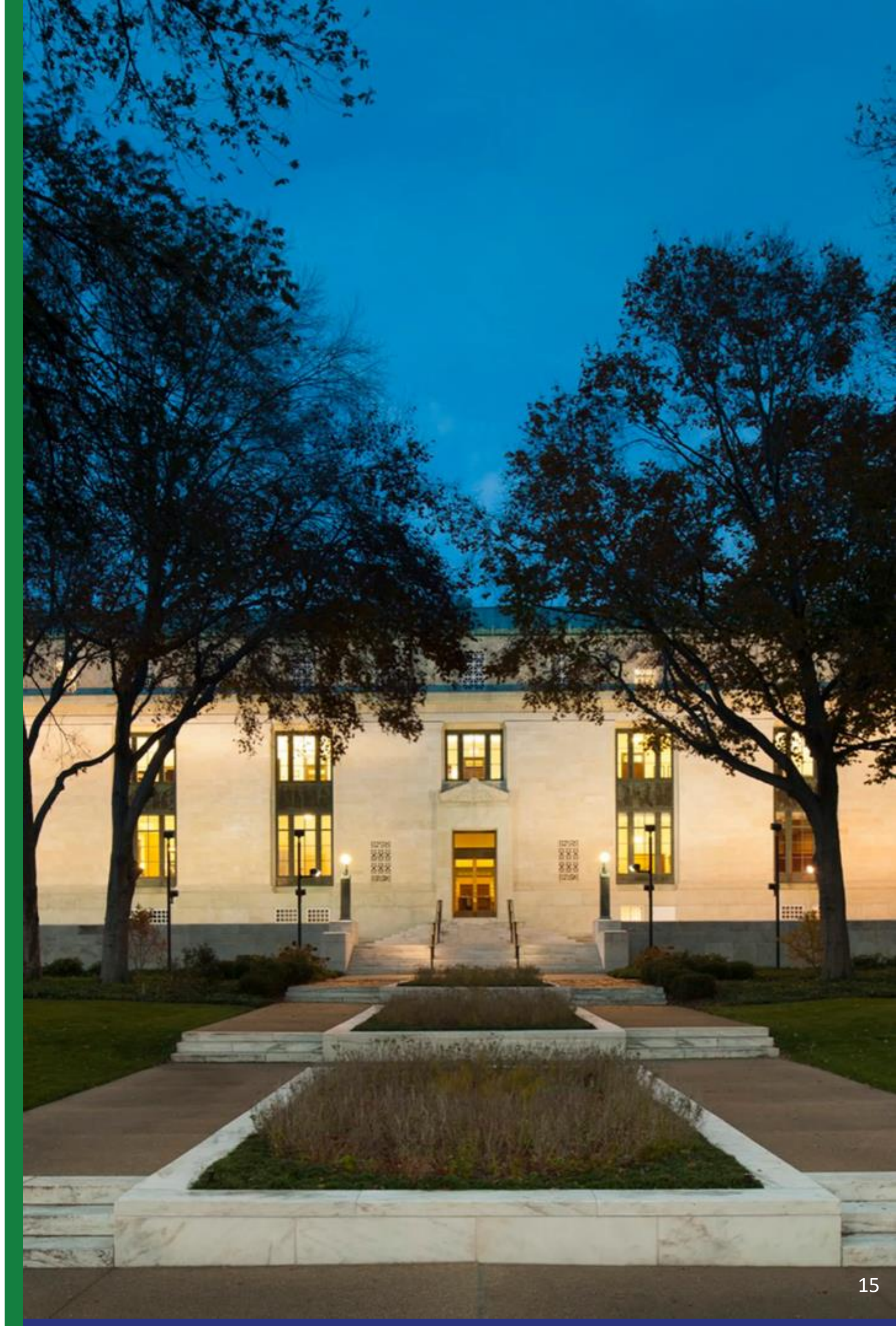
NAM Leadership Consortium
Collaboration for a Learning Health System



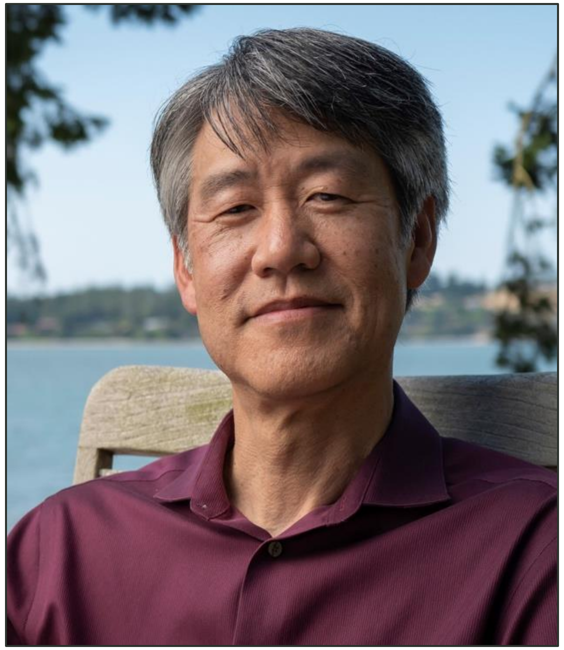
Sponsors of the NAM Healthcare AICC Initiative

- Gordon and Betty Moore Foundation (GBMF)
- Patrick J. McGovern Foundation (PJMF)
- California Health Care Foundation (CHCF)
- Epic Corporation (EPIC)
- National Institutes of Health (NIH)

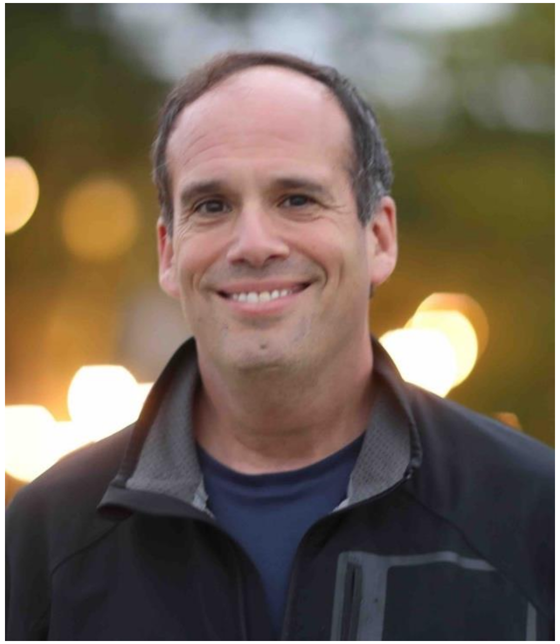
Shown in order of receipt of funding



Digital Health Action Collaborative Co-chairs



Peter Lee
Microsoft



Kenneth Mandl
Harvard

Healthcare AI Code of Conduct (AICC) Steering Committee Co-Chairs

Gianrico Farrugia,
CEO
Mayo Clinic

Bakul Patel,
Global Lead Digital Health Strategy
Google

Roy Jakobs, CEO
Royal Phillips

Healthcare AICC National Steering Committee Members

Gianrico Farrugia, President & CEO, Mayo Clinic – **Co-Chair**	**Kevin Johnson,** Professor, University of Pennsylvania
Bakul Patel, Global Lead, Digital Health Strategy, Google – **Co-Chair**	**Peter Lee,** Microsoft Research
Roy Jakobs, CEO, Philips – **Co-Chair**	**Kenneth Mandl,** Harvard Medical School
Andrew Bindman, Chief Medical Officer, Kaiser Permanente	**Kedar Mate,** President & CEO, Institute for Healthcare Improvement
Grace Cordovano, Enlightening Results	**Deven McGraw,** Co-Founder, Chief Regulatory Office, Ciitizen
Jodi Daniel, Partner, Crowell & Moring Health Group	**Philip Payne,** Chief Health AI Officer, WashU Medicine and BJC Healthcare
Wyatt Decker, EVP, and Chief Physician, UnitedHealth Group	**Vardit Ravitsky,** President, The Hastings Center
Peter Embi, Professor, Vanderbilt University Medical Center	**Suchi Saria,** Founder and CEO, Bayesian Health,
Kadija Ferryman, Core Faculty, Johns Hopkins University	**Eric Topol,** Founder & Dir., Scripps Research Translational Institute
Sanjay Gupta, Chief Medical Correspondent, CNN	**Selwyn Vickers,** CEO, Memorial Sloan Kettering Cancer Center
Eric Horvitz, Chief Scientific Officer, Microsoft	

Responding to Complexity in Health AI Adoption: Encoding Behaviors Via a Code of Conduct

- The term “code of conduct” was intentionally chosen for the NAM AICC initiative, reflecting a belief that shared behavioral expectations are essential in navigating the rapid technological changes in healthcare AI.

> **“A statement setting out guidelines regarding the ethical principles and standards of behavior expected of a professional person or (organization).”** (Oxford Reference Dictionary)

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

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,** Citizen; **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.

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

"Among the 60 publications reviewed, 3 areas of inconsistency were identified: inclusive collaboration, ongoing safety assess- ment, 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 sys- tem in pursuit of a shared vision."

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.

HealthAffairs

Topics Journals Forefront Podcasts

COMMENTARY

Artificial Intelligence In Health And Health Care: Priorities For Action

Michael E. Matheny, Jennifer C. Goldsack, Suchi Saria, Nigam H. Shah, Jacqueline Gerhart, I. Glenn Cohen, W. Nicholson Price II, Bakul Patel, Philip R. O. Payne, Peter J. Embí, ... See all authors

AFFILIATIONS

PUBLISHED: JANUARY 22, 2025 Full Access https://doi.org/10.1377/hlthaff.2024.01003

SECTIONS VIEW ARTICLE PERMISSIONS SHARE TOOLS

Abstract

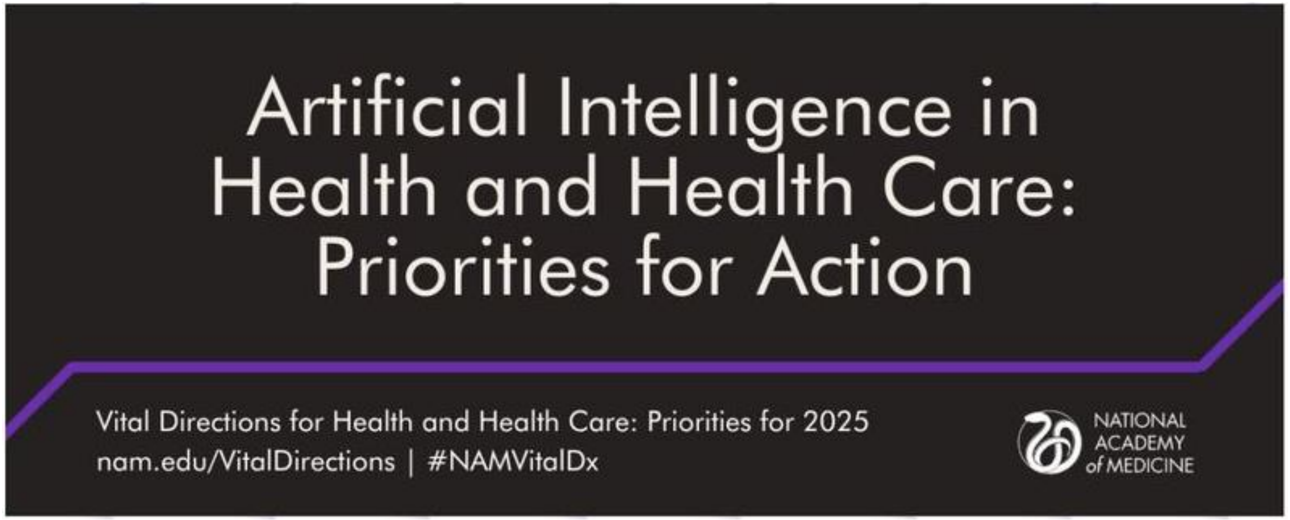
The field of artificial intelligence (AI) has entered a new cycle of intense opportunity, fueled by advances in deep learning, including generative AI. Applications of recent advances affect many aspects of everyday life, yet nowhere is it more important to use this technology safely, effectively, and equitably than in health and health care. Here, as part of the National Academy of Medicine's Vital Directions for Health and Health Care: Priorities for 2025 initiative, which is designed to provide guidance on pressing health care issues for the incoming presidential administration, we describe the steps needed to achieve these goals. We focus on four strategic areas: ensuring safe, effective, and trustworthy use of AI; promotion and development of an AI-competent health care workforce; investing in AI research to support the science, practice, and delivery of health and health care; and promotion of policies and procedures to clarify AI liability and responsibilities.

TOPICS

ARTIFICIAL INTELLIGENCE | TECHNOLOGY | LIABILITY | ORGANIZATION OF CARE | EDUCATION | MEDICAL RESEARCH | HEALTH PROFESSIONALS | DISEASES | ETHICS

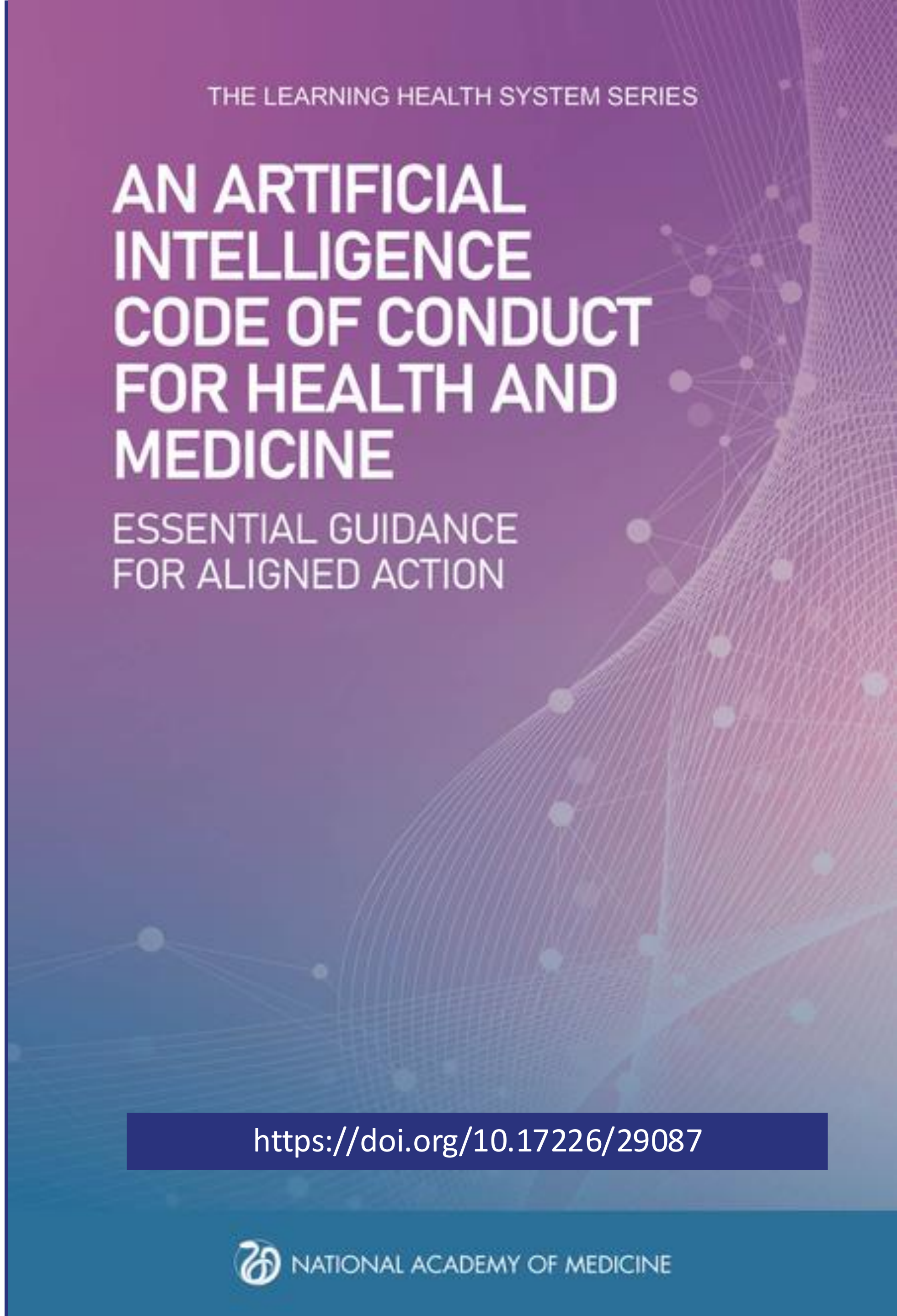
https://doi.org/10.1377/hlthaff.2024.01003

"AI is poised to transform how patients, caregivers, and health care professionals experience the management of their health, health care, and care goals. Substantial challenges remain in realizing this promise. We believe that policies in four key areas can facilitate and accelerate AI in health and health care, including promoting the safe, effective, and trustworthy use of AI; promoting the development of an AI-competent healthcare workforce; focusing investments in key research portfolios; and clarifying AI liability and responsibilities. Through these efforts, substantial societal health improvements may be achieved."



The NAM Health AI Code of Conduct Special Publication

- **Engaged:** Understanding, expressing, and prioritizing the needs, preferences, goals of people, and the related implications throughout the AI lifecycle
- **Safe:** Attendance to and continuous vigilance and controls 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 that results in reductions in resources to achieve improved health outcomes without concurrent adverse impacts on the natural environment
- **Accessible:** Ensuring that seamless stakeholder access and engagement is a core feature of each phase of the AI lifecycle 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 controls and 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



A Call To Action for the BMI Community, Building on the Healthcare AICC

- **Connect with and meaningfully engage in relevant convening and feedback activities**, including those associated with the dissemination of implementation of recommendations generated by the NAM AICC initiative, as well as current and emerging collaboratives
- **Define and propagate a culture of operational safety and continuous learning** in the context of health and healthcare-related AI by focusing on fostering shared behaviors, commitments, and outcomes among all stakeholders
- **Catalyze a community-wide dialogue** concerning both the successes and failures of AI in the clinical domain using an inclusive and transparent approach to communication, ensuring that all stakeholders—regardless of their technical expertise—can engage meaningfully in this discourse
- **Leverage past community-level experience** to address privacy and ethical concerns in the context of health-related AI via a proactive and collaborative approach to developing practical frameworks and policies.
- **Advance basic and applied BMI research agendas** to achieve the promise of contemporary AI in health and healthcare settings
- **Advocate for the adoption and use of knowledge and outcomes** resulting from the preceding items by regulatory bodies
- **Facilitate harmonization across and between recommendations, guidelines, frameworks, standards, and best practices** being produced and disseminated by convening groups, standard-setting bodies, regulators, and aligned organizations

Journal of the American Medical Informatics Association, 2025, **32(2)**, 408–412
https://doi.org/10.1093/jamia/ocae306
Advance access publication 10 December 2024
Perspective

AMIA OXFORD

Perspective

Toward an artificial intelligence code of conduct for health and healthcare: implications for the biomedical informatics community

Philip R.O. Payne, PhD*,[1], Kevin B. Johnson, MD, MS[2], Thomas M. Maddox, MD, MS[3,4], Peter J. Embi, MD, MS[5], Kenneth D. Mandl, MD, MPH[6,7], Deven McGraw, JD, MPH[8], Suchi Saria, PhD[9,10,11,12], Laura Adams, MS[13]

[1]Institute for Informatics, Data Science and Biostatistics, Washington University in St. Louis School of Medicine, St. Louis, MO 63110, United States, [2]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, United States, [3]Healthcare Innovation Lab, BJC HealthCare and Washington University in St. Louis School of Medicine, St. Louis, MO 63110, United States, [4]Division of Cardiology, Department of Medicine, Washington University in St. Louis School of Medicine, St. Louis, MO 63110, United States, [5]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37203, United States, [6]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA 02115, United States, [7]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, United States, [8]Citizen Health, San Francisco, CA 94112, United States, [9]Department of Computer Science, Johns Hopkins University Whiting School of Engineering, Baltimore, MD 21218, United States, [10]Department of Health Policy and Management, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD 21218, United States, [11]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21218, United States, [12]Bayesian Health, New York, NY 10014, United States, [13]National Academy of Medicine, Washington, DC 20001, United States

*Corresponding author: Philip R.O. Payne, PhD, Institute for Informatics, Data Science and Biostatistics, Washington University in St. Louis School of Medicine, 660 S. Euclid Ave, Campus Box 8132, St. Louis, MO 63110, United States (prpayne@wustl.edu)

Abstract

Introduction: The rapid advancement of artificial intelligence (AI) has led to significant transformations in health and healthcare. As AI technologies continue to evolve, there is an urgent need to establish a unified framework that guides the design, implementation, and evaluation of AI-driven interventions across individual and population health contexts.

Approach: In response to this need, the National Academy of Medicine (NAM) has initiated the development of an AI code of conduct (AICC) through its Digital Health Action Collaborative. This code of conduct is grounded in shared principles and commitments, aiming to actualize ethical and effective AI practices within the broader health and healthcare ecosystem. Given its specialized expertise and insight, the biomedical informatics (BMI) community plays a pivotal role in shaping and applying these guidelines.

Recommendations: We, as members of the AICC Steering Committee and the NAM Digital Health Action Collaborative, urge BMI educators, researchers, and practitioners to engage actively in refining and implementing the AICC. This involvement is critical to ensuring that the code is robust, applicable, and continuously improved to meet the evolving challenges facing health and healthcare.

Key words: artificial intelligence; biomedical informatics; policy.

Introduction

Imagine a scenario in which a faculty member at a leading academic biomedical informatics (BMI) program seeks to pilot an AI (artificial intelligence)-based chatbot designed to help families discern when symptoms of congestive heart failure warrant a hospital visit. Despite promising early in silico results, the Institutional Review Board rejects the study proposal, citing ethical concerns over testing AI as a substitute for physician judgment. This raises a critical question: what guiding principles should inform the ethical design and approval of such AI systems, and how can the healthcare community establish a shared framework to navigate these challenges?

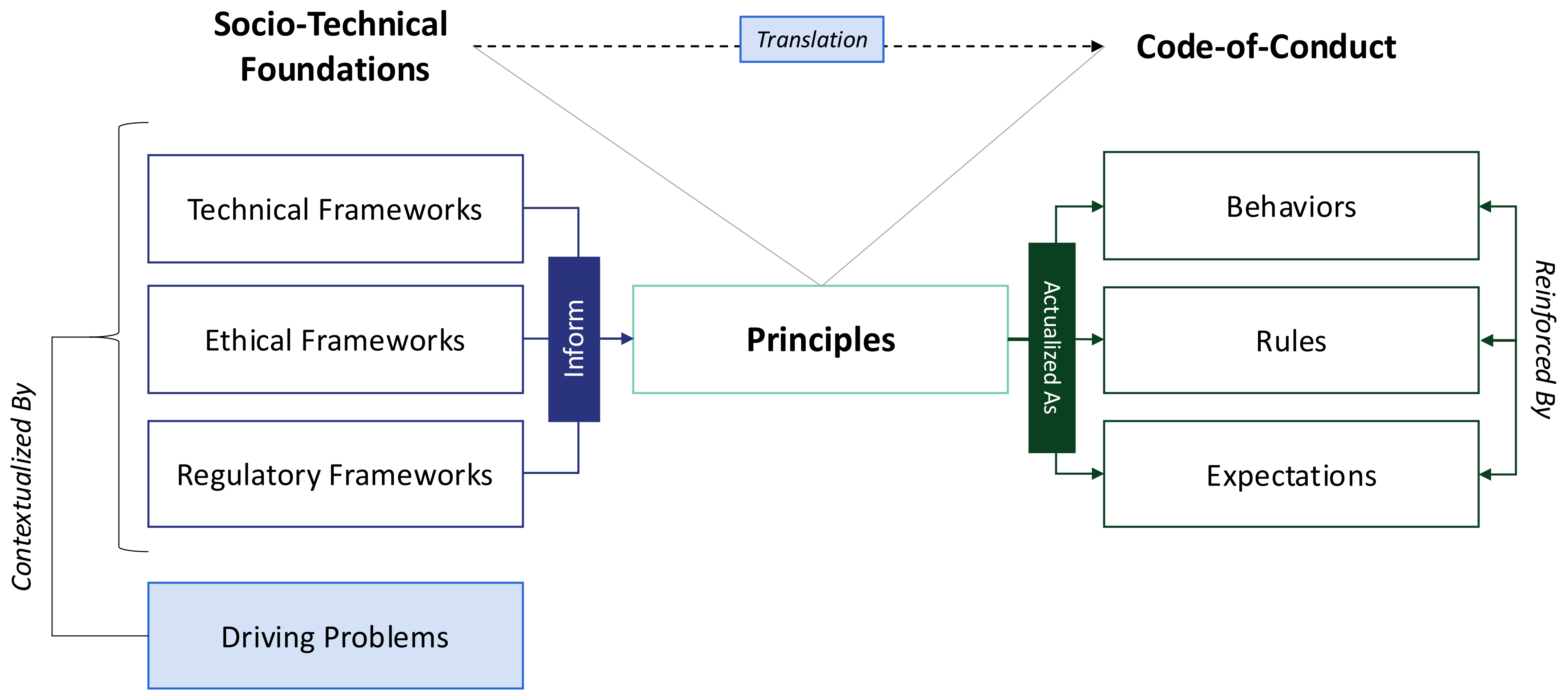
The preceding scenario underscores the pressing need for a comprehensive AI code of conduct (AICC), especially as AI technologies, including commercial generative AI-based platforms, rapidly proliferate. The availability of advanced analytical and generative AI tools offers substantial potential to enhance education, research, clinical practice, population health management, and healthcare operations.[1,2] However, adopting these technologies necessitates rigorous oversight to ensure they are developed, implemented, and evaluated responsibly.

The BMI community is uniquely positioned to lead in this effort. By leveraging its deep understanding of the computational and sociotechnical dimensions of health information systems, the BMI community can shape the development of AI in ways that maximize benefits while minimizing risks. Furthermore, the community's experience with data governance, privacy, and ethical standards provides a strong foundation for guiding the responsible use of AI in healthcare.

Received: Sept

https://doi.org/10.1093/jamia/ocae306

A Conceptual Model for Navigating the Healthcare AI Code of Conduct



Contextualizing the NAM HAICC for BMI

- Given the multidisciplinary nature of AI in healthcare—spanning research, education, clinical care, human-centered design, ethics, product development, implementation science, finance, and policy—a code of conduct that is both inclusive and adaptable is crucial.
- Such a framework can foster coordination and collaboration across diverse domains, ensuring that AI technologies evolve in ways that prioritize patient safety, equity, and continuous improvement.
- A code of conduct based on higher-order, technology-agnostic principles will accommodate the inevitable evolution of AI technologies
- By focusing on foundational principles rather than prescriptive technical standards, the AI Code of Conduct can support the ongoing advancement of AI in health and healthcare, ensuring that innovation is aligned with the best interests of patients and society.

What are the right tasks for AI in healthcare?

- Criteria and requirements for judging the feasibility, actionability, and potential clinical impact of an AI-driven intervention targeting a well-defined driving problem
- Selecting reasonable and appropriate outcome measures to justify the safety, efficacy, and value of AI, contextualized by relevant driving problems
- Ethical standards for applying the preceding dimensions in a consistent and principled manner

1

What are the right data for AI?

- Ethically, responsibly, and objectively sourcing representative data in response to a given driving problem
- Defining, declaring, measuring, and mitigating known and emergent sources of biases in those data
- Ensuring the transparency, reproducibility, and rigor of data engineering methods and protocols used in the AI application
- Adherence to privacy laws impacting the collection, use, and sharing of data for AI and methods to enable transparency to the public regarding the uses of data for AI

2

What is the right evidence standard for AI?

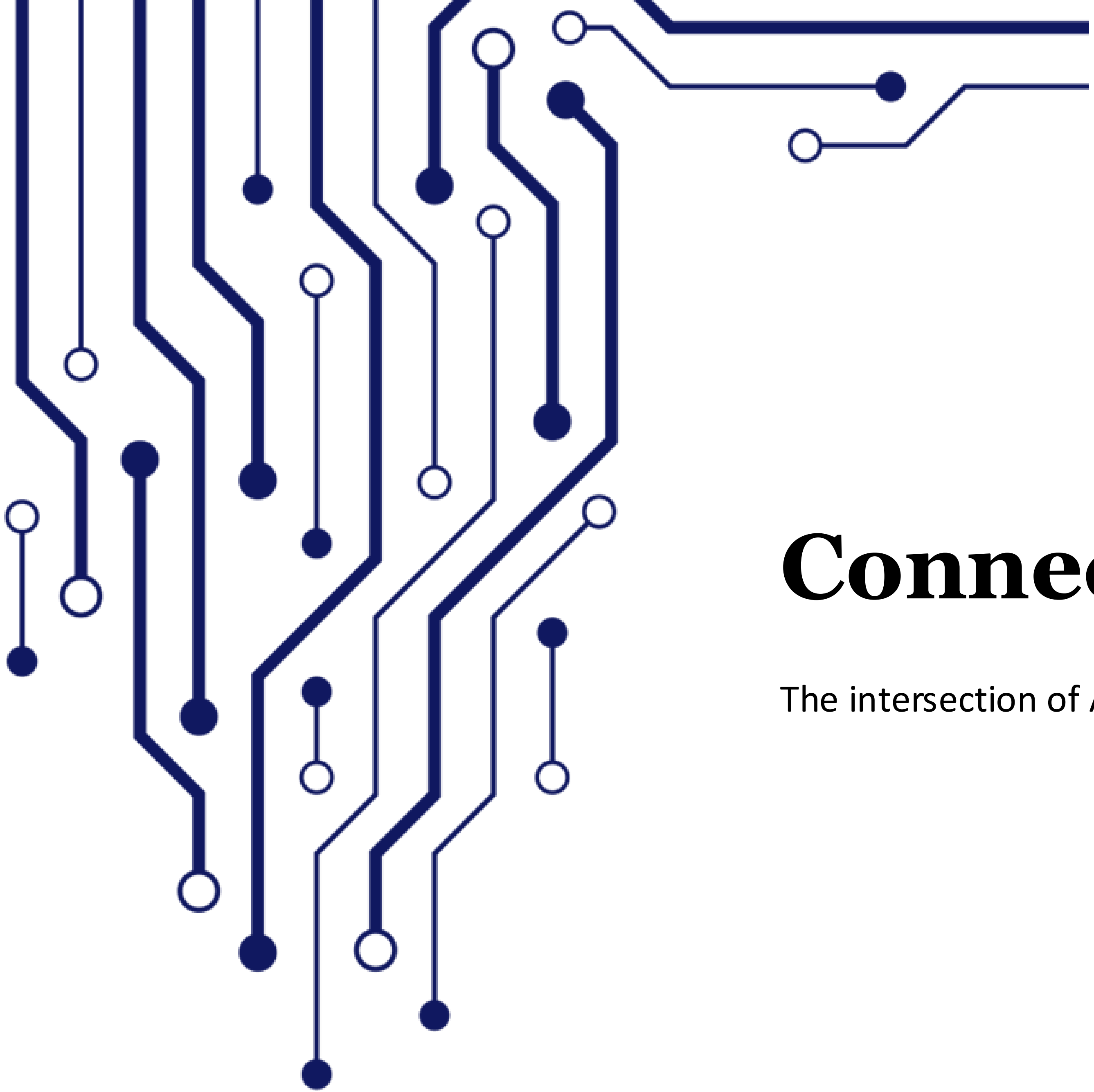
- Selecting, justifying, and applying appropriate measurement methods (eg., “measures”) to measure the efficacy of the AI intervention
- Applying and understanding the impact of AI-driven outcomes given the complex system dynamics of health and healthcare (eg, a systems-thinking approach to performance assessment)
- Selecting and applying optimal evidence reporting guidelines and standards in support of AI-focused research reports (eg, CONSORT-AI, SPIRIT-AI, DECIDE-AI, as well as emerging or future standards such as those being proposed by the VALID-AI)

3

What are the right approaches for integrating AI in clinical care?

- Contextualizing the deployment, use, and evaluation of AI in “real world” settings as a function of human–computer interaction, implementation science, and cognitive science principles
- Establishing safety and outcomes feedback mechanisms and continuous improvement cycles for AI-driven interventions (“algorithmovigilance”)
- Assessing and understanding workflow, cultural, and incentive-based structures that predispose or enable the optimal use of AI in health and healthcare settings
- Assuring clinician and patient understanding of how AI is used in clinical care and leveraging AI to improve clinician–patient shared decision-making regarding clinical interventions

4



Connecting “The Dots”

The intersection of AI and diagnostic excellence

The Intersection of the NAM Health Care AI Code of Condcut and CODEX

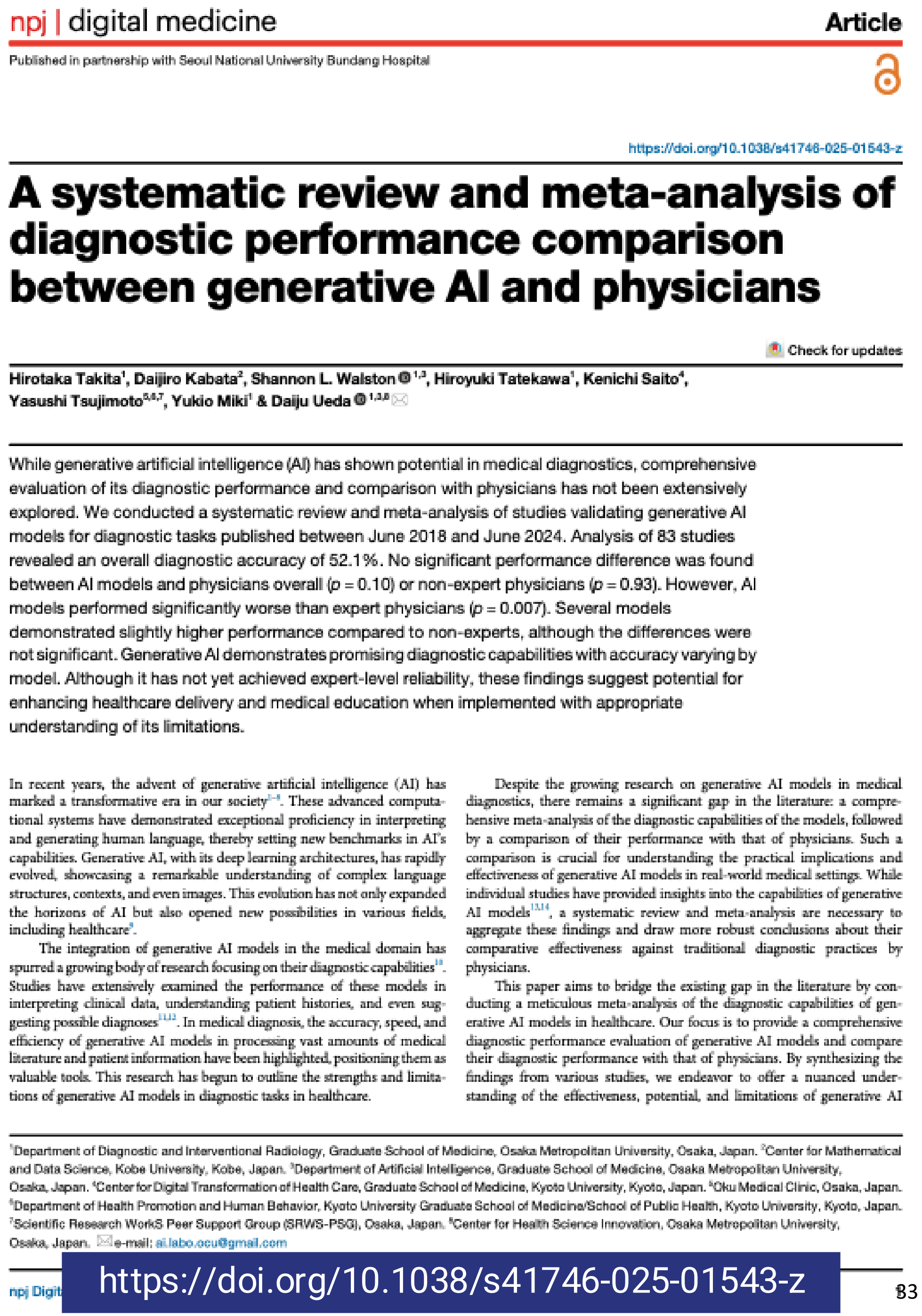
- Diagnostic errors affect nearly every person in their lifetime and account for roughly 800,000 harm events annually in the U.S.
- New evidence highlights that diagnostic errors disproportionately affect people of color and underserved communities.
- The use of AI and analogous digital health interventions in the context of reducing diagnostic error presents new opportunities to support early symptom recognition, assist in clinical reasoning, facilitate shared and patient-centered decision-making, and mitigate diagnostic odysseys.
- The recently released National Academy of Medicine **Health Care Artificial Intelligence Code of Conduct** provides a behavioral framework that bridges the gap between accelerationist and altruist viewpoints regarding responsible AI development, use, and evaluation, enabling a balanced and progressive approach to AI adoption and use in such contexts.

AI in the Context of Diagnostic Excellence

- AI can help accelerate the translation of diagnostic research into health care systems and policy interventions.
- The domains of diagnostic excellence and responsible AI development share a foundational need for collaboration among diverse clinical and technical stakeholders, representing an opportunity for co-creation, shared learning, and best practices.
- Anchoring the application of responsible AI frameworks, such as the National Academy of Medicine's Health Care Artificial Intelligence Code of Conduct, on key outcomes, including diagnostic excellence, provides an opportunity to demonstrate the real-world and patient-level impact of such systems-level approaches.

Substantial Work to Be Done to Replicate Expert Clinician Performance

"While generative artificial intelligence (AI) has shown potential in medical diagnostics, comprehensive evaluation of its diagnostic performance and comparison with physicians has not been extensively explored … No significant performance difference was found between AI models and physicians overall (p = 0.10) or non-expert physicians (p = 0.93). However, AI models performed significantly worse than expert physicians (p = 0.007). Several models demonstrated slightly higher performance compared to non-experts, although the differences were not significant. Generative AI demonstrates promising diagnostic capabilities with accuracy varying by model. Although it has not yet achieved expert-level reliability, these findings suggest potential for enhancing healthcare delivery and medical education when implemented with appropriate understanding of its limitations."


npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-025-01543-z

A systematic review and meta-analysis of diagnostic performance comparison between generative AI and physicians

Check for updates

Hirotaka Takita[1], Daijiro Kabata[2], Shannon L. Walston[1,3], Hiroyuki Tatekawa[1], Kenichi Saito[4], Yasushi Tsujimoto[5,6,7], Yukio Miki[1] & Daiju Ueda[1,3,8]

While generative artificial intelligence (AI) has shown potential in medical diagnostics, comprehensive evaluation of its diagnostic performance and comparison with physicians has not been extensively explored. We conducted a systematic review and meta-analysis of studies validating generative AI models for diagnostic tasks published between June 2018 and June 2024. Analysis of 83 studies revealed an overall diagnostic accuracy of 52.1%. No significant performance difference was found between AI models and physicians overall ($p = 0.10$) or non-expert physicians ($p = 0.93$). However, AI models performed significantly worse than expert physicians ($p = 0.007$). Several models demonstrated slightly higher performance compared to non-experts, although the differences were not significant. Generative AI demonstrates promising diagnostic capabilities with accuracy varying by model. Although it has not yet achieved expert-level reliability, these findings suggest potential for enhancing healthcare delivery and medical education when implemented with appropriate understanding of its limitations.

In recent years, the advent of generative artificial intelligence (AI) has marked a transformative era in our society[1–8]. These advanced computational systems have demonstrated exceptional proficiency in interpreting and generating human language, thereby setting new benchmarks in AI's capabilities. Generative AI, with its deep learning architectures, has rapidly evolved, showcasing a remarkable understanding of complex language structures, contexts, and even images. This evolution has not only expanded the horizons of AI but also opened new possibilities in various fields, including healthcare[9].

The integration of generative AI models in the medical domain has spurred a growing body of research focusing on their diagnostic capabilities[10]. Studies have extensively examined the performance of these models in interpreting clinical data, understanding patient histories, and even suggesting possible diagnoses[11,12]. In medical diagnosis, the accuracy, speed, and efficiency of generative AI models in processing vast amounts of medical literature and patient information have been highlighted, positioning them as valuable tools. This research has begun to outline the strengths and limitations of generative AI models in diagnostic tasks in healthcare.

Despite the growing research on generative AI models in medical diagnostics, there remains a significant gap in the literature: a comprehensive meta-analysis of the diagnostic capabilities of the models, followed by a comparison of their performance with that of physicians. Such a comparison is crucial for understanding the practical implications and effectiveness of generative AI models in real-world medical settings. While individual studies have provided insights into the capabilities of generative AI models[13,14], a systematic review and meta-analysis are necessary to aggregate these findings and draw more robust conclusions about their comparative effectiveness against traditional diagnostic practices by physicians.

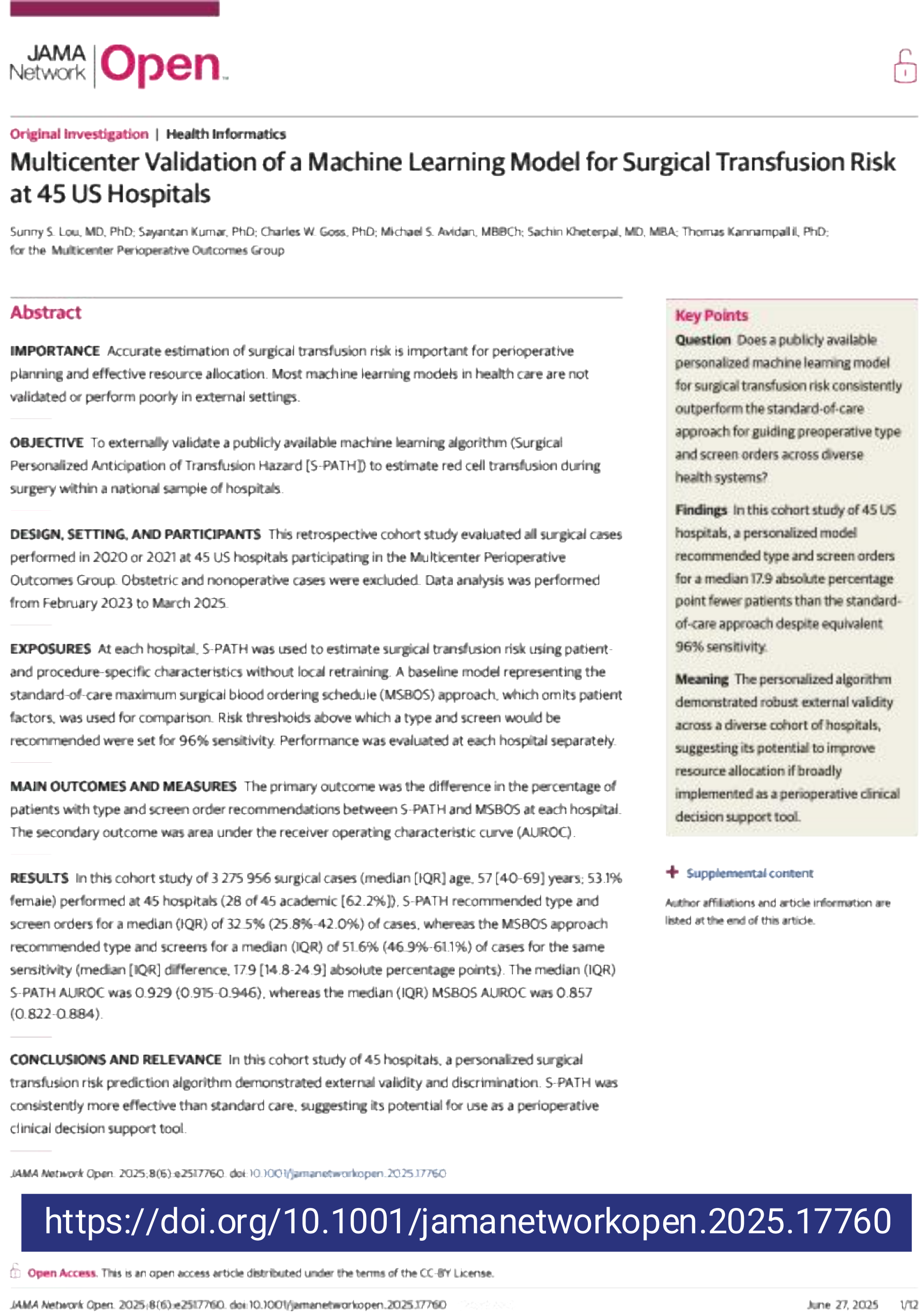
This paper aims to bridge the existing gap in the literature by conducting a meticulous meta-analysis of the diagnostic capabilities of generative AI models in healthcare. Our focus is to provide a comprehensive diagnostic performance evaluation of generative AI models and compare their diagnostic performance with that of physicians. By synthesizing the findings from various studies, we endeavor to offer a nuanced understanding of the effectiveness, potential, and limitations of generative AI

[1]Department of Diagnostic and Interventional Radiology, Graduate School of Medicine, Osaka Metropolitan University, Osaka, Japan. [2]Center for Mathematical and Data Science, Kobe University, Kobe, Japan. [3]Department of Artificial Intelligence, Graduate School of Medicine, Osaka Metropolitan University, Osaka, Japan. [4]Center for Digital Transformation of Health Care, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [5]Oku Medical Clinic, Osaka, Japan. [6]Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto University, Kyoto, Japan. [7]Scientific Research WorkS Peer Support Group (SRWS-PSG), Osaka, Japan. [8]Center for Health Science Innovation, Osaka Metropolitan University, Osaka, Japan. e-mail: ai.labo.ocu@gmail.com

https://doi.org/10.1038/s41746-025-01543-z

But also, Substantial Promise to Improve Diagnostic and Treatment Related Decision-Making

"We evaluated the performance of S-PATH, a personalized surgical transfusion risk prediction model, within a diverse cohort of 45 academic and community US hospitals serving more than 3 million surgical patients. S-PATH consistently outperformed the standard-of-care MSBOS approach at most hospitals, with higher overall discrimination and requiring fewer type and screen orders to adequately detect 96% of patients who subsequently required transfusion. These findings suggest S-PATH's generalizability and robustness and provide evidence to support its potential for pragmatic clinical value in improving resource allocation if implemented broadly."



JAMA Network Open.

Original Investigation | Health Informatics

Multicenter Validation of a Machine Learning Model for Surgical Transfusion Risk at 45 US Hospitals

Sunny S. Lou, MD, PhD; Sayantan Kumar, PhD; Charles W. Goss, PhD; Michael S. Avidan, MBBCh; Sachin Kheterpal, MD, MBA; Thomas Kannampallil, PhD; for the Multicenter Perioperative Outcomes Group

Abstract

IMPORTANCE Accurate estimation of surgical transfusion risk is important for perioperative planning and effective resource allocation. Most machine learning models in health care are not validated or perform poorly in external settings.

OBJECTIVE To externally validate a publicly available machine learning algorithm (Surgical Personalized Anticipation of Transfusion Hazard [S-PATH]) to estimate red cell transfusion during surgery within a national sample of hospitals.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study evaluated all surgical cases performed in 2020 or 2021 at 45 US hospitals participating in the Multicenter Perioperative Outcomes Group. Obstetric and nonoperative cases were excluded. Data analysis was performed from February 2023 to March 2025.

EXPOSURES At each hospital, S-PATH was used to estimate surgical transfusion risk using patient- and procedure-specific characteristics without local retraining. A baseline model representing the standard-of-care maximum surgical blood ordering schedule (MSBOS) approach, which omits patient factors, was used for comparison. Risk thresholds above which a type and screen would be recommended were set for 96% sensitivity. Performance was evaluated at each hospital separately.

MAIN OUTCOMES AND MEASURES The primary outcome was the difference in the percentage of patients with type and screen order recommendations between S-PATH and MSBOS at each hospital. The secondary outcome was area under the receiver operating characteristic curve (AUROC).

RESULTS In this cohort study of 3 275 956 surgical cases (median [IQR] age, 57 [40-69] years; 53.1% female) performed at 45 hospitals (28 of 45 academic [62.2%]), S-PATH recommended type and screen orders for a median (IQR) of 32.5% (25.8%-42.0%) of cases, whereas the MSBOS approach recommended type and screens for a median (IQR) of 51.6% (46.9%-61.1%) of cases for the same sensitivity (median [IQR] difference, 17.9 [14.8-24.9] absolute percentage points). The median (IQR) S-PATH AUROC was 0.929 (0.915-0.946), whereas the median (IQR) MSBOS AUROC was 0.857 (0.822-0.884).

CONCLUSIONS AND RELEVANCE In this cohort study of 45 hospitals, a personalized surgical transfusion risk prediction algorithm demonstrated external validity and discrimination. S-PATH was consistently more effective than standard care, suggesting its potential for use as a perioperative clinical decision support tool.

Key Points

Question Does a publicly available personalized machine learning model for surgical transfusion risk consistently outperform the standard-of-care approach for guiding preoperative type and screen orders across diverse health systems?

Findings In this cohort study of 45 US hospitals, a personalized model recommended type and screen orders for a median 17.9 absolute percentage point fewer patients than the standard-of-care approach despite equivalent 96% sensitivity.

Meaning The personalized algorithm demonstrated robust external validity across a diverse cohort of hospitals, suggesting its potential to improve resource allocation if broadly implemented as a perioperative clinical decision support tool.

+ Supplemental content

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

JAMA Network Open. 2025;8(6):e2517760. doi:10.1001/jamanetworkopen.2025.17760

https://doi.org/10.1001/jamanetworkopen.2025.17760

And, Importantly, Critical Questions Remain Concerning Human-AI Interaction in a Clinical Setting

- An accurate tool, used selectively and as a useful complement to existing records, is an important factor in gaining physician engagement.
- A clinician-to-clinician alert message will be better accepted than a machine-generated electronic health record (EHR) alert that has no clinical intermediary review or involvement.
- Providing a workflow that facilitates ease of compliance and implementation, such as when a palliative care or hospice care consult is appropriate, including the availability of relevant staff, enhances uptake and use of AI-driven CDS.
- Addressing the gap in end-of-life care training among clinicians by offering (or requiring) evidence-based training, including both didactic and experiential learning components, as well as assistance on effective and efficient use of documentation, also enhances usability and acceptance.

NEJM Catalyst | Innovations in Care Delivery

CASE STUDY

Improving End-of-Life Care through AI-Based Clinical Decision Support

Jessica Londeree Saleska, MPH, PhD, Nathan Moore, MD, Kyle Pitzer, MSW, PhD, Randi Foraker, PhD, MA, FAHA, FAMIA, FACMI, Patrick White, MD, PhD, HMDC, FACP, FAAHPM

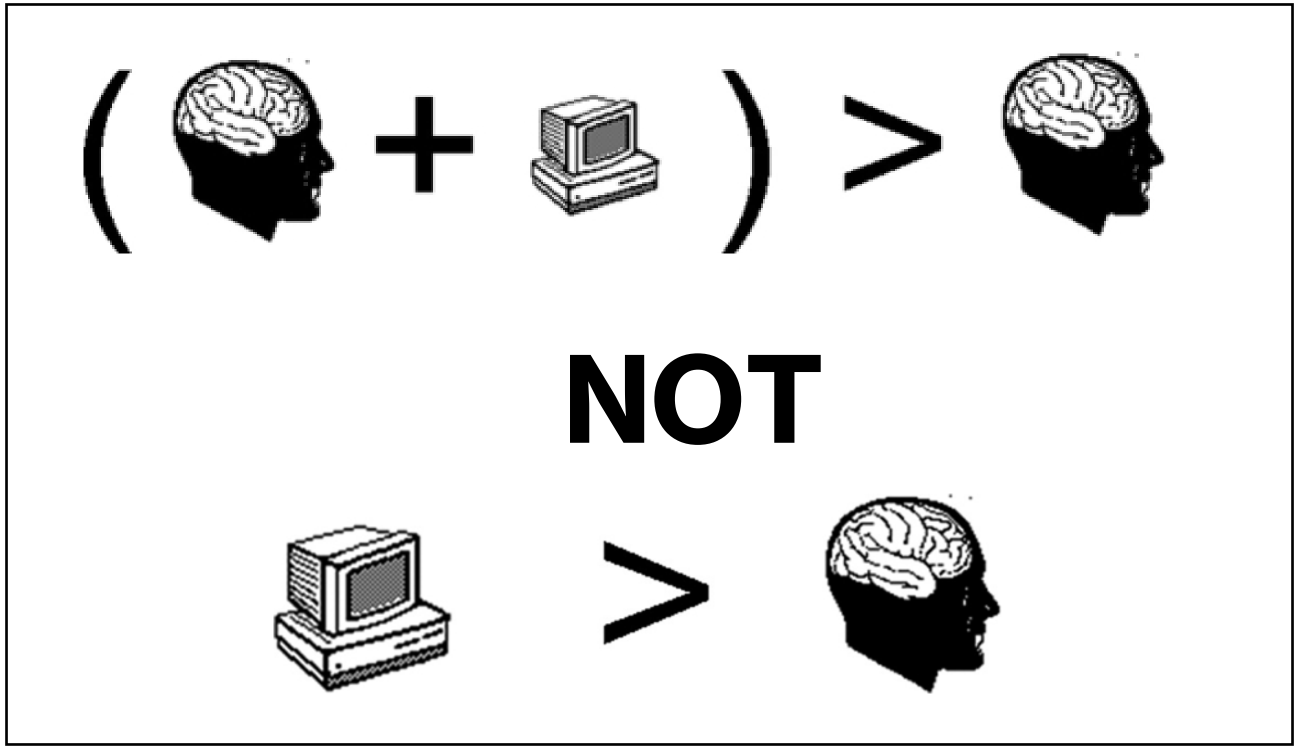
Vol. 6 No. 8 | August 2025
DOI: 10.1056/CAT.24.0392

Many patients with serious illness prioritize comfort over the prolongation of life in the final days and weeks prior to death. Goals-of-care discussions (GOCDs) can provide patients with the opportunity to express their preferred end-of-life experience, prevent aggressive and often futile interventions, improve patient satisfaction, and reduce unnecessary costs. Clinicians may feel uncomfortable initiating these conversations, however, given the lack of relevant training and difficulty identifying patients at elevated risk of mortality. The authors developed an intervention to promote GOCDs that combines mortality risk estimation using AI, clinician training, person-to-person alert messages, streamlined clinical workflows, and enhanced palliative care capacity. The program was implemented across 8 of 10 adult hospitals in the BJC HealthCare system, and data were collected between December 22, 2020 (at launch in the first hospital sites) and December 31, 2024. During this time, more than 300 clinicians were trained through the program, and they identified 13,976 patients as candidates for GOCDs. Clinicians exhibited a high response rate (93%) to the patient eligibility alert. Among patients without a prior documented GOCD at the time of messaging, 54% of responding physicians opted to initiate a GOCD with the patient, and another 24% requested palliative care to initiate the GOCD. Systemwide improvements were observed across several metrics, including a fivefold increase in GOCDs, from 1.2% of 146,257 encounters in 2021 to 6.7% of 167,681 encounters in 2024, and a 63% increase in the proportion of encounters with palliative care consults, from 2.2% of 146,257 encounters in 2021 to 3.6% of 167,681 encounters in 2024. The Vizient Mortality Index (a ratio of observed to expected mortality) also decreased by 32% during this time frame (where a lower score indicates that fewer patients died than would have been expected), from 0.92 in 2021 to 0.62 in 2024. Keys to implementation

https://doi.org/10.1056/CAT.24.0392

AI in biomedicine: enhancing human decision making

- Humans make sense of the world around them by recognizing and applying patterns
- Computers can identify patterns faster and in more significant numbers than humans can
- However, AI algorithms need to be trained:
  - Potential for bias
  - Limited by the nature of available training data
  - Costly and time-consuming
  - Continuous process
- AI is (mostly) a function of speed, as opposed to innate intelligence



Source: Friedman CP. A "fundamental theorem" of biomedical informatics. Journal of the American Medical Informatics Association. 2009 Mar 1;16(2):169-70.

Focusing on Humans

Anchoring on the humanistic needs and impact of healthcare AI

***Open-AI*:** Care team and AI collaborating to diagnose a patient

Questions?

Grace Cordovano, PhD

enlighteningresults@gmail.com

Philip R.O. Payne, PhD

prpayne@wustl.edu