

Redefining the Frontline

AI, Ethical Healthcare,
and Non-State Power



**CARNEGIE
ETHICS
FELLOWS**

**CARNEGIE
COUNCIL** *for Ethics in
International Affairs*

Contents

Introduction.....

Ethics and Power in the Age of AI in Healthcare.....

Non-State Power and Accountability in Healthcare

Trust and Operational Accountability

Moving from Principle to Practice

Conclusion.....

Notes and References.....

4

6

9

11

14

17

18

The following project was produced by a working group from the second Carnegie Ethics Fellows cohort reflecting nearly two years of convenings, collaboration, and research. Each report in this special series examines a critical issue at the intersection of ethics and international affairs. The [Carnegie Ethics Fellowship](#) aims to develop the next generation of ethical leaders across business, government, academia, and non-governmental organizations.

Cover Image Credit: Adobe/Duskcraft

Authors



REBECCA DISTLER

Carnegie Ethics Fellow

Rebecca Distler is an impact-driven leader with deep expertise at the intersection of AI and healthcare. She is a member of the Strategy team at OpenAI where she works to help translate frontier AI capabilities into public benefit through partnerships and strategic initiatives. Throughout her career, Distler has worked across technology companies, foundations, nonprofits, and governments to help ensure emerging technologies benefit the communities that stand to gain the most. Prior to OpenAI, she built and led the AI and Digital Health portfolio at the Patrick J. McGovern Foundation, a philanthropy advancing AI for public good, and served as director of Global Health Initiatives at Element, a deep learning company co-founded by Yann LeCun. Distler holds a Master's in health policy and global health from the Yale School of Public Health and a BA in political science from Yale University. She is a term member at the Council on Foreign Relations, an AI XPRIZE semifinalist, and an alumni World Economic Forum Global Shaper.



KATHY KIM

Carnegie Ethics Fellow

Kathy Kim is a mission-oriented leader engaging in AI, data, innovation, and partnership-building across public and private sectors, working with c-suite and cross-functional leaders. Through her leadership and volunteerism, Kim has been recognized as a “Rising Star” by Women in Technology, fellow with the Atlantik Bruecke Young Leaders Conference in Germany, and Global Shaper for the World Economic Forum Annual Meeting of New Champions in China. Previously, she led on the development of an enterprise AI governance system at Booz Allen Hamilton, and also led public sector engagement on AI strategy, applied data science, and technology partnerships for federal civilian and intelligence community agencies. Currently, Kim is at the Executive Office of the President Office of Management and Budget's Office of the Federal Chief CIO, where she engages chief AI officers across the federal government and contributes to government-wide technology policy, including responsible AI adoption and innovation. Her work shapes interagency collaboration and policy guidance designed to ensure AI systems are deployed safely, effectively, and in ways that benefit the public.

Carnegie Council for Ethics in International Affairs is an independent and nonpartisan nonprofit. The views expressed within this project are those of the authors and do not necessarily reflect the position of Carnegie Council or the authors' employers.

Introduction

State of AI in Healthcare

Artificial intelligence (AI) has rapidly shifted from specialized technology to general-purpose infrastructure and is becoming foundational to how people access information and conduct their day-to-day work. McKinsey's 2025 State of AI survey found that 88 percent of respondents reported regular AI use in at least one business function, up from 78 percent the prior year.¹ Pew Research Center found that 34 percent of U.S. adults had used ChatGPT as of early 2025, twice what was reported in 2023.² Among adults under 30, that figure was 58 percent.

Healthcare is one of the most visible and highly sensitive areas where this transition is unfolding. Around the world, health systems face workforce shortages, physician burnout, and unequal access to care, and are seeking cost-effective technologies to help solve these challenges. In a world where healthcare has traditionally been slow to adopt and scale new technology, AI appears to break this pattern. In 2026, the American Medical Association reported that over 80 percent of physician respondents currently use AI in a professional context, double what was reported in 2023.³

While AI has the potential to improve the efficiency, quality, and accessibility of healthcare, it is not without risks. Ethical considerations related to the development and deployment of AI in healthcare have long been documented, including concerns around bias and unrepresentative training data, privacy, data governance, accountability, liability, and the importance of human oversight.⁴

Existing ethical frameworks for AI in healthcare, as well as broader ethical AI frameworks, provide important organizing principles to guide decision-making. Yet many were developed before the widespread use of large language models (LLMs), and before the emergence of the advanced reasoning, multimodal, and agentic systems that are defining the next era of AI.

1 Singla, Alex, Alexander Sukharevsky, Bryce Hall, Lareina Yee, Michael Chui, and Tara Balakrishnan. "The State of AI in 2025: Agents, Innovation, and Transformation." McKinsey & Company, November 5, 2025. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>

2 Sidoti, Olivia, and Colleen McClain. "34% of U.S. Adults Have Used ChatGPT, About Double the Share in 2023." Pew Research Center, June 25, 2025. <https://www.pewresearch.org/short-reads/2025/06/25/34-of-us-adults-have-used-chatgpt-about-double-the-share-in-2023/>

3 American Medical Association. 2026 Physician Survey on Augmented Intelligence. Chicago: American Medical Association, Center for Digital Health and AI, March 2026. https://www.ama-assn.org/sites/ama-assn.org/files/2026-03/physician-ai-sentiment-report_0.pdf

4 Morley, Jessica, Caio C. V. Machado, Christopher Burr, Josh Cows, Indra Joshi, Mariarosaria Taddeo, and Luciano Floridi. "The Ethics of AI in Health Care: A Mapping Review." *Social Science & Medicine* 260 (2020): 113172. <https://doi.org/10.1016/j.socscimed.2020.113172>

That next phase may be arriving sooner than anticipated. AI models have already passed medical licensing exams and performed well on structured clinical assessments.^{5,6} More recent research suggests that AI systems are beginning to meet or exceed physician performance on more complex tasks. A 2026 Science study, for example, found that an LLM outperformed physician baselines on real-world emergency department case studies, a fundamentally harder challenge.⁷ These findings should not be interpreted as evidence that AI can or should replace clinicians, but they are signals of how capabilities are rapidly evolving—and of the new ethical considerations those capabilities should raise, even while deployment decisions must still depend on validation, real-world outcomes, and continuous monitoring.

Understanding this becomes more urgent when considering that non-state actors primarily lead the development of these technologies. AI labs, health technology companies, vendors, and others are increasingly shaping which problems AI is designed to solve, how this technology is integrated into health systems, and who is able to access them.

Against this backdrop, this paper explores a set of critical questions. First, as AI systems begin to reliably outperform humans in healthcare-related tasks, how do the ethical considerations shift and how should current frameworks guide this next phase? Second, assuming these capabilities and the evidence of their impact continue to advance, what new ethical obligations might arise in relation to the non-state actors that develop these systems, and in the settings where the potential health impact is the greatest?

“As AI systems begin to reliably outperform humans in healthcare-related tasks, how do the ethical considerations shift and how should current frameworks guide this next phase?”

5 Tiffany H. Kung et al., “Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models,” PLOS Digital Health 2, no. 2 (2023): e0000198, <https://doi.org/10.1371/journal.pdig.0000198>

6 Karan Singhal et al., “Large Language Models Encode Clinical Knowledge,” Nature 620 (2023): 172–180, <https://doi.org/10.1038/s41586-023-06291-2>

7 Peter G. Brodeur et al., “Performance of a Large Language Model on the Reasoning Tasks of a Physician,” Science 392, no. 6797 (2026): 524–527, <https://doi.org/10.1126/science.adz4433>

Ethics and Power in the Age of AI in Healthcare

AI Ethics Frameworks for Health and Medicine

While AI has been used in medical practice since the 1970s,⁸ the attempt to govern it through a layered ecosystem of ethical, technical, and legal frameworks is a more recent endeavor. In 2021, the World Health Organization (WHO) published the first major global framework focused specifically on the ethics and governance of AI for health. Since then, major associations such as the American Medical Association (AMA) and National Academy of Medicine (NAM) have followed with their own ethics frameworks and principles.

Table 1. Summary of AI Ethics Frameworks for Health and Medicine

Framework	Publishing Institution	Overview
Ethics and Governance of Artificial Intelligence for Health (2021) ⁹	World Health Organization (WHO)	The first major global health-specific framework on the ethics and governance of artificial intelligence; identifies ethical risks in health applications and sets out six principles for responsible use: human autonomy, safety, transparency, accountability, equity, and sustainability.
Trustworthy Augmented Intelligence in Health Care (2024) ¹⁰	American Medical Association (AMA)	Physician-facing AI principles and policy framework emphasizing ethical, equitable, responsible, accurate, and transparent AI.
Statement on Artificial and Augmented Intelligence in Medical Care (2025) ¹¹	World Medical Association (WMA)	Global physician ethics statement; emphasizes patient welfare, physician responsibility, transparency, safety, privacy, equity, professional judgment, and responsible use of AI in medical care.
An Artificial Intelligence Code of Conduct for Health and Medicine (2025) ¹²	National Academy of Medicine (NAM)	A health and medicine-specific code of conduct intended to align stakeholders around responsible development and use of AI.

8 Casimir A. Kulikowski, “Beginnings of Artificial Intelligence in Medicine,” *Yearbook of Medical Informatics* 28, no. 1 (2019): 263–270.

9 World Health Organization, *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance* (Geneva: World Health Organization, 2021), <https://www.who.int/publications/i/item/9789240029200>

10 American Medical Association, *Augmented Intelligence Development, Deployment, and Use in Health Care* (Chicago: American Medical Association, November 2024), <https://www.ama-assn.org/system/files/ama-ai-principles.pdf>

11 World Medical Association, “WMA Statement on Artificial and Augmented Intelligence in Medical Care,” adopted by the 76th WMA General Assembly, Porto, Portugal, October 2025, <https://www.wma.net/policies-post/wma-statement-on-artificial-and-augmented-intelligence-in-medical-care/>

12 National Academy of Medicine, *An Artificial Intelligence Code of Conduct for Health and Medicine: Essential Guidance for Aligned Action*, ed. Laura Adams, Elaine Fontaine, Michael Matheny, and Sunita Krishnan, NAM Special Publication (Washington, DC: National Academies Press, 2025), <https://doi.org/10.17226/29087>

These existing frameworks converge around a core set of principles. For example, all four frameworks emphasize safety, often grouping this principle with others such as “accuracy and efficacy” (WHO) and “high-quality” (AMA) that reflect the importance of ensuring AI models work as intended in healthcare contexts. Notably, NAM also links safety to preventing “potential harm.” In prioritizing safety, they also call for post-deployment and performance monitoring and ongoing assessment in real-world use.

All four frameworks also recognize AI’s potential to both reduce and potentially deepen health inequities. NAM, for example, explicitly calls for “support to low-resourced organizations and communities to ensure equitable access to the benefits of AI,” and WMA states that “new and beneficial AI healthcare tools must be developed and deployed equitably, with the goal of being accessible worldwide.”

While these frameworks provide a solid foundation for governing evolving AI capabilities, they also primarily conceive of AI as a tool that supplements human decision-making. Furthermore, they do not contemplate the potential of future autonomous systems that could take on more end-to-end clinical tasks.

As drafted, all four frameworks rely heavily on **human oversight** as a primary safeguard for AI use in healthcare. In fact, the WHO names “protect autonomy” as its first principle precisely because “people must remain in control of health-care systems and medical decisions.”¹³ AMA intentionally uses “augmented intelligence” to reflect AI’s “assistive role, emphasizing that its design enhances human intelligence rather than replaces it.” WMA goes furthest with its emphasis on “physician-in-the-loop,” stating that a licensed physician should

“review and retain final authority over all AI outputs before they shape clinical care,” to help mitigate risk and ensure patient safety.

The future of healthcare must remain human-centered. Yet, human-centered care will not always require the same level of human review of every AI-enabled action. As systems become more capable, ethical frameworks will need to calibrate oversight to level of risk—for example, requiring direct review for high-stakes decisions while allowing systems-level supervision and ongoing monitoring where appropriate.

Physicians, nurses, and other health practitioners contribute critically important cultural and contextual understanding, empathy, and relational care that AI systems should, no matter how capable, always be designed to amplify. As AI systems evolve, the question is whether every use of AI will always require the same form or intensity of human oversight, and how to evaluate and make decisions about when to move that proverbial goal post.

Today, continuous human oversight throughout the AI lifecycle is critical for ensuring that models perform as intended and that we continue to steer the development of the technology appropriately. But as increasingly powerful models become more widely available, we will need to ensure that the level of human review is proportional to the use case or decisions resulting from its use. Otherwise, AI use in critical environments could be impacted by process and overstandardization, leading to slower access to care, false assurances, administrative burden, or the reintroduction of human error into processes where the AI system was already proven to perform more reliably.

¹³ World Health Organization, Ethics and Governance of Artificial Intelligence for Health: WHO Guidance (Geneva: World Health Organization, 2021), 27, <https://www.who.int/publications/i/item/9789240029200>

This is not necessarily the state of technology today, but it is a potential future state we believe our ethical frameworks need to be prepared to address. Rather than treating human oversight as a one-size-fits-all or standardized safeguard, we need a more nuanced approach to understanding oversight of a given system, one that takes into account the system's demonstrated performance, baseline alternatives, vulnerability of affected populations, and the ability of patients and clinicians to question or override its use.

In this context, **accountability** also merits renewed consideration. All four frameworks center accountability as a core ethical principle. WHO, for example, frames accountability at a systems-level, stating that AI governance should hold “all stakeholders—in the public and private sector—accountable and responsive” to health workers, communities, and affected individuals. NAM does as well, defining accountable AI as the need to “document actions, benefits, and safeguards that ensure responsibility for AI’s effects,” effectively placing accountability not just on the clinician using the tool but on the broader system that develops, purchases, and deploys it.

On the other end of the spectrum, AMA focuses more on protecting physicians, working to ensure that they are not held responsible for tools or administrative systems they do not control. WMA attempts to do both:

emphasizing that the “final responsibility and accountability” for diagnosis and therapy must lie with the physician, while also calling for “clearly distributed accountability” and “shared and proportional” liability.

While these frameworks emphasize that AI in healthcare should remain accountable, equitable, and focused on human well-being, their treatment of oversight and accountability also raises unresolved questions. As future AI systems evolve from assisting clinicians to performing defined and complex tasks, the ethical challenge will be to clarify when human control is necessary versus when it is merely symbolic and how responsibility should be allocated across the actors who design and deploy these systems.

Many of the most consequential decisions about AI-enabled care will not necessarily be made by physicians or patients. They will be made upstream by technology companies, model developers, and other non-state actors—including philanthropies, payers, and healthcare institutions—that shape which tools get built, where they are deployed, and who benefits from their use.

The question, then, is what ethical obligations fall on the non-state actors that develop and control the infrastructure of powerful AI in healthcare.

Non-State Power and Accountability in Healthcare

The global healthcare ecosystem is composed of a wide range of stakeholders, including governments, health systems, technology companies, implementation partners, insurers, clinicians, and civil society organizations. This ecosystem has been described as one in which “the diffusion of power among state and non-state actors has transformed global health.”¹⁴

In this paper, non-state power refers to the ability of actors outside government to shape consequential decisions that impact healthcare, even when those actors do not hold formal public authority. The influence of non-state actors in healthcare is neither new nor inherently problematic. Companies and other private institutions have long contributed important scientific expertise, resources, products, and services that have helped improve human health and well-being around the world.

At the same time, tensions do arise. The commercial interests of pharmaceutical companies shape research priorities and access, insurers influence which services are covered and under what conditions, and philanthropic organizations shape global health priorities through their funding decisions and technical contributions. Across each of these examples, actors without formal governmental authority exert substantial influence on

technology use and quality of care even as states remain ultimately responsible for regulation, the protection of their citizens, and the stewardship of their health systems.

As a result, non-state healthcare actors have long had to assume certain responsibilities around safety and quality of care. Pharmaceutical companies, for example, generally develop a particular product for a defined medical purpose and bear formal obligations to demonstrate its safety and efficacy. These obligations are supplemented by industry frameworks, such as the International Federation of Pharmaceutical Manufacturers and Associations (IFPMA) Code of Practice, that establish expectations for research, promotion of products, and interactions with healthcare professionals and patient organizations.¹⁵

At the same time, the broader ethical responsibilities for these non-state actors are less clearly defined. Continuing with the example of pharmaceutical companies, considerations such as whether a company should ensure equitable access, prioritize conditions affecting underserved populations, or invest in lower-income markets do not always have comparable regulatory requirements. Rather, these types of obligations are often shaped by voluntary commitments, external incentives, and reputational considerations.

14 Julio Frenk and Suerie Moon, “Governance Challenges in Global Health,” *New England Journal of Medicine* 368, no. 10 (2013): 936–942, <https://doi.org/10.1056/NEJMr1109339>.

15 International Federation of Pharmaceutical Manufacturers & Associations (IFPMA), *IFPMA Code of Practice 2019* (Geneva: IFPMA, 2019), https://www.ifpma.org/wp-content/uploads/2018/09/IFPMA_Code_of_Practice_2019_EN.pdf.

For example, frameworks like the UN Special Rapporteur's Human Rights Guidelines for Pharmaceutical Companies address pricing, licensing, and access to medicines, but do not necessarily create enforceable obligations.¹⁶

To that end, there are gaps between what a non-state actor is legally required to do and what it may be ethically expected to do because of the influence its product or services exert over health outcomes. This gap is central to understanding non-state power in healthcare: the more the actor exerts that influence, whether intentional or not, the more important it is to ask what ethical responsibilities should reasonably follow and what are the appropriate frameworks for enforcing or upholding certain behaviors.

Naturally, AI introduces new dimensions to longstanding questions about non-state actors and power dynamics in healthcare. However, not all AI use in healthcare fits the paradigms described above. Some AI systems are developed for defined clinical purposes, such as an AI-enabled diagnostic. These systems may resemble other healthcare products: their intended use can be specified, performance can be evaluated against defined benchmarks and standards, and they can be regulated as a medical device, with obligations¹⁷ attached to the developers and purchasing institutions—even as many of these frameworks and policies are still nascent. AI's self-improving and rapid iteration cycles make oversight and governance a unique challenge compared to more static healthcare products.

General-purpose AI and the agentic systems that society is rapidly moving toward invite a different set of considerations. Broadly capable AI models and systems are not developed as healthcare products nor for singular clinical purposes. Instead, they may be adapted or used by downstream application developers, vendors, health systems, researchers, clinicians, or patients for a wide range of health-related and health-adjacent tasks. While upstream AI model developers shape important features of the AI system such as training and safeguards, the decisions that determine its impact in healthcare are often established by others.

This distinction makes the allocation of responsibility across non-state actors less straightforward. A pharmaceutical company generally knows the condition and regulatory pathway for which a medicine is being developed. A developer of a general-purpose AI model may not know or control every healthcare application built on top of it. Conversely, a downstream developer or health system may determine the clinical use but lack full control over the underlying model. Responsibility therefore cannot rest exclusively with either the upstream developer or the downstream user.

¹⁶ Paul Hunt, "Human Rights Guidelines for Pharmaceutical Companies in Relation to Access to Medicines," annex to The Right of Everyone to the Enjoyment of the Highest Attainable Standard of Physical and Mental Health, UN General Assembly, A/63/263, August 11, 2008, pp. 15–25.

¹⁷ U.S. Food and Drug Administration, "Artificial Intelligence in Software as a Medical Device," content current as of March 25, 2025, <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device>

At the same time, the distinction described above may become less clear over time as general-purpose AI systems advance. A model may not have been designed or authorized to diagnose a disease, and yet as examined in our introduction, it may outperform human clinicians on certain diagnostic tasks. At certain points in performance, ignoring a system's potential use could itself carry ethical consequences if the model could otherwise materially improve accuracy or access to healthcare.

Given these advancing capabilities, stakeholders understandably seek answers as to whether a general-purpose model is capable of performing a clinical task. We position that a separate, but important, set of questions include what evidence, safeguards, and institutional conditions should be required before those capabilities are relied on in the real world. This underscores the need for ethical frameworks to more precisely articulate how accountability is differentiated and distributed across the non-state actors that directly influence healthcare.

Upstream developers are best positioned to evaluate the model's underlying behavior and limitations and provide information that supports responsible downstream use. Application developers, vendors, and health systems may be better positioned to validate a particular healthcare use, establish appropriate human oversight, and monitor performance during implementation. Governments and regulators retain responsibility for setting the legal standards and enforcement mechanisms that govern these relationships and, ultimately, for determining when a broadly capable system crosses into a clinical use case.

Inevitably, more capable AI will put pressure on this distributed model. As systems move from generating information to performing more complex tasks, upstream vs. downstream may blur, and choices about model design and technical safeguards are likely to have more consequential effects. At the same time, decisions by health systems and other end users will continue to shape how those capabilities impact providers, patients, and the relationships between them.

Trust and Operational Accountability

The model of distributed accountability described above has implications for societal trust, both in AI systems broadly and in healthcare. Patient care is already changing as more people use AI both in daily life and in healthcare settings. KFF's Tracking Poll on Health Information and Trust found that 32 percent of adults had used AI for physical- or mental-health information or advice in the previous year. Among those who had used AI for physical-health advice, 69 percent said they trusted AI tools "a great deal" or "a fair amount" to provide reliable health information.¹⁸

18 Alex Montero et al., "KFF Tracking Poll on Health Information and Trust: Use of AI for Health Information and Advice," KFF, March 25, 2026, <https://www.kff.org/public-opinion/kff-tracking-poll-on-health-information-and-trust-use-of-ai-for-health-information-and-advice/>

Notably, the same report found that 65 percent of those who had uploaded personal medical information still reported concerns about the privacy of medical data entered into AI tools. This suggests that trust in AI is not straightforward. Users may trust a system to provide useful or reliable health information while remaining wary of how their data is collected or used. In this sense, trust and confidence in the technology's performance coexists with mistrust of the institutions responsible for developing it, and thus can influence perceptions of accountability.

Trust is also relevant to accountability because patients and clinicians may use the same AI system in different ways. Clinicians may interpret an AI output alongside other medical evidence, professional judgment, and knowledge of the patient, while individuals using a model may treat its response as complete or authoritative. As AI systems become more capable, widely available, and able to perform a broader range of health-related tasks without professional intermediaries, individual differences and preferences across AI use may also become more consequential in shaping accountability.

This underscores the need to operationalize accountability within ethics frameworks. Existing frameworks frequently identify accountability as a core principle, but stating accountability as a value does not establish how it should function in practice. As Donia observes, "While accountability is often framed

as a normative good, less clear is exactly what kind of normative work accountability is expected to do, and how it is expected to do it."¹⁹ Another unresolved question also centers on how accountability may shift when multiple actors hold responsibility and AI systems become increasingly capable, autonomous, and broadly deployed at a rapid pace.

Donia identifies five distinct "normative logics" of algorithmic accountability in health systems, recognizing that the socio-cultural norms influence how processes and their expected outcomes, including the governance of these processes, are developed. These normative logics of algorithmic accountability include (1) accountability as verification, (2) accountability as participation, (3) accountability as social license, (4) accountability as fiduciary duty, and (5) accountability as compliance. Each corresponds, respectively, to demonstrating that systems perform as expected, involving affected groups, establishing legitimacy and trust, fulfilling professional obligations, and meeting legal requirements.

These logics reflect different assumptions about the purpose of accountability and the actors responsible for carrying it out. In Table 2, we compare Donia's logics to the ethical frameworks previously summarized in Table 1, highlighting the relationship to trust and operational accountability, as well as the applicable normative logics.

19 Joseph Donia, "Algorithmic Accountabilities and Health Systems: A Review and Sociomaterial Approach," *Big Data & Society* 12, no. 2 (2025): 1-14, <https://doi.org/10.1177/20539517251334099>

Table 2. Review of Ethical Frameworks for Healthcare AI and Normative Logics of Algorithmic Accountability

<i>Ethical Frameworks for Healthcare AI</i>	Relevance to Themes of Trust and Operational Accountability	Applicable “Normative Logics”
WHO Ethics and Governance of Artificial Intelligence for Health (2021)	Refer to <i>Principle 5.3</i> “ensure transparency, explainability and intelligibility” and <i>Principle 5.4</i> “foster responsibility and accountability.”	Verification, Social License, Fiduciary Duty, Compliance
AMA Trustworthy Augmented Intelligence in Health Care (2024)	Refer to <i>Table 1 Common themes from AI guidelines and regulations</i> , highlighting “accountability—AI systems should be subject to oversight during development and deployment; right remedies should be provided if harm occurs” and “professional responsibility—Individuals and teams involved in the development and deployment of AI systems take responsibility for the performance and effects of those systems.” Refer to <i>Table 2 Crosscutting responsibilities of developers, deployers, and end users in fulfilling commitments to ethics, evidence, and equity</i> , especially highlighting the physician decision-making authority and ability to make clinical decisions such as diagnosis and treatment, and to override the AI system.	Verification, Social License, Fiduciary Duty
WMA Statement on Artificial and Augmented Intelligence in Medical Care (2025)	Refer to the defined “physician-in-the-loop (PITL)” principle recognizing the physician’s authority to “review and retain final authority over all AI outputs before they shape clinical care.” Refer to the <i>Guiding Principles for AI in Healthcare</i> such as “accountability” recognizing that “the prevalence of these tools necessitates clearly distributed accountability . . . among all stakeholders, including but not limited to developers, healthcare organizations, regulators, researchers and clinicians. ”	Verification, Participation, Social License, Fiduciary Duty, Compliance
NAM An Artificial Intelligence Code of Conduct for Health and Medicine (2025)	Refer to <i>Table 3-1 Crosswalk of Draft AICC Code Principles and Commitments</i> , recognizing the applicability of the commitment to “Engage people as partners with agency in every stage of the lifecycle” relevance with the following AICC Code Principles: “engaged,” “equitable,” “accessible,” “transparent,” and “accountable.”	Verification, Participation, Social License, Fiduciary Duty, Compliance

Together, these frameworks establish important foundations for accountability, including professional oversight, informed consent, and patient autonomy. At the same time, they are less clear about how accountability should work in practice—particularly how patients and other affected groups should help shape the design and deployment of AI in healthcare.

As previously described, and with increasingly autonomous AI systems, relying on physician oversight alone will not be enough. There needs to be clearly defined roles, opportunities for participation, and processes for distributing responsibility across many actors. Ethical frameworks must move beyond stating principles to explaining who is responsible for what, how affected people will have a meaningful voice, and what happens when the system causes harm.

Moving from Principle to Practice

Illustrative Example: Clinical Decision Support Systems in Low-Resource Settings

Low-resource healthcare settings provide an important test case for the ethical questions surfaced in this paper. These settings are often characterized by large populations in need of care, shortages of clinicians and specialists, limited diagnostic capacity, constrained budgets, unreliable electricity or Internet connectivity, long distances between patients and facilities, limited infrastructure, and substantial variation in the availability and quality of care.

The World Health Organization (WHO) projects a global shortfall of 11 million health workers by 2030, concentrated primarily in low- and lower-middle-income countries.²⁰ These shortages exist within already challenging healthcare environments: Between 5.7 million and 8.4 million deaths in low- and middle-income countries are attributed to poor-quality healthcare each year.²¹

These conditions may suggest that AI can expand access to expertise and improve the quality of care, but they also heighten the consequences of deploying systems that are not contextually appropriate. AI models can reproduce biases present in their training data, and performance may vary across languages, particularly where medical data in local languages are limited.^{22,23} These risks make local evaluation essential,

including assessment across the populations, languages, epidemiology, and existing care pathways in which an AI system will be deployed.

At the same time, evaluating the trade-offs of AI use in low-resource settings requires an honest assessment of the care patients can currently access. In some settings, the available alternative may be an overextended generalist, prolonged waits for specialist review, or no access to care at all. Recognizing this baseline is not an argument for weaker ethical standards in low-resource healthcare settings: It is a call to action to ensure that AI systems are intentionally designed, contextually appropriate, and assessed against the realities of comparative risk and safeguards needed to realize their potential benefits in these contexts.

Clinical decision support systems (CDSS) illustrate this tension. CDSS use patient-specific clinical information to provide assessments or recommendations that support evidence-based decisions at the point of care.²⁴ AI has expanded these capabilities by enabling systems to analyze larger and more complex forms of health data, including medical images, electronic health records, laboratory information, and clinical text.²⁵ In low-resource settings, these systems have the potential to make clinical knowledge and specialist interpretation available to health workers and patients who would otherwise have limited or no access to them.

20 World Health Organization, "Health Workforce," accessed July 17, 2026, <https://www.who.int/health-topics/health-workforce>.

21 World Health Organization, "Quality Health Services," May 19, 2025, <https://www.who.int/news-room/fact-sheets/detail/quality-health-services>.

22 Stephen R. Pfohl et al., "A Toolbox for Surfacing Health Equity Harms and Biases in Large Language Models," *Nature Medicine* 30 (2024): 3590–3600, <https://doi.org/10.1038/s41591-024-03258-2>.

23 Samuel Rutunda et al., "Large Language Models for Frontline Healthcare Support in Low-Resource Settings," *Nature Health* 1 (2026): 191–197, <https://doi.org/10.1038/s44360-025-00038-1>.

24 Agency for Healthcare Research and Quality, "Clinical Decision Support Systems," Patient Safety Network, accessed July 17, 2026, <https://psnet.ahrq.gov/primer/clinical-decision-support-systems>.

25 World Health Organization, *Generating Evidence for Artificial Intelligence-Based Medical Devices: A Framework for Training, Validation and Evaluation* (Geneva: World Health Organization, 2021), <https://iris.who.int/handle/10665/349093>.

Computer-aided detection (CAD) of tuberculosis (TB) on chest X-rays is a powerful example. In settings where radiologists are scarce, these systems can expand TB screening capacity and help identify patients who might otherwise go undetected. WHO now recommends qualifying CAD software as an alternative to human interpretation of digital chest X-rays for TB screening and triage among people aged fifteen and older; in 2025, it approved six software products for this use and described chest X-ray screening with CAD as an essential approach to closing gaps in TB case detection.²⁶

The rapid adoption of AI in radiology suggests that there is a threshold between AI experimentation and routine clinical use that can be and already has been crossed. Radiology accounts for more than three-quarters of AI-enabled medical devices authorized by the U.S. Food and Drug Administration, by far the largest clinical category.^{27,28} A 2024 survey of European Society of Radiology members found that 48 percent of respondents were using AI and another 25 percent planned to do so.²⁹ These figures do not demonstrate even adoption across countries or health systems, but illustrate how purpose-built AI systems can become embedded in routine care once their performance is sufficiently demonstrated and institutions develop workflows for their use.

The use of AI in radiology raises forward-looking ethical questions. If an AI system consistently performs a particular task more accurately, and more affordably and accessibly than a human being, continued reliance or insistence solely on human review may not always best serve patients. Under those conditions, the role of the clinician may shift from independently reviewing every image toward supervising the broader diagnostic process, resolving complex or uncertain cases, or helping clinicians and their patients navigate the results.

WHO's recommendation on CAD for TB screening illustrates the importance of balancing critical questions of access with human accountability. TB is the world's leading infectious disease killer, sickening over 10.7 million people and claiming over 1.2 million lives annually.³⁰ Access to screening and diagnosis is essential for improving care for millions around the world. In the example of CAD, requiring specialist review of every image would defeat the purpose and public health benefit the technology is designed to help unlock. Instead, people who screen positive must undergo confirmatory testing before treatment, helping maintain human accountability across the broader care pathway.

This example illustrates the broader considerations presented in this paper: As AI systems become capable of reliably performing certain healthcare tasks, human-centered care may sometimes require less direct human control at certain decision points and stronger

26 World Health Organization, "WHO Approves Six Software Products for Computer-Aided Detection of TB on Chest X-Ray," June 11, 2025, <https://www.who.int/news/item/11-06-2025-who-approves-six-software-products-for-computer-aided-detection-of-tb-on-chest-x-ray>

27 U.S. Food and Drug Administration, "Artificial Intelligence-Enabled Medical Devices," accessed July 17, 2026, <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices>

28 Ram Sivakumar, Brian Lue, and Shinjini Kundu, "FDA Approval of Artificial Intelligence and Machine Learning Devices in Radiology: A Systematic Review," JAMA Network Open 8, no. 11 (2025): e2542338, <https://doi.org/10.1001/jamanetworkopen.2025.42338>

29 Moreno Zanardo et al., "Impact of AI on Radiology: A EuroAIM/EuSoMII 2024 Survey among Members of the European Society of Radiology," Insights into Imaging 15 (2024): 240, <https://doi.org/10.1186/s13244-024-01801-w>

30 World Health Organization, Global Tuberculosis Report 2025 (Geneva: World Health Organization, 2025), "Top Findings and Messages in the 2025 Report."

“As AI systems become capable of reliably performing certain healthcare tasks, human-centered care may sometimes require less direct human control at certain decision points and stronger institutional accountability around the conditions of use.”

institutional accountability around the conditions of use. In low-resource settings, that shift may be especially consequential because AI is not just incrementally improving an existing service, it could be delivering a critically important service where there are not enough healthcare workers or resources to provide it.

These questions become more complex as CDSS moves from bounded tools toward general purpose models capable of supporting a wider range of tasks. A general-purpose AI system could assist with history-taking, diagnosis, treatment planning, documentation, referral decisions, and patient communication within a single encounter. And unlike a system developed for one disease or diagnostic task, its role may change across clinicians, facilities, and patient populations so its effects may be more difficult to define and assign to any single actor.

In low-resource healthcare settings, there is significant interest in and debate around how CDSS can extend forms of expertise and help improve quality of care. For example, real-world research conducted by Penda Health and OpenAI across nearly 40,000 patient visits found that clinicians with access to an LLM-based clinical copilot made 16 percent fewer diagnostic errors and 13 percent fewer treatment errors than clinicians without access to the system.³¹ Notably, a subsequent cluster-randomized trial provided evidence that while the AI system improved the quality of clinical documentation, diagnostic reasoning, and treatment planning, it did not significantly reduce treatment failure.³² This underscores a set of points we will not explore in depth in this paper, but are worth highlighting:

model performance is an incomplete measure of impact, as outcomes also depend on whether patients can obtain medicines, return for follow-up care, and access the services needed to act on those decisions, as well as our ability to measure and evaluate these complex factors in a real-world healthcare system.

Ultimately, the risks of adopting an AI system in any healthcare context must be evaluated carefully, including the risk of error, bias, poor localization, over-reliance, and weak oversight of implementation. But as AI systems become more capable, the evaluation should also account for the clinical context and the potential consequences of non-adoption. When the existing alternative is delayed or entirely absent care, a system shown to perform reliably may help reduce the harms associated with existing gaps in care.

As AI systems become more capable and as we build the evidence base around their use, questions may shift increasingly to include not only whether their use is appropriate, but also the consequences of non-use. This underscores the need for ethical frameworks to evaluate AI against the actual standard of care available in a particular setting, rather than against an idealized clinical environment.

Such frameworks must also recognize that, as AI becomes more capable, accountability must extend beyond whether the model produces an accurate output to whether the state and non-state actors involved have created the conditions required for that output to benefit patients.

31 Robert Korom et al., “AI-Based Clinical Decision Support for Primary Care: A Real-World Study,” arXiv, July 22, 2025, <https://doi.org/10.48550/arXiv.2507.16947>

32 Ambrose Agweyu et al., “Generative AI-Enabled Clinical Decision Support System in Primary Care: A Pragmatic, Cluster-Randomized Trial,” Nature Medicine (2026), <https://doi.org/10.1038/s41591-026-04503-6>

Conclusion

Existing ethics frameworks for AI in healthcare provide a strong foundation for governing the next era of AI, with broad agreement that these systems should be safe, equitable, transparent, accountable, and designed to support human well-being. While these principles remain essential, they were largely developed for a world in which AI primarily assists clinicians—not one in which increasingly capable systems operate with greater autonomy, support complex decision-making across care pathways, or interact directly with patients without a professional intermediary.

As AI capabilities advance, the ethical questions will change. Human-centered care should remain the goal, but this does not equate to requiring human oversight of every AI output. Rather, it requires a broader evaluation of how human-centered principles are applied throughout an AI lifecycle. For example, AI-assisted decision-making by physicians could lead to faster and more accurate diagnoses, decrease false reassurance, and improve access to care, particularly in low-resource settings.

Operationalizing accountability in healthcare AI will also require further development. The World Medical Association states that “AI integration does not diminish physician responsibility for patient welfare and advocacy” and that final responsibility for diagnosis and therapy must remain with the physician. At the same time, it maintains that “responsibility should be appropriately allocated among all stakeholders,” including developers, healthcare organizations, regulators, researchers, and clinicians.³³ This

captures a central tension: multiple actors may hold legitimate responsibilities, but those responsibilities will not remain static as the relative influence of those different actors shifts, as highlighted by Donia with the normative logics of accountability.

The unresolved question is how accountability should shift as systems become more capable and assume more consequential roles. Low-resource settings make these questions particularly visible and urgent, especially as there may be additional risks to non-adoption. AI may extend access to diagnostic interpretation and clinical expertise where the alternative is inconsistent or absent care. Realizing that potential, however, requires technologies that are contextually appropriate, locally evaluated, and supported by the infrastructure needed to translate AI outputs into better health outcomes. As systems demonstrate reliable performance, failing to assess or deploy beneficial capabilities may itself contribute to avoidable harm.

The next generation of healthcare AI ethics should review how human oversight and accountability are addressed throughout the AI lifecycle. Ethical frameworks also must be revisited and updated to account for scenarios in which systems are provided permissions to exercise even greater autonomy and capability to enhance patient care. Ultimately, the measure of an ethical framework should be whether it enables increasingly capable AI to improve human health while preserving safety, equity, trust, and meaningful responsibility across the full system of care.

³³ World Medical Association, “WMA Statement on Artificial and Augmented Intelligence in Medical Care,” adopted by the 76th WMA General Assembly, Porto, Portugal, October 2025, para. 9.

Notes and References

Singla, Alex, Alexander Sukharevsky, Bryce Hall, Lareina Yee, Michael Chui, and Tara Balakrishnan. "The State of AI in 2025: Agents, Innovation, and Transformation." McKinsey & Company, November 5, 2025. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>.

Sidoti, Olivia, and Colleen McClain. "34% of U.S. Adults Have Used ChatGPT, About Double the Share in 2023." Pew Research Center, June 25, 2025. <https://www.pewresearch.org/short-reads/2025/06/25/34-of-us-adults-have-used-chatgpt-about-double-the-share-in-2023/>.

American Medical Association. 2026 Physician Survey on Augmented Intelligence. Chicago: American Medical Association, Center for Digital Health and AI, March 2026. https://www.ama-assn.org/sites/ama-assn.org/files/2026-03/physician-ai-sentiment-report_0.pdf.

Morley, Jessica, Caio C. V. Machado, Christopher Burr, Josh Cowls, Indra Joshi, Mariarosaria Taddeo, and Luciano Floridi. "The Ethics of AI in Health Care: A Mapping Review." *Social Science & Medicine* 260 (2020): 113172. <https://doi.org/10.1016/j.socscimed.2020.113172>.

Kung, Tiffany H., et al. "Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models." *PLOS Digital Health* 2, no. 2 (2023): e0000198. <https://doi.org/10.1371/journal.pdig.0000198>.

Singhal, Karan, et al. "Large Language Models Encode Clinical Knowledge." *Nature* 620 (2023): 172–180. <https://doi.org/10.1038/s41586-023-06291-2>.

Brodeur, Peter G., et al. "Performance of a Large Language Model on the Reasoning Tasks of a Physician." *Science* 392, no. 6797 (2026): 524–527. <https://doi.org/10.1126/science.adz4433>.

Kulikowski, Casimir A. "Beginnings of Artificial Intelligence in Medicine." *Yearbook of Medical Informatics* 28, no. 1 (2019): 263–270.

World Health Organization. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva: World Health Organization, 2021. <https://www.who.int/publications/i/item/9789240029200>.

American Medical Association. Augmented Intelligence Development, Deployment, and Use in Health Care. Chicago: American Medical Association, November 2024. <https://www.ama-assn.org/system/files/ama-ai-principles.pdf>.

World Medical Association. "WMA Statement on Artificial and Augmented Intelligence in Medical Care." Adopted by the 76th WMA General Assembly, Porto, Portugal, October 2025. <https://www.wma.net/policies-post/wma-statement-on-artificial-and-augmented-intelligence-in-medical-care/>.

National Academy of Medicine. An Artificial Intelligence Code of Conduct for Health and Medicine: Essential Guidance for Aligned Action. Edited by Laura Adams, Elaine Fontaine, Michael Matheny, and Sunita Krishnan. NAM Special Publication. Washington, DC: National Academies Press, 2025. <https://doi.org/10.17226/29087>.

World Health Organization. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva: World Health Organization, 2021, 27. <https://www.who.int/publications/i/item/9789240029200>.

Frenk, Julio, and Suerie Moon. "Governance Challenges in Global Health." *New England Journal of Medicine* 368, no. 10 (2013): 936–942. <https://doi.org/10.1056/NEJMr1109339>.

International Federation of Pharmaceutical Manufacturers & Associations. IFPMA Code of Practice 2019. Geneva: IFPMA, 2019. https://www.ifpma.org/wp-content/uploads/2018/09/IFPMA_Code_of_Practice_2019_EN.pdf.

Hunt, Paul. "Human Rights Guidelines for Pharmaceutical Companies in Relation to Access to Medicines." Annex to The Right of Everyone to the Enjoyment of the Highest Attainable Standard of Physical and Mental Health. UN General Assembly, A/63/263, August 11, 2008, 15–25.

U.S. Food and Drug Administration. “Artificial Intelligence in Software as a Medical Device.” Content current as of March 25, 2025. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device>.

Montero, Alex, et al. “KFF Tracking Poll on Health Information and Trust: Use of AI for Health Information and Advice.” KFF, March 25, 2026. <https://www.kff.org/public-opinion/kff-tracking-poll-on-health-information-and-trust-use-of-ai-for-health-information-and-advice/>.

Donia, Joseph. “Algorithmic Accountabilities and Health Systems: A Review and Sociomaterial Approach.” *Big Data & Society* 12, no. 2 (2025): 1-14. <https://doi.org/10.1177/20539517251334099>.

World Health Organization. “Health Workforce.” Accessed July 17, 2026. <https://www.who.int/health-topics/health-workforce>.

World Health Organization. “Quality Health Services.” May 19, 2025. <https://www.who.int/news-room/fact-sheets/detail/quality-health-services>.

Pfohl, Stephen R., et al. “A Toolbox for Surfacing Health Equity Harms and Biases in Large Language Models.” *Nature Medicine* 30 (2024): 3590–3600. <https://doi.org/10.1038/s41591-024-03258-2>.

Rutunda, Samuel, et al. “Large Language Models for Frontline Healthcare Support in Low-Resource Settings.” *Nature Health* 1 (2026): 191-197. <https://doi.org/10.1038/s44360-025-00038-1>.

Agency for Healthcare Research and Quality. “Clinical Decision Support Systems.” Patient Safety Network. Accessed July 17, 2026. <https://psnet.ahrq.gov/primer/clinical-decision-support-systems>.

World Health Organization. *Generating Evidence for Artificial Intelligence-Based Medical Devices: A Framework for Training, Validation and Evaluation*. Geneva: World Health Organization, 2021. <https://iris.who.int/handle/10665/349093>.

World Health Organization. “WHO Approves Six Software Products for Computer-Aided Detection of TB on Chest X-Ray.” June 11, 2025. <https://www.who.int/news/item/11-06-2025-who-approves-six-software-products-for-computer-aided-detection-of-tb-on-chest-x-ray>.

U.S. Food and Drug Administration. “Artificial Intelligence-Enabled Medical Devices.” Accessed July 17, 2026. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices>.

Sivakumar, Ram, Brian Lue, and Shinjini Kundu. “FDA Approval of Artificial Intelligence and Machine Learning Devices in Radiology: A Systematic Review.” *JAMA Network Open* 8, no. 11 (2025): e2542338. <https://doi.org/10.1001/jamanetworkopen.2025.42338>.

Zanardo, Moreno, et al. “Impact of AI on Radiology: A EuroAIM/EuSoMII 2024 Survey among Members of the European Society of Radiology.” *Insights into Imaging* 15 (2024): 240. <https://doi.org/10.1186/s13244-024-01801-w>.

World Health Organization. *Global Tuberculosis Report 2025*. Geneva: World Health Organization, 2025. “Top Findings and Messages in the 2025 Report.”

Korom, Robert, et al. “AI-Based Clinical Decision Support for Primary Care: A Real-World Study.” *arXiv*, July 22, 2025. <https://doi.org/10.48550/arXiv.2507.16947>.

Agweyu, Ambrose, et al. “Generative AI-Enabled Clinical Decision Support System in Primary Care: A Pragmatic, Cluster-Randomized Trial.” *Nature Medicine* (2026). <https://doi.org/10.1038/s41591-026-04503-6>.

World Medical Association. “WMA Statement on Artificial and Augmented Intelligence in Medical Care.” Adopted by the 76th WMA General Assembly, Porto, Portugal, October 2025, para. 9. <https://www.wma.net/policies-post/wma-statement-on-artificial-and-augmented-intelligence-in-medical-care/>.

THE CARNEGIE ETHICS FELLOWSHIP

In today's world of geopolitical upheaval and global economic transformation, where can young leaders go to understand the power of ethical leadership, its impact on multilateral cooperation and collaboration, and how it applies to their professional and personal lives?

The Carnegie Ethics Fellowship is a space for talented young professionals to develop their capabilities and be examples of values-driven responsible leadership. Fellows collaborate on projects curated by Carnegie Council for Ethics in International Affairs, giving them the opportunity to contribute to work that has deep connections to both New York and the broader world. The two-year Fellowship is structured to develop the next generation of ethical leaders from business, government, academia, and nongovernmental organizations. The Fellowship is part of Carnegie Council's significant commitment to developing ethics in leadership and to the communities of experts that work toward this end, aligning the power of decision-making with reflective right action.

ABOUT CARNEGIE COUNCIL

Carnegie Council for Ethics in International Affairs is an independent nonprofit that works to empower ethics globally by identifying and addressing the most critical ethical issues of today and tomorrow. Founded by Andrew Carnegie over a century ago, we set the global ethical agenda and work for an ethical future by convening leading experts, building active communities, producing agenda-setting resources, and catalyzing the creation of ethical solutions to global problems. Join us in using the power of ethics to build a better world. Carnegie Council is a nonprofit 501(c) (3) institution. For more information, please visit CarnegieCouncil.org