QUADERNI DI SAN GIORGIO

39

# GLOBAL HEALTH IN THE AGE OF AI

LEO S. OLSCHKI

fondazione GIORGIO CINI ONLUS

# QUADERNI DI SAN GIORGIO

The publication of the *Quaderni di San Giorgio*, more than seventy years after the first volume, bears witness to Fondazione Giorgio Cini's desire to actively contribute to contemporary thought and to promote the dissemination of the outcomes of the scholarly gatherings held on the island of San Giorgio. These pages bring together voices from politics, science, philosophy, and all the humanities, with contributions from the most authoritative figures in international culture, that continue to shape what has always been defined on the island of San Giorgio as "dialogue between cultures." This edition of the Quaderni, while renewed, retains the same editorial style and scientific aims as in the past. "In San Giorgio we study, and from San Giorgio we speak and write. The Quaderni are therefore intended to document not the studies carried out in San Giorgio as such, but rather the ideas that San Giorgio would like to spread across the world."

fondazione GIORGIO CINI ONLUS

QUADERNI DI SAN GIORGIO

*President*

Gianfelice Rocca

QUADERNI DI SAN GIORGIO

39

# GLOBAL HEALTH IN THE AGE OF AI

Curated by

Luciano Floridi

Leo S. Olschki

Casa Editrice Leo S. Olschki
Viuzzo del Pozzetto, 8
50126 Firenze
www.olschki.it

This volume is the outcome of the “Global Health in the Age of A.I.” symposium organized by Fondazione Giorgio Cini in November 2024.

The symposium was curated by
Luciano Floridi
With Jessica Morley

Versione cartacea - Printed version
ISBN 978 88 222 9539 2

Versione digitale - Digital PDF
ISBN 978 88 222 8693 2

DOI: 10.82026/9788822286932

SUMMARY

# PROLOGUE

The Fondazione Giorgio Cini has always been a privileged place for reflection and encounter, capable of intertwining historical memory and future projections, humanist culture and the worlds of science. Founded in the wake of World War II, at a time of not only material but also ideal reconstruction, the Foundation has proved able to interpret major global transformations with the ambition of promoting dialogue between various fields of knowledge, fostering the circulation of innovative ideas. Over the decades, the Foundation has addressed issues that have ushered in crucial historical transitions, from jurisprudence to the space race, from social upheavals to relations between an East and West divided by the Iron Curtain, as well as collisions between the Global North and South. It has always done so with the same spirit: striving to address the present in all its complexity and, at the same time, to offer paths for research and to plot out long-term strategies.

The Symposium dedicated to the relationship between artificial intelligence and global health was also very much in line with this approach. Rather than exploring the scientific and technological potential of AI, the Foundation sought to examine the ethical responsibilities, the social and geopolitical implications as well as the choices that humanity is called upon to make today to shape the future. The focus on this area, rather than other areas of AI application, was not accidental. Health is the most universal, urgent and fundamental aspect of life. It therefore provides the perfect litmus test for measuring the impacts, conflicts and questions that new technologies present us with.

Today's healthcare systems are travelling along two divergent tracks. On the one hand, an aging population is leading to an increase in the demand for care and services. On the other hand, technological innovation is providing us with more and more options, but this leads to a further rise – not a reduction – in costs. Thus, rather than offsetting each other, these two trends amplify one another: healthcare systems are destined to become ever more expensive. In a changing world, monitoring the chain of transmission between innovation and its practical applications, for example through digitalisation and artificial intelligence, enables us to embrace change and to correct mistakes. The crucial point, in my opinion, is not to succumb to what scholars call 'techno-solutionism': the illusion that technology can solve all our problems. This is not and cannot be the case: AI can, first and foremost, provide us with unprecedented data analisys, but it cannot in any way replace the professionalism and assessments of doctors and healthcare personnel. The future of AI in healthcare – as the scholars who participated in the Symposium on the Island of San Giorgio remind us – does not depend solely on what is technically possible, but also on what is ethically acceptable and environmentally and socially sustainable. Philosopher Luciano Floridi, who guided the debate with great clarity, reminded us that we still have time to act. This is not an indefinite period of time, but a limited historical window in which it is still possible to set a course for how artificial intelligence will be applied to the field of health. As a global community, are we truly capable of taking responsibility for the ethical orientation of AI, or are we in danger of letting the grass grow under our feet, only to then find powerful tools already bent to the logic of exclusion and competition?

The heart of the problem lies in data. So-called big data, which feeds AI, may prove to be an unprecedented resource for medicine and public health: it allows us to identify hidden patterns, improve diagnosis and prevention, and understand the long-term dynamics – long data – that influence the spread of disease and the management of epidemics. But the collection and use of data raise crucial questions: who controls it? Which communities benefit from it? Does its use reinforce equality or fuel new inequalities? The risk is clear: if AI is developed and used merely as a tool for economic and geopolitical competitiveness, it will end up creating new divisions. It will become a commodity reserved for the few, the preserve of nations and institutions that have the resources to accumulate and process enormous amounts of data, leaving entire countries and communities behind. Instead of being a common good, AI risks becoming a factor of exclusive power, accentuating already deep imbalances.

The Symposium focused on the need for a change of perspective: thinking of AI not only as a technology at the service of precision in the manufacture of drugs in the richest countries, but as a tool for global health. It is a challenge that affects politics, economics and culture. It requires imagining a collective use of knowledge and resources, capable of including rather than excluding, of reducing inequalities rather than exacerbating them. This is why healthcare truly represents an exemplary testing ground. The management of healthcare data does not only concern the individual dimension of the patient, but reflects collective dynamics of enormous scope: the ability to predict and contain epidemics, to manage global health crises, to rethink prevention and treatment systems from a long-term perspective. The analysis of long data, which collects and interprets information distributed over decades and across large populations, becomes a strategic resource for the life sciences, but also a question of power, insofar as it determines who can set priorities and lines of action. In short, the stakes are not only economic or political. They are, above all, human. They are linked to the dignity of people and the universal right to health. Technological evolution is proceeding at a speed that risks leaving ethical reflection and political debate behind. AI is no longer a research topic confined to laboratories: it is already an integral part of national strategies, market dynamics and geopolitical challenges. Every delay increases the risk that its use will be decided by a few actors, according to particular interests.

So, is it possible to imagine and build an AI that is oriented towards the common good, open and capable of serving humanity? How can political institutions, the scientific community, health organisations and civil society all be involved? And finally, what role can the European dimension play with its tradition of universal welfare available to all? The Fondazione Giorgio Cini offers its contribution, providing a framework for reflection and a compass. Another lantern along the path, one which does not point to the destination but which lights up the terrain, helping us not to get lost and offering the world a cultural shelter in which to pause along our journey.

This volume marks the revival of a great publishing tradition of the Fondazione Giorgio Cini: that of the *Quaderni* ('notebooks') *di San Giorgio*. Between 1955 and 1984, the Foundation issued thirty-eight volumes of this prestigious series, published by Sansoni Editore, which brought together essays and contributions from the most authoritative figures in international culture. Its pages interwove voices from politics, science, philosophy, literature, art and the humanities, giving shape to what on the island of San Giorgio has always been called 'the dialogue between cultures'. The inside flap of the back cover bore a motto that perfectly summed up this vocation: "In San Giorgio, one studies, and from San Giorgio, one speaks and writes. The *Quaderni di San Giorgio* are the written message, just as the *Voce di San Giorgio* is the spoken one." Today, resuming publication of the *Quaderni* means reaffirming the Foundation's desire to actively contribute to contemporary thought, to offer tools for understanding our times, and to leave a lasting mark for future generations. This first issue, dedicated to the Symposium on Global Health in the Age of AI, renews the original spirit of the *Quaderni*: that of a culture that dialogues with and observes the world without prejudice, and that through discussion, seeks to trace a direction for all of humanity.

Gianfelice Rocca
President Fondazione Giorgio Cini

# GLOBAL HEALTH IN THE AGE OF AI: CHARTING A COURSE FOR ETHICAL IMPLEMENTATION AND SOCIETAL BENEFIT

Jessica Morley,[1*] Emmie Hine,[1,2] Huw Roberts,[3] Renée Sirbu,[1] Hutan Ashrafian,[4] Charlotte Blease,[5,6] Marisha Boyd,[7] John L. Chen,[8,9] Alexandre Chiavegatto Filho,[10] Enrico Coiera,[11] Glenn I. Cohen,[12] Amelia Fiske,[13] Nandini Jayakumar,[14] Angeliki Kerasidou,[15] Federica Mandreoli,[16] Melissa D. McCradden,[17,18] Stella Namuganza,[19] Elaine O. Nsoesie,[20] Ravi B. Parikh,[21] Sandeep Reddy,[22] Jana Sedlakova,[23,24] Tamara Subul,[25] Sophie van Baalen,[26] Hannah van Kolfschooten,[27] Luciano Floridi[1,2]

Abstract – Artificial Intelligence (AI) presents unprecedented opportunities to transform healthcare worldwide, from improving diagnostic accuracy to expanding access in underserved regions. Despite this potential and growing investment, a significant gap persists between AI's theoretical promise and its realised benefits in healthcare settings. This article examines the complex barriers impeding AI benefits realization in global health contexts, including ethical uncertainties, data infrastructure limitations, evidence quality concerns, and regulatory ambiguities. We analyze current initiatives addressing these challenges and highlight how technological solutions alone cannot resolve fundamental healthcare inequities. Drawing on the interdisciplinary perspectives and insights presented at the Global Health in the Age of AI Symposium hosted by the Cini Foundation

---

**1.** Digital Ethics Center, Yale University, New Haven, CT, USA; **2.** Department of Legal Studies, University of Bologna, Italy; **3.** Oxford Internet Institute, University of Oxford, UK; **4.** Institute of Global Health Innovation, Imperial College London, UK; **5.** Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden; **6.** Digital Health, Department of Psychiatry, Beth Israel Deaconess Medical Center, USA; **7.** Nuffield Department of Primary Care Health Sciences, University of Oxford, UK; **8.** Department of Radiology, University of Cambridge, UK; **9.** AI Centre for Value Based Healthcare, London, UK; **10.** Department of Epidemiology,

and Yale Digital Ethics Center,[1] we propose five core infrastructure requirements necessary for ethical AI implementation: robust data exchange; epistemic certainty with staff autonomy; actively protected healthcare values; validated outcomes with meaningful accountability; and environmental sustainability. These requirements form the foundation for a systems approach that balances technological advancement with ethical imperatives, contextual adaptability, and global equity considerations. We conclude that the successful integration of AI into healthcare demands coordinated action across sectors and borders, with careful attention to avoiding technological colonialism and ensuring AI serves as a force for health equity rather than widening existing disparities.

## 1. Introduction

Artificial Intelligence (AI) represents the newest chapter in the ongoing digital revolution aimed at addressing numerous global healthcare challenges, including: insufficient funding; clinician burnout; workforce shortages; structural inequities in access; unwarranted variations in care; and avoidable patient harm from medical errors. Specifically, AI is posited to help mitigate these challenges by developing new treatments and making them more targeted; providing remote or automated access to care; automating administrative and operational tasks; offering decision support

School of Public Health, University of Sao Paulo, Brazil; **11.** Australian Institute of Health Innovation, Faculty of Medicine Health and Human Sciences, Macquarie University, Australia; **12.** Harvard Law School, Cambridge, MA, USA; **13.** Institute of History and Ethics in Medicine, Department of Preclinical Medicine, TUM School of Medicine and Health, Technical University of Munich, Germany; **14.** Department of Sociology, University of Cambridge, UK; **15.** The Ethox Centre, Nuffield Department of Population Health, University of Oxford, UL; **16.** University of Modena and Reggio Emilia, Modena, Italy; **17.** Australian Institute for Machine Learning, Adelaide SA, Australia; Women's and Children's Health Network, Adelaide SA, Australia; **18.** SickKids Research Institute, Toronto ON, Canada; **19.** Hertie Institute for AI in Brain Health, University of Tübingen, Germany; **20.** School of Public Health, Boston University, Boston, MA, USA; **21.** Emory University School of Medicine, Atlanta, GA, USA; **22.** School of Public Health and Social Work, Queensland University of Technology, Brisbane, Australia; **23.** Department of Informatics, University of Zurich, Zurich, Switzerland; **24.** Digital Society Initiative, University of Zurich, Switzerland; **25.** Johns Hopkins Aramco Healthcare, Saudi Arabia; **26.** Rathenau Instituut, The Hague, The Netherlands; **27.** Amsterdam Law School, University of Amsterdam, The Netherlands; * Corresponding author: Jessica.morley@yale.edu.

[1] https://www.cini.it/en/eventi/global-health-in-the-age-of-ai-charting-a-course-for-ethical-implementation-and-societal-benefit/.

to limit the effects of cognitive overload; coordinating care more seamlessly to improve the patient experience; identifying unwarranted variations in care in real-time; conducting disease surveillance; and more.

These are not purely hypothetical use cases of AI. The AI literature is replete with examples of AI helping radiologists read images faster and identify pathology earlier; emergency departments triage patients; administrators manage waitlists more efficiently; and public health bodies predict disease outbreaks. And this is far from a definitive list. Following the advent of Generative AI (GenAI), the range of potential health uses of AI is ever-growing. Consequently, enthusiasm among policymakers, regulators, and healthcare providers is rising. This is evidenced by the large number of global healthcare strategies that place AI at, or near the centre of their plans (see supplementary material Table 1);[1] the growing number of AI solutions approved as medical devices by regulators like the US Food and Drug Administration (FDA);[2] and indicators of informal adoption such as a recent survey reporting that 20% of British General Practitioners (GPs) now use GenAI as part of their routine clinical practice.[3]

Enthusiasm alone cannot ensure the global healthcare community fully capitalises on AI's opportunities. Examples of AI working effectively in real-world healthcare settings, outside research environments, or well-funded academic hospitals, remain limited. Many, if not most, of the AI solutions described in the research literature, are not executable on the frontlines of care,[4] and development typically stalls before the real-world testing phase can be reached.[5] Solutions that do make it 'out of the lab' are concentrated in (i) narrowly focused domains of medicine rather than healthcare writ-large; and (ii) well-resourced regions, nationally and internationally, where the potential gains from AI-enabled care are less, but the foundational infrastructure is more readily available, rather than under-resourced regions where the potential gains from AI-enabled care could be much higher. As a result, there are growing concerns that a large swath of the global population may be excluded from the benefits of AI[6] and that overreliance on AI as a 'silver bullet' might come at the expense of resourcing interventions that tackle the upstream non-disease or non-medical focused determinants of health.[7]

At the moment, AI is exacerbating the digital divide, not reducing it. The reasons for this 'benefit gap' are numerous and, to an extent, well-rehearsed. At a practical level, these include: the complex and often opaque AI development journey that resists straightforward scaling;[8] inadequate infrastructure in many healthcare systems – even in well-resourced countries; skepticism among healthcare providers and patients regarding both the usability and utility of AI solutions;[9] and healthcare regulatory frameworks struggling to keep pace with rapid technological advancement.[10] At a higher level of abstraction, benefit-impediments include the extent to which crucial AI design choices are being left to Big Tech companies that are more directly incentivised to expand their user base and increase profit than to improve global health;[11] and the insufficiency – because of lack of enforcement or complete absence – of existing values- and rights-based frameworks for dealing with complex trade-offs related to accountability, cybersecurity, privacy, accuracy, and intellectual property protection and innovation.[12]

Awareness of these challenges has not yet translated into internationally agreed or consensus-driven solutions. This is partly because contextual variations between different global healthcare systems are significant, and 'what works' in one context might not work in another. However, it is also because conversations and research about the different facets of the benefits gap – technical, legal, ethical, and social – are typically conducted in siloes, inside isolated research institutions dedicated to one domain or one geographic region. This creates a 'can't see the wood for the trees' situation and leaves the suggested solutions vulnerable to the limitations of groupthink and confirmation bias. It was the aim of the Global Health in the Age of AI Symposium, hosted on San Giorgio Maggiore by the Cini Foundation and the Yale Digital Ethics Center, to help overcome these limitations and begin to foster a more interdisciplinary and international discussion.

The Symposium was held in November 2024 and brought together more than 40 leading experts in global health and AI from all over the world for three days of focused discussion. The following pages provide a summary of that discussion. The aim is not to recant every detail nor to give an answer to every question that was raised – this would be impossible – but to provide a high-level overview as it emerged from the discussion: the opportunities, the challenges, the open questions, and the available solutions. By design, this might mean that some Sections feel overly simplified to the expert in that area, like the Section on regulation for legal scholars. In such instances, the reader is encouraged to skip the relevant Section (if they are already familiar with the content) or to follow the references provided to gain deeper insight (if they are unfamiliar). The hope is that, by the end of Section 8, everyone can share the same foundation for further discussion and research.

The structure of the remainder of this article is as follows: Section 2 provides an overview of the potential benefits of AI use in healthcare. Section 3 highlights the ethical risks that can negate the benefits of AI use in healthcare, unless they are proactively mitigated. Section 4 describes the other factors (technical, social, and legal) contributing to the benefits gap. Section 5 offers some insight into various ongoing initiatives designed to overcome these technical, social, and legal barriers and close the benefits gap. Section 6 discusses why AI should not be seen as a silver bullet to all existing healthcare challenges, even if the benefits gap can be closed. Section 7 identifies a series of system-level requirements that, if met, chart a course for ethical implementation and societal benefit. Section 8 concludes the article.

## 2. The Opportunity

Understanding the widespread enthusiasm for AI in healthcare requires first clarifying what AI is. At its most fundamental level, AI can be defined as "the science of making machines [or software] do things that would require intelligence if done by humans".[13] To put it differently, AI represents a novel form of agency – distinct from human agency – that decouples problem-solving capability from the requirement for intelligence.[14] This separation may eliminate historical constraints, such as the need for local clinical expertise, that traditionally limit healthcare system design and performance. It creates significant opportunities to reimagine healthcare delivery, making it more effective, efficient, and equitable. Table 2, in the supplementary material, provides specific examples of AI's reconfiguring potential and the resulting opportunities across different healthcare contexts.

Once this reimaging power is acknowledged, it becomes clear that AI can transform everything about global healthcare, from what counts as evidence of illness in an individual patient, to how international healthcare organizations conduct disease surveillance. Therefore, growing enthusiasm for AI from national and international policymakers is unsurprising. Indeed, the potential benefits of AI is so significant that health is listed as a priority for AI development in multiple national AI strategies (e.g., Australia, Argentina, Brazil, China, Chile, Columbia, Japan, Peru, Malaysia, the UK, the Republic of Korea, and Uruguay).[15] This is good news. However, it does not mean that capitalizing on these opportunities in the 'right' way is straightforward. Introducing AI into local, national, and international healthcare systems also introduces new risks. If the use of AI in healthcare is to be successful, these risks must be mitigated proactively so that they do not detract from the potential benefits. The first step in mitigation is understanding the problem. This is the task of the next Section.

## 3. The Ethical Risks

While AI's potential to transform healthcare delivery across local, national, and international scales is evident, this transformation is not guaranteed to be positive. Without careful governance, AI's reimagining power may pose significant threats to patient safety, clinical authority, and healthcare's overall reliability and trustworthiness. Consequently, AI's promising potential for healthcare comes with substantial ethical risks. In brief, they arise from (a) flaws in the evidence base of AI (epistemic concerns related to AI inputs); the potential of AI to change the definition of illness, the nature of healthcare, and the practice of medicine (normative concerns related to AI outputs); and (c) the 'black box' nature of the AI development pipeline which undermines the effectiveness of existing accountability mechanisms (traceability concerns).[16,17] Examples of each of these 'types' of ethical risk, and the consequences of each, are provided in Table 3 in the supplementary material.

Awareness of these ethical risks has grown. In late 2024, the OECD and the World Medical Association surveyed medical associations' perspectives – representing countries from North and South

America, Asia/Pacific, Europe, the Middle East, North Africa, and Sub-Saharan Africa – regarding the introduction of AI tools. Of these, 94% expressed concern about the ethical risks of using AI in healthcare.[18] In response to this growing awareness, there has been a proliferation of primarily principle-based ethical frameworks and audit methodologies produced by academics, stakeholder groups, medical academies, national governments, and international organizations, to help healthcare providers assess – and proactively mitigate – the ethical risks of AI adoption and use. In the supplementary material Table 4 shows how different ethical risks map to the principles most often featured in ethical frameworks, and Table 5 provides some example frameworks from the FUTURE-AI Consortium,[19] the Coalition for Health AI,[20] the American National Academy of Medicine,[21] the Pan-Canadian AI for Health Guiding Principles,[22] and the World Health Organization.[23]

These frameworks reassure healthcare providers: 72% of medical associations believe AI's potential benefits greatly outweigh its ethical risks.[18] However, the overall impact of these frameworks is currently limited by four key limitations.

### 3.1. *The volume problem*

Framework proliferation has created an overwhelming landscape for AI developers. This abundance complicates framework selection for specific use cases and increases risks of ethics washing, shopping, shirking, dumping, or lobbying.[24] Pro-ethical design becomes performative rather than substantive.

### 3.2. *The operationalization problem*

High-level principles prove difficult to operationalise, focusing predominantly on the 'what' rather than the 'how.' This leaves AI developers – many lacking ethics training – with minimal practical guidance. Limitations persist even when frameworks provide specific guidance (e.g., CHAI or FUTURE-AI). For example, explainability requirements are commonly promoted but rarely achieved meaningfully through technical solutions alone. So-called "explainable AI" techniques often provide "false hope" of explainability while failing to support informed clinical decisions,[25] or patient consent, potentially undermining clinicians' ability to fulfill their fiduciary duties and patients' ability to trust in the healthcare system.[26]

### 3.3. *The tension problem*

Insufficient attention is given to instances where principles conflict with each other in principle-based frameworks, leaving developers (and other stakeholders) without guidance when facing inevitable trade-offs. This is problematic as several tensions, including those between explainability and accuracy, explainability and autonomy, privacy and equity, accuracy and accessibility, accuracy and autonomy, in arise frequently. Table 6, in the supplementary material, provides more details.

### 3.4. *The context problem*

Contextual and cultural blind spots emerge from the narrow parameters of most ethical frameworks. Current approaches predominantly focus on evaluating individual AI applications or even individual aspects of individual AI applications in isolation – assessing, for example, the ethical justifiability of specific models for specific tasks – rather than examining the broader ethical implications of systemic AI integration into healthcare systems. This myopic focus overlooks important questions about cumulative effects, institutional transformation, and healthcare values at scale. As[27] argue, meaningful ethical accountability requires looking across systems rather than merely inside them. Their perspective, which echoes that of Dobbe and Wolters (2024), and McCradden *et al.*, (2025),[28,29] conceptualises healthcare AI systems as sociotechnical assemblages rather than just collections of code and data – encompassing human actors, institutional contexts, professional practices, and power dynamics that collectively shape how AI technologies function in healthcare settings and their ethical impact. This complexity is rarely fully acknowledged in existing ethical frameworks.

Moreover, these frameworks primarily emerge from Western bioethical traditions, with their particular emphasis on individual autonomy, informed consent, and risk-benefit calculations.[30,31] This cultural positioning limits their global relevance and applicability. As AI adoption spreads globally, this Western-centric approach risks imposing inappropriate ethical standards while failing to incorporate diverse ethical perspectives that might better address local healthcare priorities and cultural contexts.

Finally, ethical risks can vary depending on the specific healthcare domain. Mental healthcare, for example, presents unique challenges given its cultural sensitivity, differing conceptualization across regions, and the highly relational nature of care. Mental health AI applications – particularly conversational AI for therapy – often embed Western treatment paradigms that may be incongruent with other cultural contexts. Furthermore, while developers may label such tools as "self-help" patients frequently perceive them as therapeutic interventions, exacerbating the risk of harm. Generic principle-based frameworks often fail to account for such contextual complexity.

These limitations leave crucial questions unanswered: what should the clinician's future role be?[32,33] What constitutes an ideal therapeutic relationship involving clinicians, patients, and AI? How can universal coverage be achieved? What balance between human-led, algorithmic-led, and augmented care is appropriate? Without addressing these questions, clinician and patient adoption often remains hesitant. Ethical uncertainty is, therefore, a significant contributing factor to the benefits gap. It is, however, far from being the *only* contributing factor. The purpose of the following Section is to provide insight into the other technical, social, and legal contributing factors.

## 4. The Benefits Gap

The benefits gap in healthcare AI is evident internationally – with less than 2% of all AI models reaching beyond the prototyping phase[34] – and nationally. Recent surveys, for example, revealed that only 27% of Australian healthcare providers have reported using AI in care delivery,[35] and that in India, even simple AI tools remain confined to well-resourced private hospitals.[36] While there is some early evidence to suggest that the use of generative AI (e.g., large language models) for administrative purposes may buck this trend,[3] claims about increasing adoption at a local scale can sometimes be over-extrapolated, and increasing use of administrative AI alone will not deliver the transformative benefits outlined in Section 2.

More positively, this measured adoption could be interpreted as the precautionary principle in action.[37] Historical examples like Watson for Oncology at Manipal Health Enterprises show the risks of premature adoption. Implemented in 2016, it was deemed a failure within a year when its recommendations proved incompatible with local patient populations and clinical practices.[18,36] This example illustrates how, without proper guidelines, AI integration into clinical workflows can introduce errors that compromise patient safety. Such failures can create lasting distrust or 'chilling effects' that discourage the use of even well-designed AI systems. This is a concern given the trend of clinicians using generative AI tools despite known risks of hallucinations and biases, which suggests its adoption is significantly outpacing the development of necessary safeguards. However, this concern must be balanced against the fact that excessive caution carries its own risks.

One such risk is a potential AI winter if healthcare improvements fail to materialise.[38,39] This would not just be a problem for investors and shareholders waiting for a return on their investment; freezing investment in AI development would be accompanied by significant opportunity costs and suggest that funds already invested had been wasted. Therefore, it is imperative that effective strategies for closing the benefits gap in a way that appropriately balances the risks of premature adoption of potentially harmful AI and overly cautious adoption of potentially life-saving or life-changing AI, be developed. This requires creating an understanding of the primary causes of the benefits gap beyond the ethical uncertainties already discussed: (i) data availability, quality, and accessibility; (ii) missing and legacy infrastructure; (iii) evidence quality; (iv) public trust; and (v) regulatory clarity. Let's see them separately.

### 4.1. *Data availability, quality, and accessibility*

Data is the foundation of all AI. The adage "rubbish in, rubbish out" is so oft-repeated that it has become cliché. If AI is trained, validated, evaluated, and deployed on poor quality, non-representative data, plagued with inaccuracy, missingness, and messiness, then it will underperform to the detriment of patient safety and healthcare utility, as highlighted in Section 3. This is just as true for administrative uses of AI as for clinician or patient-facing uses. AI used to predict which individuals are likely to miss their appointment, predict staffing requirements, write discharge summaries, or prioritise waiting lists is just as vulnerable to poor-quality inputs as AI used to diagnose or triage. For this reason,[40] state that [lack of] data availability and accessibility is the primary barrier to successfully utilizing AI in healthcare.

Healthcare has been data-driven since antiquity – Hippocrates emphasised recording all patient symptoms – yet high-quality, representative healthcare data remains surprisingly limited.[18,41] Despite the range of potential data sources – including surveys, disease registries, insurance records, genome sequences, wearables, apps, epidemic surveillance systems, imaging, and electronic health records[42] – most healthcare systems struggle to collect and maintain data of sufficient quality and representativeness for AI applications. The barriers to the availability of quality data represent a multi-layered problem.

#### 4.1.1. Lack of digitization

While well-resourced countries like the UK saw 96% of electronic health records digitised by 1996,[43] many regions still rely on paper records or non-structured electronic systems. In Brazil, approximately 20% of healthcare providers lacked electronic health records as recently as 2019.[44] In India, even in large hospitals, it is still common for a new registration number or patient file to be created every time a patient visits, for prescriptions to be handwritten, and for frontline health workers to record patient histories in notebooks using their own systems of annotation.[36]

#### 4.1.2. Representation biases

Recorded data often reflect historical patterns of discrimination or access inequities. Clinical trial data systematically under-represent women, ethnic minorities, and elderly populations. In India, documented instances show women from lower castes being denied healthcare and thus excluded from records.[36] In the UK, ethnicity recording in electronic health records varies by geography.[45] A 2020 WHO review of health information systems of 133 countries, found that less than 50% of countries record data related to mental health disorders.[46] These systemic gaps can be a substantial source of bias in AI.[42]

#### 4.1.3. Interoperability challenges

Solutions to missingness and data quality variation within and between countries often involve data sharing and record linkage. For example, Kim *et al.*, (2020)[47] developed an algorithm to recognise breast cancer combined mammography datasets from South Korea, the US, and the UK.[48] Doing this well is, however, extremely challenging. International standards for semantic interoperability, technical formats, and data governance remain inconsistent.[49] This results in incompatibilities between datasets using different terminologies, measurement units, and coding systems, increasing error risks, and preprocessing costs.[49]

#### 4.1.4. Fragmentation and privacy concerns

Healthcare data typically resides in siloed databases, often controlled by private entities.[6] Combining these datasets requires complex negotiations around value sharing, consent, and intellectual property.[50] Even when these hurdles are overcome, privacy and security risks persist. A 2023 of NHS data flows in England found that electronic health records were being shared with more than 460 subsequent academic, commercial, and public data consumers,[51] while the China Internet Network Security Report 2020 documented nearly 5 million exports of medical imaging data through domestic networks.[50] While these data are 'de-identified', no deidentification technique is entirely resistant to reidentification,[52] and, once data have left controlled healthcare environments, there is little control over what these data are used for.

#### 4.1.5. Undervaluation of data work

Transforming raw healthcare data into a format that can support AI applications requires tremendous effort, expertise, and resources. This 'data work' encompasses technical data processing and contextual interpretation. Consequently, the introduction of AI into healthcare also generates new and mostly unarticulated forms of work for clinicians, patients, clinical coders, healthcare analysts, and AI developers.[53] This work is often undervalued and rarely compensated for or adequately accounted for in calculations of the resource requirements for AI. Consequently, the data work aspects of AI development can put even greater strain on already resource-constrained health systems.

These issues combine to contribute to the problem of data poverty: "the inability for individuals, groups, or populations to benefit from a discovery or innovation due to insufficient data that are adequately representative".[54]

4.2. *Missing and legacy infrastructure*

Frontline AI deployment in any capacity – whether patient-facing, clinician-facing, or administrative – requires robust digital infrastructure (e.g., electronic health records that can support AI-based clinical decision support) and high-speed internet connectivity.[18] Without these foundational elements, AI's utility and usability become effectively nonexistent. This essential infrastructure is frequently inadequate for two distinct reasons.

4.2.1. Missing foundations

In less-resourced regions, the primary challenge is insufficient basic digital infrastructure. Building the requisite systems can be prohibitively expensive against competing priorities.[36] Consequently, necessary investments often remain unmade, resulting in patchy or nonexistent digital infrastructure. In Latin America, despite Argentina, Brazil, Chile, and Colombia launching AI initiatives and internet penetration reaching 70%, significant disparities in connectivity, computing capacities, and download speeds persist across health systems.[55] Similarly, in Nepal, where AI could improve healthcare access in remote mountainous regions, even basic internet and telephone services remain insufficient for reliable telemedicine.[56] These gaps can create dependencies on external infrastructure, introducing power imbalances – for example, India has 6 supercomputers compared to China's 63.[2]

[2] These numbers are based on the TOP500 list which updates biannually, these numbers are from November 2024 https://top500.org/lists/top500/2024/11/.

#### 4.2.2. Legacy Integrations

Conversely, the challenge stems from outdated systems in better-resourced countries where healthcare digitization began decades ago. Many healthcare facilities operate with decades-old electronic health record (EHR) systems that were never designed to interface with modern AI solutions. These legacy systems often use proprietary data formats and outdated communication protocols that make real-time data access and processing extremely difficult, yet are difficult to replace for cost or contractual reasons. In England, for example, NHS hospitals utilise multiple different EHR systems, separately configured and often running on unsupported operating systems. This fragmentation forces AI deployment to occur in non-integrated, ad-hoc implementations that may require multiple logins or manual data entry when direct access to necessary raw data is impossible. These implementations can increase cognitive burden through alert fatigue, poor timing, inappropriate system placement, and limited access to essential data at the point of care.[57-60] Such legacy issues help explain why AI performance often deteriorates when deployed at scale across different healthcare settings with varying technical infrastructure.

These infrastructure deficiencies led the WHO to conclude that none of the 133 countries surveyed in their healthcare information systems report had fully performant or sustainable information systems.[46]

### 4.3. *Evidence quality*

Clinician adoption of new healthcare technologies depends on demonstrated utility, usability, safety, effectiveness, and cost efficiency.[61] This is problematic for AI as, despite significant research and development investment, demonstrating AI's economic and clinical benefits remains challenging. It has been acknowledged that computational validation of model accuracy is insufficient for clinical deployment. As McCradden, Stephenson, *et al.*, (2020)[62] state, "accurate predictions within a static database do not translate into accurate predictions of health events in a non-stationary clinical context that includes shifting patient trends, operational and/or procedural changes, and practice updates." Yet, there is no standardised process for generating evidence of effectiveness and safety to an agreed-upon standard outside computer validation yet exists. Consequently, much of the evidence surrounding claims made by AI developers is relatively poor evidence quality. This is another complex issue.

### 4.3.1. Randomised controlled trials

Randomised Controlled Trials have traditionally been considered the gold standard of high-quality pre-market (or pre-deployment) evidence of clinical effectiveness and safety.[63] Examples of RCTs involving AI interventions do exist: the SPEC-AI RCT in Nigeria evaluated AI-guided screening for cardiomyopathy in obstetric patients.[64] Often, these RCTs fall short of expected standards. Van De Sande *et al.*, (2024),[34] found that among 64 AI-related RCTs, the most commonly reported outcome was 'fidelity' – how closely the implementation matched developer intentions – while outcomes more directly relevant to clinical utility appeared in fewer than 10% of the RCTs reviewed. However, the overarching problem is that, given how expensive and slow-paced RCTs are, they remain rare in AI,[65] and no alternative gold standard of evidence has been agreed upon.

### 4.3.2. Unreported results

Alternative evaluation approaches do exist, even if what is to become the new gold standard is not yet known, but the impact of these is minimised by reporting gaps. As Nsoesie (2018)[66] stated, there is "a dearth of published studies on the evaluation of AI diagnostic tools used in clinical settings." Microsoft's 2019 announcement of screening over 200,000 people in Indira for cardiovascular disease using an "AI-powered API across Apollo Hospitals" was, for example, never followed by published evaluation results. Moreover, when publications appear, they often do not contain key information but feature problematic reporting practices. Andaur Navrro *et al.*, (2023)[67] found that most machine learning prediction model studies contained multiple instances of "spin" using overly optimistic language like "outperformed" "improved" or "superior" without external validation, while still recommending clinical implementation.

#### 4.3.3. Inadequate evaluation metrics for generative AI

The recent proliferation of generative AI tools, particularly Large Language Models (LLMs) and chatbots for healthcare applications, has introduced further evidence challenges. As Abbasian *et al.*, (2024)[68] highlight, current evaluation frameworks primarily assess these systems using language-based metrics such as BLEU (Bilingual Evaluation Understudy) and ROUGE (Recall-Oriented Understudy for Gisting Evaluation), which focus on how closely the words and sentence structures produced by LLLMs match reference texts, rather than how well LLMs meet healthcare-specific needs. As such, these metrics fail to capture crucial elements, including: (i) clinical accuracy and groundedness (whether statements align with factual medical knowledge); (ii) empathy and emotional support (the ability to provide compassionate responses to healthcare interactions); (iii) trustworthiness indicators (including safety, privacy, bias, and interpretability); (iv) performance constraints (including latency, memory efficiency, and computational requirements); and (v) risks of hallucination. These limitations in evaluation metrics create substantial uncertainty about generative AI's real-world safety and effectiveness, even as these technologies are being rapidly adopted.[68]

#### 4.3.4. Expectation meets reality

Promising pre-deployment results frequently fail to translate to real-world benefits. AI models with higher diagnostic accuracy than humans in controlled settings often show limited impact on patient outcomes, staff satisfaction, or costs after implementation.[69] A survey of AI use in Czech hospitals found that while approximately 45% of representatives reported using AI in clinical practice, 15% found it either misleading or ineffective.[70] These disparities stem partly from context-dependent performance variations – the effectiveness of AI tools depends heavily on implementation approach, usage context, and user interaction.[71] For example, AI clinical decision support effectiveness hinges on how it influences end-user decisions – effects rarely tested pre-deployment.[72] With so many confounding factors, identical AI systems deployed across different settings seldom achieve consistent outcomes.

These evidence challenges collectively create ambiguity about AI systems' practical value, hindering adoption.

### 4.4. *Public trust*

Patient and public trust is essential for successful AI adoption, yet often overlooked in technical development processes. Without trust, even technically superior AI systems may face patient resistance – a significant impediment to benefits realization. Maintaining trust is a multidimensional challenge that extends beyond technical performance, and relies upon stakeholders (a) obtaining the 'social license' to develop, deploy, and use AI; and (b) ensuring the alignment between AI systems and healthcare values. Rising to this challenge necessitates paying attention to several different factors.

#### 4.4.1. The meaning of public benefit

As already stressed, data is the foundation of all AI. In healthcare, these foundational data are sensitive patient data used for secondary purposes – i.e., for purposes beyond direct care. AI development is thus dependent on patients being willing to provide AI developers with access to their data. Aitken *et al.*, (2018)[73] found that the perception of public benefit is often the primary factor influencing public acceptance of data sharing and use. When meaningful benefits can be demonstrated, patients and the public are generally more willing to accept various data practices that might otherwise raise concerns. However, defining "public benefit" or "public interest" presents significant challenges. These concepts are inherently value-laden, context-dependent, and often contested. What one stakeholder group perceives as beneficial may differ substantially from another's perspective. For instance, efficiency gains valued by healthcare administrators may not align with clinicians' priorities for improved care quality, or patients' desires for more personalised attention.

#### 4.4.2. The importance of relational values

Kerasidou (2020)[74] emphasises that empathy and compassion are foundational elements of patient-centered care that must be protected in AI-augmented healthcare environments if patients are expected to trust its use. Research has consistently demonstrated the benefits of empathetic care, including improved patient satisfaction, increased treatment adherence, and better health outcomes. The introduction of AI creates a paradox for these relational values. While AI promises to increase efficiency, allegedly freeing healthcare professionals to focus on empathetic engagement, it fundamentally differs from previous medical technologies. Unlike earlier innovations that extended healthcare professionals' physical capabilities, AI aims to augment or replace aspects of human reasoning and decision-making in a way that might entirely transform the therapeutic relationship between clinician and patient and threaten patient trust.

#### 4.4.3. The recurring need for public involvement

Patient and Public Involvement and Engagement (PPIE) is essential for maintaining trust throughout the AI lifecycle, as it ensures technologies align with patient values and addresses concerns before they become barriers to adoption. Effective PPIE should be participatory, inclusive (accessible to all, especially marginalised groups), discursive (creating safe spaces for learning and debate), meaningful (where participant input genuinely shapes direction), and recurring (responsive to evolving opinions and contexts).[43] However, implementing these principles in AI healthcare contexts presents unique challenges due to technical complexity limiting accessibility, commercial pressures restricting transparency, rapid development cycles hampering sustained engagement, and difficulties in representing diverse perspectives when AI's impacts may disproportionately affect already underrepresented communities.

### 4.5. *Regulatory clarity*

The regulatory context strongly influences the likelihood of successful technology adoption across healthcare systems at all scales. Clear legal obligations for all stakeholders – clinicians, insurers, and technology developers – promote adoption, while ambiguous or rapidly evolving regulatory requirements typically drive abandonment rather than adoption.[61] AI follows this same pattern. Its adoption faces significant barriers because it introduces unprecedented challenges across regulatory domains – including data protection, privacy, cybersecurity, intellectual property, liability, and medical device regulation.[75] The resulting legal uncertainty for all stakeholders impedes widespread adoption and subsequent benefits realization.[10] Whilst our focus over the following few paragraphs is on medical device regulation and liability uncertainties when AI use results in patient harm (the topics that came up most frequently during the symposium), readers should recognise that the central challenge remains broader: how regulation at large can keep pace with innovation without impeding it.

### 4.5.1. Medical device regulation

Medical device regulation aims to ensure medical devices are safe, effective, and provide clinical benefit that outweighs potential risks to patients. According to the WHO, a medical device is defined as "an article, instrument, apparatus or machine that is used in the prevention, diagnosis, or treatment of illness or disease, or for detecting, measuring, restoring, correcting or modifying the structure or function of the body for some health purpose".[76] National regulatory frameworks,[3] typically involve both pre-market approval demonstrating safety and efficacy, and post-market surveillance monitoring real-world performance

In most jurisdictions, AI tools fall within the purview of medical device regulations by meeting the definition of Software as a Medical Device (SaMD). The International Medical Devices Regulatory Forum (IMDRF),[4] defines SaMD as "software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device". Despite this, three key challenges complicate AI medical device regulation.

---

[3] For example, those produced by US Food and Drugs Administration (FDA); UK Medicines and Healthcare products Regulatory Agency (MHRA); China National Medical Products Administration (NMPA); Japan Pharmaceuticals and Medical Devices Agency (PMDA); Australia Therapeutic Goods Administration (TGA); and India Central Drugs Standard Control Organization (CDSCO).

[4] The IMDRF includes regulatory authorities from Australia, Brazil, Canada, China, Europe, Japan, Russia, Singapore, South Korea, the UK, and the US.

*Classification ambiguity*

Meeting the classification of AI as SaMD is straightforward for AI used in disease prediction or diagnosis. However, many healthcare AI applications fall into regulatory grey areas.[75] Furthermore, regulatory frameworks typically determine requirements (e.g., the type of evidence required to demonstrate safety and efficacy) based on the intended use and risk level. Contemporary AI tools often have flexible functionalities that evolve, making intended use difficult to define consistently.[77] Risk classification is equally problematic. The EU Medical Device Regulations (MDR), for example, consider SaMD to be moderate risk (class IIa) if it is used to make low-risk diagnostic or therapeutic decisions, high-risk (class IIb) if used to make diagnostic or therapeutic decisions that lead to severe deterioration or surgical intervention, and highest-risk (class III) if used to make decisions that can cause death or irreversible worsening of the patient. A tool like IBM Watson for Oncology could, therefore, be classified as class IIa, IIb, or even III, depending on the impact of the outputted decision.[75]

*Adaptive capabilities*

Unlike traditional medical devices, some AI systems continuously learn after deployment, producing seemingly emergent and poorly understood outputs.[77] This adaptability challenges many processes embedded within medical device approval designed to regulate static devices with consistent functionality and performance. Pre-market approval alone cannot guarantee consistent performance across an AI system's lifecycle. While enhanced post-market surveillance offers one solution, some jurisdictions require repeated evaluation for software updates. In Japan, for example, pre-market approval of SaMD can take over a year, with each subsequent update potentially triggering new reviews -creating such developer frustration that, in 2023, sixteen AI companies formed a consortium specifically to advocate for regulatory reform.[78]

*Generative AI Challenges*

Generative AI tools, including Large Language Models (LLMs), present novel regulatory difficulties due to their non-deterministic outputs, broad functionality range, internet-scale training data mixing facts with misinformation, inability to self-assess accuracy, and hallucination tendencies.[79,80] These properties mean even seemingly low-risk applications may carry significant risks – LLM-based scribes might introduce hallucinated diagnoses or falsely documented tests.[80] Despite these concerns, most public LLMs remain largely unregulated as they fall outside traditional SaMD definitions. Weissman *et al.*, (2025),[81] found that despite disclaimers against medical use, both GPT-3 and Llama-3 readily produced "device-like" clinical decision support when asked about medical emergencies: 100% of GPT-4 and 52% of Llama-3 responses included specific diagnoses and treatments that would legally qualify them as medical devices under existing frameworks. Neither single-shot nor multi-shot prompts based on FDA guidance reliably constrained these behaviors.

#### 4.5.2. Liability regulation

Healthcare liability law establishes responsibility and consequences when patient harm occurs, balancing patient redress against clinicians' need to practice without excessive litigation fear. AI introduces novel liability questions that existing frameworks struggle to address, causing clinician concern. The OECD and World Medical Association survey found that 89% of medical associations worried that AI applications may increase clinicians' liability if they fail to evaluate AI recommendations.[18] Two distinct liability challenges emerge:

*Standard of care conflicts*

AI recommendations may appropriately deviate from standard care guidelines by incorporating real-time medical advances or personalised treatment plans. While these deviations should theoretically improve outcomes, clinicians may hesitate to follow recommendations that differ from conventional practice due to liability concerns,[82] potentially missing opportunities to improve care. Conversely, if AI becomes the de facto standard against which clinicians are measured, practitioners may feel pressured to accept recommendations they believe are incorrect, increasing both automation bias and patient harm risks.

*Over-burdening*

Medical malpractice frameworks typically place the burden of proof on patients, requiring them to establish that the clinician's care fell below the required standard of care and that this deviation *directly* caused their harm. With AI-influenced decisions, patients must demonstrate harm resulting specifically from either the AI recommendation or the clinician's interpretation – a nearly impossible task given AI opacity. As Molnár-Gábor (2020)[83] explains, AI creates a complex "network of human-machine relationships between manufacturers, programmers, physicians as operators and patients as users" making error sources "not clearly identifiable". Furthermore, modern AI systems, especially deep learning models, operate with internal reasoning processes that are inscrutable even to developers.[84] While some legal systems offer relief through "fully controllable risk" concepts,[83] these principles apply poorly to AI systems that clinicians neither fully control nor understand. These evidentiary challenges raise the bar for patients seeking compensation for AI-related harms.

These liability issues affect all healthcare systems but become particularly acute in resource-constrained settings where frontline workers "may not have the knowledge, training, or confidence to be able to interpret and challenge AI-generated results".[36] Similarly, high-volume clinical environments offer limited opportunity to thoroughly interrogate AI outputs before acting upon the theme, while patients in resource-constrained settings may lack empowerment or means to seek redress for AI-related harm.

### 4.6. *Benefits gap: here to stay?*

Having examined these barriers, it should now be clear that the AI benefits gap has no single cause. Rather, in addition to ethical uncertainty, the gap exists due to inadequate technical and data infrastructure, insufficient evidence, and regulatory uncertainty. Fortunately, while formidable, these barriers to benefits realization are not insurmountable, and, as the following Section highlights, several initiatives are already underway to help lower them.

## 5. The Solutions

The multifaceted nature of the AI benefits gap necessitates diverse solutions rather than a singular approach. Consequently, numerous initiatives are underway across scales – from local to international – addressing various aspects of benefits realization. The scope of these efforts is too extensive to comprehensively detail every technical, legal, social, and policy development within this article. Readers should consult resources like the OECD AI Policy Observatory for comprehensive coverage. Instead, this Section focuses on five key progress areas directly addressing previously identified barriers: (1) infrastructure and interoperability investments targeting technical limitations; (2) educational interventions for addressing adoption willingness and liability concerns; (3) methodological innovations and reporting standards improving evidence quality; (4) evolving medical device regulatory frameworks balancing innovation with safety; and (5) emerging AI-specific regulatory approaches.

### 5.1. *Data infrastructure and interoperability initiatives*

To overcome the barriers to benefits-realization associated with data availability, accessibility, and quality detailed in Section 4.1, individual nations and international organizations are investing significant sums into data infrastructure. A 2022 survey[5] of the European Member States of the WHO found that 73% had allocated dedicated funding to "digital health" (including data) infrastructure during the COVID-19 pandemic. More specifically, in 2024, the Brazilian President, launched a four-year, $4 billion plan across government ministries to promote the development of AI-related infrastructure, governance, and regulation.[55] Similarly, the Singaporean Ministry of Health launched a five-year $150 million plan to embed AI at scale across the health system.[85] Although there are contextual variations, these infrastructure investments typically focus on four key areas: electronic health records (EHRs), data access environments, interoperability initiatives, and analytics platforms. Table 7 in the supplementary material provides more detail on each of these areas and some example initiatives of each.

### 5.2. *Educational interventions*

To address concerns highlighted in Section 4.4.2 about clinicians' ability to evaluate AI outputs and their potential liability, ethi-

---

[5] This is 73% of the 98% of European member states that responded.

cal governance documents (referenced in Section 3) typically require that clinical AI tools be 'explainable.' However, the broader strategy being employed by individual countries, is to provide clinicians with access to educational resources that might enable them to gain the requisite skills to maintain epistemic authority over algorithmic tools. Such educational initiatives aim to develop technical competence and strengthen clinicians' critical thinking abilities, enabling them to appropriately question, validate, or override AI recommendations when their professional judgment indicates it's necessary – thereby addressing both liability concerns and maintaining clinical autonomy. Table 8, in the supplementary material, provides some examples.

### 5.3. *Methodological innovations and reporting guidelines for evidence generation*

To tackle the issue of poor quality evidence of effectiveness and safety highlighted in Section 4.3, academic researchers, stakeholder groups, research integrity organizations, and international bodies, have been investing in the development of (a) guidance setting out the steps involved in evaluating AI tools; (b) new methodologies, other than the RCT, for evaluating AI tools; and (c) reporting guidelines for papers reporting evaluation results. Tables 9-10, in the supplementary material, provide examples of a and b, whilst an overview of reporting guidelines can be found in Kolbinger et al (2024) and Dijkstra et al (2025).[86,87]

### 5.4. *Evolving medical device regulations*

To tackle the issue of unclear medical device regulation highlighted in Section 4.4.1, various regulatory authorities, international organizations, and stakeholder groups have been investing in the development of enhanced regulatory frameworks that (a) better define when AI tools qualify as medical devices and what requirements they must meet; (b) outline processes for dealing with the adaptative nature of AI models that continuously learn; and (c) establish consistent quality, safety, and compliance measures across AI medical device lifecycles. Table 11, in the supplementary material, provides some examples of these ongoing developments.

### 5.5. *Horizontal regulation*

Beyond sector-specific regulations for healthcare AI, several horizontal AI regulatory frameworks are designed to govern AI applications across multiple sectors, including health, in development. The most comprehensive of these is the EU AI Act. The AI Act establishes a harmonised risk-based approach to regulating AI systems – categorizing them as unacceptable risk (banned), high-risk (subject to strict requirements), limited risk (transparency obligations), or minimal risk (minimal regulation). Many healthcare AI applications that fall within the scope of category II or III medical devices, according to the EU Medical Device Regulations (MDR), also fall into the high-risk category of the AI Act. Consequently, developers of these devices must not only meet the requirements related to safety and effectiveness set out in the MDR, but must also meet the AI Act's additional requirements related to risk management, technical documentation, human oversight, and conformity assessments before market approval. Aboy *et al.*, (2024)[88] provide a detailed overview of how the EU MDR and the EU AI Act intersect and, Table 12, in the supplementary material, draws from a systematic review conducted by Busch et al (2025)[89] to provide examples of other horizontal AI regulations in development outside the EU that are likely to exist with other international medical device regulations in similar ways.

### 5.6. *Limitations of current solutions*

This overview of current infrastructure, educational, methodological, and regulatory initiatives indicates the global healthcare AI community's commitment to overcoming barriers to benefits-realization – an encouraging development worthy of recognition. However, critical limitations threaten the effectiveness of these initiatives if not addressed.

#### 5.6.1. Guidance overload and standardization challenges

As first highlighted in Section 3 regarding ethics frameworks, the proliferation of guidance documents, policy frameworks, and standards has created an overwhelming landscape that AI developers, especially SMEs, struggle to navigate.[90] One systematic review of reporting standards (e.g., STARD-AI) identified over 220 items recommended for documenting model performance alone.[91]

Organizations like the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) are attempting to address this fragmentation through comprehensive standards such as ISO 13485 for medical device quality management, IEC 62304 for software lifecycle processes, and ISO/IEC 42001 for AI management systems.[90] However, while these international standards may eventually improve consistency, their global harmonization remains challenging, limiting near-term effectiveness.

### 5.6.2. Fragmented infrastructure development

Infrastructure investments, though substantial, often proceed without cohesive planning for the entire AI lifecycle. While electronic health records and data access environments receive significant attention, their availability remains inconsistent, particularly in less-resourced settings.[92] Furthermore, interoperability initiatives and analytics platforms frequently fail to address end-to-end infrastructure needs spanning development, deployment, evaluation, and ongoing monitoring. This piecemeal approach risks creating new silos and incompatibilities that may exacerbate existing implementation barriers rather than reduce them.

### 5.6.3. Data extractivism and technological colonialism

Infrastructure inequities are compounded by emerging patterns of data extractivism – where large technology companies harvest data from vulnerable populations to build and train AI systems without meaningful benefit-sharing.[93] The COVID-19 pandemic accelerated this trend, with Aouragh *et al.*, (2020)[94] documenting how contact-tracing apps and increased smartphone dependence enabled unprecedented data collection from populations worldwide. The resulting algorithms often performed poorly when deployed in the very communities whose data trained them,[95] creating a troubling cycle where data flows from under-resourced communities to technology companies without returning to data contributors. These patterns of technological colonialism threaten to deepen rather than reduce global health inequities.

### 5.6.4. Regulatory inconsistencies and arbitrage

At the international level, regulatory inconsistencies create additional challenges. While medical device regulatory frameworks generally share common elements, such as risk-based approaches and safety requirements, they diverge significantly in structure, implementation timelines, enforcement mechanisms, and technical specifications.[90]

These variations reflect differing interpretations of high-level design principles, creating tensions around jurisdictional autonomy in implementing frameworks established by bodies like the European Commission or the International Medical Device Regulators Forum. These inconsistencies create obstacles for developers entering global markets, introduce compliance ambiguities, and impede multi-centre trials necessary for robust evidence generation.

Perhaps most concerning is the risk of regulatory arbitrage – where companies strategically exploit less stringent regulatory environments to circumvent more demanding requirements elsewhere. The current AI healthcare market structure exacerbates this issue, as development capabilities concentrate on entities whose priorities might not align with those of patients or health systems. This power imbalance relegates many nations – particularly those with fewer resources – to being "rule takers" rather than "rule makers" in the global AI landscape, often resulting in inferior outcomes for citizens in jurisdictions with more permissive oversight.

These limitations underscore a critical insight: without coordinated cross-national action, fragmented AI solutions risk exacerbating existing digital inequities, hindering innovation sharing, and undermining AI's potential to address global health challenges like rare diseases or emerging public health emergencies (OECD, 2024). The broader implications of these challenges for equitable healthcare AI implementation form the focus of the next Section.

## 6. The Bigger Picture

While Section 5 addressed solutions to practical barriers impeding AI benefits realisation, successful AI development, deployment, and use face additional challenges beyond technical considerations – specifically socio-political tensions related to equity, power, and control. AI innovations alone cannot restructure the incentives underpinning existing healthcare practices, and an overemphasis on technological solutions often diverts attention from more fundamental social and structural healthcare challenges. Shipton and Vitale (2024)[7] argue that healthcare AI's expansion reveals a "politics of avoidance" – global health's tendency to favour technological interventions as workarounds for addressing more contentious commercial, economic, and political determinants of health. This approach exemplifies what Lock and Nguyen (2018)[96] call the routine objectification of health-related matters "as technical problems, to be solved through the application of technology and the conduct of science."

Such techno-optimism can mask how AI in healthcare may reinforce multiple intersections of power, including racism, patriarchy, and neoliberalism. For example, Chinn *et al.*, (2023)[97] note that while health is related to personal decisions and lifestyles, it also depends on structural factors like access to affordable healthcare, preventative care, and education. In countries where healthcare access is not universal, framing health as an individual responsibility – potentially reinforced by personalised AI health tools – can undermine support for necessary structural changes. This individualisation of health responsibility through AI tools risks obscuring the social determinants that fundamentally shape health outcomes, particularly for marginalised communities. Paying attention to these risks is particularly important given that important AI design choices are increasingly left to Big Tech companies that are "more directly incentivized to expand their user base than to expand human knowledge" potentially reinforcing biases rather than challenging them.[11] This is why, as McCoy *et al.*[98] argue, "Space, staff, stuff, systems, and support" (5S) are essential considerations for integrating AI into healthcare. A comprehensive overview of this framework and how it applies to AI is provided by McCoy *et al.*;[98] here, we focus on physical infrastructure (space), systemic shortages (staff and stuff), and social dynamics (systems and support).

### 6.1. *Physical infrastructure*

Attention to the 5S's illustrates how, just as much as missing or legacy data infrastructure can act as impediments to the benefits realisation of healthcare AI, so too can physical infrastructure significantly impede the equitable benefits realisation of AI in healthcare. Many Global Majority countries struggle to provide dignified, safe, reliable spaces with essential utilities like water and electricity; health clinics serving over one billion people globally – particularly in rural Asia, Africa, and Latin America – lack reliable electricity.[99] The physical infrastructure required for AI systems also raises additional environmental justice concerns. AI development is increasingly resource-intensive and environmentally damaging, as evidenced by Microsoft's reported 34% increase in water consumption, primarily attributed to data centres.[100] The environmental burden of this infrastructure disproportionately affects low-resource communities – such as the majority-Black South Memphis neighbourhood experiencing increased smog from xAI's new data centre[101] – and Global Majority communities, who are already more vulnerable to climate change impacts.[102,103]

### 6.2. *Systemic shortages*

The World Health Organization (WHO) projects a shortfall of 11 million healthcare workers in 2030, with the most severe shortages in lower- and middle-income countries.[104] While AI can assist with scheduling and other administrative tasks, it cannot replace the need for trained professionals. This risks creating a two-tiered system between those who can afford to see, or have access to, a human to manage their care and those who must make do with purely algorithmic care. As Shipton and Vitale (2024)[7] note, the benevolent framing of AI as a solution to healthcare worker shortages masks the political-economic drivers of these shortages in LMICs. Similarly, AI cannot resolve shortages in medical supplies, which are especially acute in low- and middle-income countries that rely on donated pharmaceuticals. These systemic shortages highlight the limitations of relying on AI to address structural resource deficits.

### 6.3. *Social dynamics*

Social factors are also vital to the successful deployment of AI in healthcare. Priority must be given to local leadership and expertise to implement appropriate solutions, avoid imposing approaches inappropriate for local contexts, and ensure that systems are trusted. Implementation science recognises the necessity of stakeholder involvement and "local champions".[105] However, current AI development patterns, dominated by the US, EU, and China, often fail to address local needs, perpetuating technological colonialism and global health inequities.

Without global systems for ensuring appropriate stakeholder engagement in AI development projects before cross-border deployment, this pattern of development risks not only implementing ineffective solutions but also imposing Western healthcare values and delivery models onto different cultural contexts – potentially undermining local healthcare practices, traditional healing systems, and cultural approaches to health and wellbeing. As Bailey *et al.*, (2020)[106] highlight, the material embeddedness of algorithms in healthcare organisations raises important questions about governance, authority, and responsibility. When algorithmic technologies require additional human work that is often informal and invisible, this can create tension with professionals' caregiving duties to which they are formally held accountable.

6.4. *Beyond technological solutionism*

The purpose of highlighting these challenges related to physical infrastructure, systemic shortages, and social dynamics is not to give the impression that investing in the development of healthcare AI is inherently problematic, nor to suggest that AI is failing because it cannot solve these more foundational issues. The purpose is, instead, to highlight three fundamental truths: (1) while technology is part of the solution to health problems, tunnel-visioned support for it may come at the expense of resourcing interventions that target upstream causes of ill health; (2) even if all the hurdles to healthcare AI benefits realisation can be overcome, it will still be essential to address the social determinants of health – such as food security, housing, and economic stability – as AI solutions cannot substitute for these basic needs; and (3), as Cohen *et al.*, (2020)[12] state, AI should not be viewed as standalone products, but as complex socio-technical systems with many interacting components that might harm rather than improve healthcare quality without consideration of the entire system.

With these truths in mind, we approach the penultimate Section focused on identifying the requirements that, if met, chart a course for AI implementation that maximises societal benefit while addressing the complex socio-technical challenges we have identified throughout this paper.

## 7. The Requirements for Benefits Gap Closure

The preceding exploration of AI in healthcare has revealed that (a) AI holds enormous promise to improve access, make healthcare more personalised, enable efficiency gains, and support clinical decision-making; and (b) those seeking to get its development, deployment, and use 'right' face significant challenges from everything from lacking data infrastructure, to inconsistent medical device regulation, to sociocultural issues related to trust and the changing dynamics of healthcare. What has emerged is that the overriding impediment to realising the relevant benefits is not a lack of knowledge regarding how to develop and deploy AI in healthcare contexts practically, but a lack of agreement regarding how to ensure that AI serves broader health equity goals rather than diverting resources and attention from them. The patterns we have identified – from entrenched data quality issues to the politics of technological avoidance – show that without intentional, coordinated efforts, AI risks exacerbating rather than ameliorating global health inequities.

While various guidelines, frameworks, and policies are being developed at national and institutional levels, until there is global coordination, achieving the overarching goal of equitable AI benefits realisation will remain difficult. The fragmented landscape of AI governance, combined with the concentration of AI development in a few countries and companies, threatens to perpetuate colonial patterns of technology transfer that impose solutions ill-suited to diverse healthcare contexts.

What is needed is a systems design approach focused on a set of core requirements that all countries can be encouraged to meet in ways that are contextually specific yet internationally interoperable and overseen by global organisations with equal representation across regions. This approach requires moving away from simply identifying barriers and enablers towards identifying the fundamental requirements necessary to create healthcare system infrastructures capable – from the start – of supporting successful AI benefits realisation that is technically feasible, socially acceptable, ethically justifiable, and legally compliant.

Below, we outline some of these essential requirements, interpreting 'infrastructure' broadly to encompass all the elements identified by Sittig and Singh (2010)[107] (see table 13 in the supplementary material), and noting that these requirements are necessary but not sufficient conditions for realising the potential of AI in healthcare while mitigating its risks. This systems-level approach neither attempts to control inherent sociotechnical complexity nor reduce it, but instead aims to make it more manageable by: (a) disaggregating the system elements, (b) identifying what new system infrastructure needs to provide for these elements to interact effectively, and (c) identifying the policy, legislative, regulatory, and strategic interventions required to realise this infrastructure.

**Requirement 1:** Design infrastructure components related to hardware, software, and clinical content around the core concept of 'robust data exchange.'

This should help ensure healthcare infrastructure facilitates rather than impedes the trustworthy and compassionate exchange of clinical data (including symptoms, diagnoses, treatments, and phenotype/genotype profiles) at both individual and aggregate levels among three key stakeholder groups: (a) AI developers; (b) policymakers and regulators who determine care standards; and (c) frontline clinicians.[108-111]

To realise this requirement, interventions must be designed to ensure that:

1. As much clinical data is recorded as accurately and comprehensively as possible at source in a way that does not interfere with or undermine the compassionate and caring exchange of information (i.e., 'data') between clinicians and patients during clinical consultation. This should help minimise the need for multiple separate sources of input data.
2. AI developers can access clinical data for AI training purposes in a way that is transparent, auditable, privacy-preserving, and

not reliant on trust-damaging deidentification and dissemination methods of exchange.

3. System-wide protection from 'vendor lock-in'. No one tech company should be able to own large parts of the global supply chain of healthcare hardware and software. This should help prevent data assets from becoming ever-more siloed, and help protect against the accumulation of further technical debt by making it more feasible for system designers to adopt a more modular 'plug-and-play' approach to technical development.
4. Seamless point-of-care integration for AI that guarantees the 'five rights' (right patient, right clinician, right information, right time, right format).[112]

Table 14, in the supplementary material, provides some example interventions, including: develop speciality-specific EHR templates; prioritise model-to-data data access mechanisms; mandate compliance with interoperability standards; establish international protocols to mitigate the risk of data colonialism and ensure data sovereignty; invest in the development of transfer learning and domain adaptation techniques; and explore alternative data sources for underrepresented regions.

**Requirement 2:** Design infrastructure components related to HCI, people, workflow, and communication, around the core concepts of 'epistemic certainty' and 'autonomous staff and patients'

This should help ensure that (a) clinicians are confident that AI outputs are valid, valuable, and constructed to improve their capacity to make shared decisions that will ultimately lead to better outcomes for their patients; (b) the autonomy of clinicians over their decisions regarding their patients is maintained; (c) clinicians remain confident in their right to this autonomy even when using AI; (d) patients feel confident in the reliability of AI tools used independently of clinicians; and (d) patients feel comfortable interpreting the outputs of AI tools used independently of clinicians.

To realise this requirement, interventions must be designed to ensure that:

1. The acceptance of epistemological responsibility for data quality and data interpretability by all relevant stakeholders, rests on the

recognition that "datasets that are of poor quality, incomplete, or unrepresentative are not merely a technical issue, but an ethical concern".[113]

2. The creation, collation, and curation of high-quality, representative, multi-modal datasets specifically for AI development is a high priority.
3. AI outputs are designed to be meaningfully interpretable, and actionable, by clinicians – not 'just explainable.'
4. Clinicians can verify the relevance and accuracy of AI outputs during clinical consultations.
5. Clinicians support shared decision-making and *combine* clinical expertise with AI insights rather than *replacing* clinical expertise with AI insights, in keeping with the tenets of evidence-based medicine and recognising that healthcare complexity cannot be reduced to simple algorithmic rules.[114]
6. The limitations and assumptions of specific AI algorithms are easily discoverable by all relevant stakeholders – including clinicians at the point of care.
7. The role of clinicians as 'learned intermediaries' is appropriately protected.
8. Decisions to develop AI for specific purposes are made in response to identified clinical or broader healthcare system needs, and are not 'just' based on what data are available and what is technically possible.
9. Decisions to implement AI in specific clinical settings are appropriately informed by technical and clinical expertise, and are not 'just' managerial decisions.

Table 15 in the supplementary material provides some example interventions, including: value data work and invest in AI-ready datasets; incorporate data literacy into medical education and continuing professional development; protect the clinician's right to override AI advice; and conduct root cause analysis to help address distrust.

**Requirement 3:** Design the infrastructure components related to people, and internal policies, procedures, and cultures around the core concept of actively protected healthcare values.

This should help ensure that universal healthcare values are not only upheld by AI, but also actively protected from all trade-offs and dilutions.[115] More specifically, this is so that healthcare systems can protect and promote universal healthcare values, including equal access, patient-centricity, trustworthiness, and high-quality care, following AI by ensuring these values are embedded in AI design at various decision points in the development, deployment, and use pipeline from what information is included in the datasets used to train AI, to how AI is built and who by, and to how AI is used to guide care.

To realise this requirement, interventions must be designed to ensure that:

1. AI is not built to meet only the healthcare needs of a small proportion of the population.
2. AI protects and enhances patient autonomy rather than overrides it.
3. AI respects all patient rights – including the meta right 'not to know' – and preferences regarding their care.
4. Patients *wish* AI to be used in their care rather than tolerate it.

Table 16 in the supplementary material provides some example interventions, including: mandate AI is built by diverse teams with a range of lived experiences; mandate patient and public involvement in all stages of the AI lifecycle; prioritise the collection, curation, standardisation, and integration of 'qualitative' health data (e.g., patient-reported outcomes) in addition to 'quantitative' health data; integrate AI initiatives with broader health system strengthening; develop coordinated policies connecting AI to social determinants of health; and prioritise AI development for conditions with high unmet needs.

**Requirement 4:** Design the infrastructure components related to external policies, management systems, and structures around the core concepts of requirements, such as 'validated outcomes' and 'meaningful accountability.'

This should help ensure that AI will (a) enable clinicians to uphold and enhance the tenets of evidence-based medicine; (b) ensure all care is delivered to an appropriate standard following the best evidence to deliver the best possible health outcome for the specific patient in question; and (c) be supported by a proportionate governance framework that sufficiently protects clinicians, patients, and the public from harm.

To realise this requirement, interventions must be designed to ensure that:

1. Information governance frameworks are simplified and made less arbitrary.
2. The risks presented by all types of AI are not underestimated by medical device regulation.
3. The risks of regulatory arbitrage are appropriately managed.
4. The healthcare system does not fall foul of 'AI exceptionalism' and subjects all AI to robust efficacy, safety, and reliability testing.
5. Liability frameworks delineate responsibilities between AI developers, healthcare organisations, and clinicians.
6. Consumer protection mechanisms provide clear routes to redress for patients harmed by AI.
7. Assumptions about the 'value' of AI are tested before being taken as given.
8. Liability frameworks fairly distribute responsibility among AI developers, healthcare organisations, and clinicians, preventing undue burden on any stakeholder while ensuring patient recourse.
9. Global regulatory interoperability is pursued to facilitate consistent standards while respecting contextual differences across healthcare systems.

Table 17, in the supplementary material, provides some example interventions, including: broaden and standardise the definition of AI as a medical device, so as often as possible, AI used in clinical care is classified as Software as a Medical Device, registered, and regularly reported publicly; adapt Health Technology Assessment (HTA) processes to be suitable for AI and ensure AI is subject to HTA before being procured by individual healthcare organisations (where applicable); ensure processes for testing robustness, reliability, accuracy, efficacy, and safety are centralised, streamlined, and transparent; mandate that all decisions along the AI development pipeline be clearly and publicly documented; embed sociotechnical evaluation in all relevant evaluation and regulatory requirements; and design AI evaluations that consider "human plus algorithm" rather than "human versus algorithm".

**Requirement 5:** Design all infrastructure components with environmental sustainability as a core consideration.

This should help ensure that AI in healthcare contributes to, rather than detracts from, broader societal goals of environmental stewardship and sustainable development. More specifically, this helps ensure that the environmental impact of AI systems is considered throughout their lifecycle – from development to deployment, and ongoing use – balancing technological advancement with ecological responsibility.

To realise this requirement, interventions must be designed to ensure that:

1. AI systems energy consumption and carbon footprints are measured, reported, and minimised.
2. Hardware resources are used efficiently and sustainably, with appropriate lifecycle management.
3. AI development prioritises necessary and high-impact implementations rather than wasteful or duplicative efforts.
4. Environmental impacts are considered alongside clinical and economic impacts in evaluation frameworks.

Table 18, in the supplementary material, provides some example interventions, including: mandate environmental impact assessments for large-scale AI deployments; establish energy efficient standards for AI; require lifecycle planning for AI hardware; develop frameworks for prioritizing high-impact use cases; and develop mobile first and offline-capable AI applications.

### 7.1. *From Requirements to Reality: A Collaborative Path Forward*

Combined, these five requirements are designed to make benefits realisation by showing how to embrace, rather than ignore, the fact that the technical performance of AI cannot be separated from human practices, social dynamics, and organisational contexts. Though the requirements provided are necessary but not sufficient, and the example interventions are illustrative but not exhaustive, we hope they provide a solid foundation upon which different stakeholders, be they local, national, or international, can work collaboratively to close the benefits gap and fully capitalise on the reimagining power of AI. In short, we hope these requirements are the start, not the end, of the conversation on how the health AI benefits gap might be closed and the reimagining power realised in full.

## 8. Conclusion

This article has analysed the rapidly evolving landscape of healthcare AI, examining its transformative potential and the complex challenges that must be addressed to realise this potential ethically and equitably. We have shown how AI can fundamentally reimagine healthcare delivery while highlighting the substantial gap between theoretical promises and practical benefits realisation. The path from initial enthusiasm to effective real-world deployment is neither straightforward nor inevitable.

The persistent benefits gap in healthcare AI stems from a complex interplay of factors: ethical uncertainties, data quality and accessibility issues, infrastructure limitations, evidence quality concerns, and regulatory ambiguities. These challenges are magnified by broader sociopolitical tensions related to equity, power, and control in healthcare. Our analysis has shown that without intentional, coordinated efforts, AI risks exacerbating rather than ameliorating global health inequities – perpetuating patterns of technological colonialism and resource concentration that have historically shaped global health.

The recommendations outlined in Section 7 offer a systems-level, design-based approach to addressing these challenges, focusing on core infrastructure requirements that span technical, social, ethical, and regulatory dimensions. However, these recommendations should be understood as necessary but still insufficient conditions for success. Actual progress requires moving beyond technological solutionism to address the fundamental social determinants of health and the structural inequities that shape healthcare access and outcomes globally.

As AI technologies continue to evolve at an unprecedented pace, particularly with the emergence of foundation models and generative AI, the healthcare community faces both extraordinary opportunities and responsibilities. The choices made today by policymakers, regulators, healthcare organisations, technology developers, clinicians, and patients will shape whether AI becomes a force for equity and improved health outcomes globally or another technological intervention that primarily benefits the already privileged.

The path forward demands humility regarding technology's limitations, vigilance about unintended consequences, and unwavering commitment to healthcare's core values. By approaching AI benefits realisation as a complex sociotechnical system design challenge, we can create a future where these powerful technologies address humanity's fundamental health needs rather than reinforcing existing inequities. This represents our essential collective task as we navigate the age of AI in global health.

## Acknowledgements

This research is the outcome of the three-day symposium, entitled "Global Health in the Age of AI", organised by the Fondazione Giorgio Cini, in collaboration with Yale's Digital Ethics Center, on 7-9 November 2024, in the Island of San Giorgio Maggiore, Venice. The authors are very grateful for the support they received from the Foundation. They declare no conflict of interests.

## References

1. Karpathakis, K., Morley, J. & Floridi, L. A Justifiable Investment in AI for Healthcare: Aligning Ambition with Reality. *Minds & Machines* 34, 38 (2024).
2. FDA. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. *FDA* (2024).
3. Blease, C. R., Locher, C., Gaab, J., Hägglund, M. & Mandl, K. D. Generative artificial intelligence in primary care: an online survey of UK general practitioners. *BMJ Health Care Inform* 31, (2024).
4. Panch, T., Mattie, H. & Celi, L. A. The "inconvenient truth" about AI in healthcare. *npj Digit. Med.* 2, 77 (2019).
5. Pucher, G. *et al.* Why implementing machine learning algorithms in the clinic is not a plug-and-play solution: a simulation study of a machine learning algorithm for acute leukaemia subtype diagnosis. *eBioMedicine* 111, 105526 (2025).
6. Alberto, I. R. I. *et al.* The impact of commercial health datasets on medical research and health-care algorithms. *The Lancet Digital Health* 5, e288–e294 (2023).

7. Shipton, L. & Vitale, L. Artificial intelligence and the politics of avoidance in global health. *Social Science and Medicine* 359, (2024).
8. Boag, W. *et al.* The algorithm journey map: a tangible approach to implementing AI solutions in healthcare. *npj Digit. Med.* 7, 87 (2024).
9. Ayorinde, A. *et al.* Health Care Professionals' Experience of Using AI: Systematic Review With Narrative Synthesis. *J Med Internet Res* 26, e55766 (2024).
10. Bouderhem, R. Shaping the future of AI in healthcare through ethics and governance. *Humanit Soc Sci Commun* 11, 416 (2024).
11. Kostick-Quenet, K. M. A caution against customized AI in healthcare. *npj Digit. Med.* 8, 13 (2025).
12. Cohen, I. G., Evgeniou, T., Gerke, S. & Minssen, T. The European artificial intelligence strategy: implications and challenges for digital health. *The Lancet Digital Health* 2, e376–e379 (2020).
13. McCarthy, J., Minsky, M., Rochester, N. & Shannon, C. *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf (1955).
14. Floridi, L. AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis. *Philos. Technol.* 38, 30, s13347-025-00858–9 (2025).
15. Maslej, N. *et al.* Artificial Intelligence Index Report 2023. Preprint at https://doi.org/10.48550/ARXIV.2310.03715 (2023).
16. Morley, J. *et al.* The ethics of AI in health care: A mapping review. *Social Science & Medicine* 260, 113172 (2020).
17. Morley, J. & Floridi, L. The Ethics of AI in Health Care: An Updated Mapping Review. Preprint at https://doi.org/10.2139/ssrn.4987317 (2024).
18. OECD. *Artificial Intelligence and the Health Workforce: Perspectives from Medical Associations on AI in Health*. vol. 28 https://www.oecd.org/en/publications/artificial-intelligence-and-the-health-workforce_9a31d8af-en.html (2024).
19. Lekadir, K. *et al.* FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ* e081554 (2025) doi:10.1136/bmj-2024-081554.
20. CHAI. Responsible AI Guide. *CHAI - Coalition for Health AI* https://chai.org/responsible-ai-guide/ (2025).
21. Adams, L. *et al.* Artificial Intelligence in Health, Health Care, and Biomedical Science: An AI Code of Conduct Principles and Commitments Discussion Draft. *National Academy of Medicine* https://nam.edu/perspectives/artificial-intelligence-in-health-health-care-and-biomedical-science-an-ai-code-of-conduct-principles-and-commitments-discussion-draft/#:~:text=The%20draft%20AI%20Code%20of,health%20systems%2C%20payers%2C%20patients%2C (2024).
22. Canada, H. Pan-Canadian AI for Health (AI4H) Guiding Principles. https://www.canada.ca/en/health-canada/corporate/transparency/health-agreements/pan-canadian-ai-guiding-principles.html (2025).
23. *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Executive Summary*. (World Health Organization, Geneva, 2021).

24. Floridi, L. Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philos. Technol.* 32, 185–193 (2019).

25. Ghassemi, M., Oakden-Rayner, L. & Beam, A. L. The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health* 3, e745–e750 (2021).

26. McCradden, M. D., Joshi, S., Mazwi, M. & Anderson, J. A. Ethical limitations of algorithmic fairness solutions in health care machine learning. *The Lancet Digital Health* 2, e221–e223 (2020).

27. Ananny, M. & Crawford, K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20, 973–989 (2018).

28. Dobbe, R. & Wolters, A. Toward Sociotechnical AI: Mapping Vulnerabilities for Machine Learning in Context. *Minds & Machines* 34, 12 (2024).

29. McCradden, M. D. *et al.* CANAIRI: the Collaboration for Translational Artificial Intelligence Trials in healthcare. *Nat Med* 31, 9–11 (2025).

30. Amugongo, L. M., Bidwell, N. J. & Corrigan, C. C. Invigorating Ubuntu Ethics in AI for healthcare: Enabling equitable care. in 583–592 (2023). doi:10.1145/3593013.3594024.

31. Ferlito, B., Segers, S., De Proost, M. & Mertes, H. Responsibility Gap(s) Due to the Introduction of AI in Healthcare: An Ubuntu-Inspired Approach. *Science and Engineering Ethics* 30, (2024).

32. Blease, C. *et al.* Artificial Intelligence and the Future of Primary Care: Exploratory Qualitative Study of UK General Practitioners' Views. *J Med Internet Res* 21, e12802 (2019).

33. Ogolodom, M. P. *et al.* Knowledge and perception of healthcare workers towards the adoption of artificial intelligence in healthcare service delivery in Nigeria. *AG Salud* 1, 16 (2023).

34. Van De Sande, D. *et al.* To warrant clinical adoption AI models require a multi-faceted implementation evaluation. *npj Digit. Med.* 7, 58 (2024).

35. Shinners, L., Aggar, C., Stephens, A. & Grace, S. Healthcare professionals' experiences and perceptions of artificial intelligence in regional and rural health districts in Australia. *Australian J Rural Health* 31, 1203–1213 (2023).

36. Parry, C. & Aneja, U. Artificial Intelligence for Healthcare: Insights from India | 3. AI in Healthcare in India: Applications, Challenges and Risks. https://www.chathamhouse.org/2020/07/artificial-intelligence-healthcare-insights-india/3-ai-healthcare-india-applications (2023).

37. Beyleveld, D. & Brownsword, R. Emerging Technologies, Extreme Uncertainty, and the Principle of Rational Precautionary Reasoning. *Law, Innovation and Technology* 4, 35–65 (2012).

38. Floridi, L. AI and Its New Winter: from Myths to Realities. *Philos. Technol.* 33, 1–3 (2020).

39. Floridi, L. Why the AI Hype is Another Tech Bubble. *Philos. Technol.* 37, 128, s13347-024-00817-w (2024).

40. Kumar, D., Ingole, A. & Choudhari, S. G. Towards Ideal Health Ecosystem With Artificial Intelligence-Driven Medical Services in India: An Overview. *Cureus* (2023) doi:10.7759/cureus.48482.

41. Panch, T., Mattie, H. & Atun, R. Artificial intelligence and algorithmic bias: implications for health systems. *Journal of Global Health* 9, 010318 (2019).

42. Näher, A.-F. *et al.* Secondary data for global health digitalisation. *The Lancet Digital Health* 5, e93–e101 (2023).

43. Goldacre, B. & Morley, J. Better, broader, safer: using health data for research and analysis. *GOV.UK* https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-research-and-analysis (2022).

44. Barbalho, I. M. P. *et al.* Electronic health records in Brazil: Prospects and technological challenges. *Front. Public Health* 10, 963841 (2022).

45. The OpenSAFELY Collaborative *et al.* Consistency, completeness and external validity of ethnicity recording in NHS primary care records: a cohort study in 25 million patients' records at source using OpenSAFELY. *BMC Med* 22, 288 (2024).

46. *SCORE for Health Data Technical Package: Global Report on Health Data Systems and Capacity 2020*. (World Health Organization, Geneva, 2021).

47. Kim, H.-E. *et al.* Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *The Lancet Digital Health* 2, e138–e148 (2020).

48. Seastedt, K. P. *et al.* Global healthcare fairness: We should be sharing more, not less, data. *PLOS Digit Health* 1, e0000102 (2022).

49. *Collective Action for Responsible AI in Health*. vol. 10 https://www.oecd.org/en/publications/collective-action-for-responsible-ai-in-health_f2050177-en.html (2024).

50. Wang, C., Zhang, J., Lassi, N. & Zhang, X. Privacy Protection in Using Artificial Intelligence for Healthcare: Chinese Regulation in Comparative Perspective. *Healthcare* 10, 1878 (2022).

51. Zhang, J. *et al.* Mapping and evaluating national data flows: transparency, privacy, and guiding infrastructural transformation. *The Lancet Digital Health* 5, (2023).

52. Gadotti, A., Rocher, L., Houssiau, F., Crețu, A.-M. & De Montjoye, Y.-A. Anonymization: The imperfect science of using data while preserving privacy. *Sci. Adv.* 10, eadn7053 (2024).

53. Fiske, A., Prainsack, B. & Buyx, A. Data Work: Meaning-Making in the Era of Data-Rich Medicine. *J Med Internet Res* 21, e11672 (2019).

54. Ibrahim, H., Liu, X., Zariffa, N., Morris, A. D. & Denniston, A. K. Health data poverty: an assailable barrier to equitable digital health care. *The Lancet Digital Health* 3, e260–e265 (2021).

55. Regazzoni, J. Health, Latin America, and the Promise of Artificial Intelligence. *Think Global Health* https://www.thinkglobalhealth.org/article/health-latin-america-and-promise-artificial-intelligence (2024).

56. Dahal, D. P. & Sharma, P. Exploring the Challenges and Opportunities of Implementing artificial intelligence in healthcare settings in Nepal: A literature review. *Kathmandu Univ Med J (KUMJ)* 21, 436–443 (2023).

57. Bowman, S. Impact of electronic health record systems on information integrity: quality and safety implications. *Perspect Health Inf Manag* 10, 1c (2013).

58. Jankovic, I. & Chen, J. H. Clinical Decision Support and Implications for the Clinician Burnout Crisis. *Yearb Med Inform* 29, 145–154 (2020).

59. Olakotan, O. O. & Mohd Yusof, M. The appropriateness of clinical decision support systems alerts in supporting clinical workflows: A systematic review. *Health Informatics J* 27, 14604582211007536 (2021).

60. Sarkar, U. & Samal, L. How effective are clinical decision support systems? *BMJ* m3499 (2020) doi:10.1136/bmj.m3499.

61. Greenhalgh, T. *et al.* Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. *J Med Internet Res* 19, e367 (2017).

62. McCradden, M. D., Stephenson, E. A. & Anderson, J. A. Clinical research underlies ethical integration of healthcare artificial intelligence. *Nat Med* 26, 1325–1326 (2020).

63. Burns, P. B., Rohrich, R. J. & Chung, K. C. The Levels of Evidence and Their Role in Evidence-Based Medicine: *Plastic and Reconstructive Surgery* 128, 305–310 (2011).

64. Adedinsewo, D. A. *et al.* Contextual challenges in implementing artificial intelligence for healthcare in low-resource environments: insights from the SPEC-AI Nigeria trial. *Front. Cardiovasc. Med.* 12, 1516088 (2025).

65. Angus, D. C. Randomized Clinical Trials of Artificial Intelligence. *JAMA* 323, 1043 (2020).

66. Nsoesie, E. O. Evaluating Artificial Intelligence Applications in Clinical Settings. *JAMA Netw Open* 1, e182658 (2018).

67. Andaur Navarro, C. L. *et al.* Systematic review finds "Spin" practices and poor reporting standards in studies on machine learning-based prediction models. *Journal of Clinical Epidemiology* S0895435623000756 (2023) doi:10.1016/j.jclinepi.2023.03.024.

68. Abbasian, M. *et al.* Foundation metrics for evaluating effectiveness of healthcare conversations powered by generative AI. *npj Digit. Med.* 7, 1–14 (2024).

69. Joshi, S. *et al.* AI as an intervention: improving clinical outcomes relies on a causal approach to AI development and validation. *Journal of the American Medical Informatics Association* 32, 589–594 (2025).

70. CzechTrade. Two-Thirds of Czech Hospitals Utilise Artificial Intelligence. https://www.czechtradeoffices.com/au/news/two-thirds-of-czech-hospitals-utilise-artificial-intelligence (2024).

71. Coiera Enrico. Assessing Technology Success and Failure Using Information Value Chain Theory. in *Studies in Health Technology and Informatics* (IOS Press, 2019). doi:10.3233/SHTI190109.

72. Lyell, D., Lustig, A., Denyer, K., Vedantam, S. & Magrabi, F. Using Clinical Simulation to Evaluate AI-Enabled Decision Support. in *Studies in Health Technology and Informatics* (eds. Bichel-Findlay, J., Otero, P., Scott, P. & Huesing, E.) (IOS Press, 2024). doi:10.3233/SHTI230975.

73. Aitken, M., Porteous, C., Creamer, E. & Cunningham-Burley, S. Who benefits and how? Public expectations of public benefits from data-intensive health research. *Big Data & Society* 5, 205395171881672 (2018).

74. Kerasidou, A. Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. *Bull. World Health Organ.* 98, 245–250 (2020).

75. Gerke, S., Minssen, T. & Cohen, G. Ethical and legal challenges of artificial intelligence-driven healthcare. in *Artificial Intelligence in Healthcare* 295–336 (Elsevier, 2020). doi:10.1016/B978-0-12-818438-7.00012-5.

76. WHO. Medical devices. https://www.who.int/health-topics/medical-devices (2025).

77. Onitiu, D., Wachter, S. & Mittelstadt, B. How AI challenges the medical device regulation: patient safety, benefits, and intended uses. *Journal of Law and the Biosciences* lsae007 (2024) doi:10.1093/jlb/lsae007.

78. Tochibayashi, N. & Kutty, N. Three AI tools revolutionising healthcare in Japan. *World Economic Forum* https://www.weforum.org/stories/2023/12/three-ai-tools-setting-the-stage-for-a-tech-revolution-by-japans-entrepreneurial-doctors/ (2023).

79. Harrer, S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *eBioMedicine* 90, 104512 (2023).

80. Warraich, H. J., Tazbaz, T. & Califf, R. M. FDA Perspective on the Regulation of Artificial Intelligence in Health Care and Biomedicine. *JAMA* 333, 241 (2025).

81. Weissman, G. E., Mankowitz, T. & Kanter, G. P. Unregulated large language models produce medical device-like output. *npj Digit. Med.* 8, 148 (2025).

82. Price, W. N., II, Gerke, S. & Cohen, I. G. Potential Liability for Physicians Using Artificial Intelligence. *JAMA* 322, 1765–1766 (2019).

83. Molnár-Gábor, F. Artificial Intelligence in Healthcare: Doctors, Patients and Liabilities. in *Regulating Artificial Intelligence* (eds. Wischmeyer, T. & Rademacher, T.) 337–360 (Springer International Publishing, Cham, 2020). doi:10.1007/978-3-030-32361-5_15.

84. Yu, K.-H., Beam, A. L. & Kohane, I. S. Artificial intelligence in healthcare. *Nat Biomed Eng* 2, 719–731 (2018).

85. Ang, A. Singapore invests $150M for public health genAI adoption. https://www.healthcareitnews.com/news/asia/singapore-invests-150m-public-health-genai-adoption (2024).

86. Kolbinger, F. R., Veldhuizen, G. P., Zhu, J., Truhn, D. & Kather, J. N. Reporting guidelines in medical artificial intelligence: a systematic review and meta-analysis. *Commun Med* 4, 71 (2024).

87. Dijkstra, P., Greenhalgh, T., Mekki, Y. M. & Morley, J. How to read a paper involving artificial intelligence (AI). *bmjmed* 4, e001394 (2025).

88. Aboy, M., Minssen, T. & Vayena, E. Navigating the EU AI Act: implications for regulated digital medical products. *npj Digit. Med.* 7, 237 (2024).

89. Busch, F. *et al.* AI regulation in healthcare around the world: what is the status quo? Preprint at https://doi.org/10.1101/2025.01.25.25321061 (2025).

90. Reddy, S. Global Harmonization of Artificial Intelligence-Enabled Software as a Medical Device Regulation: Addressing Challenges and Unifying Standards. *Mayo Clinic Proceedings: Digital Health* 3, 100191 (2025).

91. Ong, J. C. L. *et al.* Regulatory Science Innovation for Generative AI and Large Language Models in Health and Medicine: A Global Call for Action. Preprint at https://doi.org/10.48550/arXiv.2502.07794 (2025).

92. McLennan, S., Fiske, A. & Celi, L. A. Building a house without foundations? A 24-country qualitative interview study on artificial intelligence in intensive care medicine. *BMJ Health Care Inform* 31, e101052 (2024).

93. Chagnon, C. & Hagalani-Albov, S. Data extractivism: Social pollution and real-world costs. in *The European Digital Economy: Drivers of Digital Transition and Economic Recovery* (eds. Lubacha, J., Mäihäniemi, B. & Wisła, R.) (Routledge, London, 2023). doi:10.4324/9781003450160.

94. Aouragh, M., Gürses, S., Pritchard, H. & Snelting, F. The extractive infrastructures of contact tracing apps. *Journal of Environmental Media* 1, 9.1-9.9 (2020).

95. Chekroud, A. M. *et al.* Illusory generalizability of clinical prediction models. *Science* 383, 164–167 (2024).

96. Lock, M. M. & Nguyen, V.-K. *An Anthropology of Biomedicine*. (Wiley Blackwell, Hoboken, New Jersey, 2018).

97. Chinn, S., Hasell, A. & Hiaeshutter-Rice, D. Mapping Digital Wellness Content: Implications for Health, Science, and Political Communication Research. *Journal of Quantitative Description: Digital Media* 3, 1–56 (2023).

98. McCoy, L. G. *et al.* Building health systems capable of leveraging AI: applying Paul Farmer's 5S framework for equitable global health. *BMC Glob. Public Health* 3, 39 (2025).

99. WHO, World Bank & IRENA & SEforALL. *Energizing Health: Accelerating Electricity Access in Health-Care Facilities: Executive Summary*. https://www.who.int/publications/i/item/9789240066984 (2023).

100. Moss, S. Microsoft's water consumption jumps 34 percent amid AI boom. *Data Center Dynamics* https://www.datacenterdynamics.com/en/news/microsofts-water-consumption-jumps-34-percent-amid-ai-boom/ (2023).

101. Kerr, D. How Memphis became a battleground over Elon Musk's xAI supercomputer. *KERA News* https://www.keranews.org/2024-09-11/how-memphis-became-a-battleground-over-elon-musks-xai-supercomputer (2024).

102. Ludwig, F. *et al. Climate Change Impacts on Developing Countries - EU Accountability*. https://library.wur.nl/WebQuery/wurpubs/359693 (2007).

103. Fiske, A. Climate Change and Health: The Next Challenge of Ethical AI. Accepted with Lancet Global Health. *Lancet Global Health* (Forthcoming).

104. WHO. Health workforce. https://www.who.int/health-topics/health-workforce (2025).

105. Lu, A. D. *et al.* Implementation Strategies for Frontline Healthcare Professionals: People, Process Mapping, and Problem Solving. *J GEN INTERN MED* 36, 506–510 (2021).

106. Bailey, S., Pierides, D., Brisley, A., Weisshaar, C. & Blakeman, T. Dismembering organisation: The coordination of algorithmic work in healthcare. *Current Sociology* 68, 546–571 (2020).

107. Sittig, D. F. & Singh, H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. *Quality and Safety in Health Care* 19, i68–i74 (2010).
108. Detmer, D. E. Building the national health information infrastructure for personal health, health care services, public health, and research. *BMC Med Inform Decis Mak* 3, 1 (2003).
109. Dixon, B. E. & Grannis, S. J. Information Infrastructure to Support Public Health. in *Public Health Informatics and Information Systems* (eds. Magnuson, J. A. & Dixon, B. E.) 83–104 (Springer International Publishing, Cham, 2020). doi:10.1007/978-3-030-41215-9_6.
110. Mashima, D. & Ahamad, M. Enabling Robust Information Accountability in E-healthcare Systems. in (2012).
111. Praveen, S. P. *et al.* A robust framework for handling health care information based on machine learning and big data engineering techniques. *International Journal of Healthcare Management* 1–18 (2022) doi:10.1080/20479700.2022.2157071.
112. Obermeyer, Z. & Emanuel, E. J. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med* 375, 1216–1219 (2016).
113. Lee, S. S.-J. Obligations of the "Gift": Reciprocity and Responsibility in Precision Medicine. *The American Journal of Bioethics* 21, 57–66 (2021).
114. Bezemer, T. *et al.* A Human(e) Factor in Clinical Decision Support Systems. *J Med Internet Res* 21, e11732 (2019).
115. Baron, J. & Spranca, M. Protected Values. *Organizational Behavior and Human Decision Processes* 70, 1–16 (1997).

# SUPPLEMENTARY MATERIAL

| Goals for AI set out in global healthcare strategies | |
|---|---|
| Australia, Canada, US, Italy, Republic of Korea | Enable more targeted treatments for specific conditions, particularly for mental health conditions, and reduce medication errors |
| Italy, Netherlands | Predict the onset and progression of the disease |
| Canada, India, Republic of Korea, Singapore, UK, Uruguay | Increase the speed and accuracy of diagnostic pathways |
| UK | Improve access to clinical knowledge and "empower" patients to manage their own care |
| Canada | Prioritise waiting lists to reduce hospital and surgery waiting times |
| Republic of Korea | Enable preventative care |

Table 1. Example goals for AI set out in global healthcare strategy documents adapted from.[1]

| Redesign Opportunity | Description | Example |
|---|---|---|
| Separating symptoms & diagnosis | AI allows the identification of new causal mechanisms of disease and new symptoms indicating disease presence; this can improve the accuracy of diagnosis, prognosis, and treatment planning. | GNS Healthcare, in partnership with the Alliance for Clinical Trials in Oncology, wanted to improve the survival rate from colorectal cancer by better understanding the underlying disease mechanism. They built causal models with data from a clinical trial with more than 2,000 patients who were being treated with two different drugs. They were able to identify both molecular and clinical causal drivers that can serve as biomarkers for survival, so doctors can better target the right treatment to the right patient [2] |
| Separating clinical knowledge & clinician | AI can distribute specialist knowledge to non-specialist providers and underserved regions, allowing high-quality clinical decision support where specialists are unavailable. This can help improve access to care. | In rural India, the Aravind Eye Care System deployed an AI screening tool called REFER that enables non-specialist community health workers to identify diabetic retinopathy using smartphone cameras. The system achieved 85.2% sensitivity and 92% specificity, allowing early detection in remote areas where ophthalmologists are scarce, and has screened over 50,000 patients who would otherwise require travel to distant specialists for diagnosis [3] |
| Separating care & clinic | AI enables care delivery outside traditional clinical settings, allowing monitoring, diagnosis, and treatment to occur in homes and communities. This can also help improve access to care. | The SPEC-AI Nigeria trial used AI-guided screening with portable digital stethoscopes to double the detection rate of cardiomyopathy in obstetric patients compared to usual care [4] |
| Joining patient & data | AI integrates disparate patient data sources (genetic, behavioral, environmental, and social) | China's "Agent Hospital" demonstrates how AI can integrate and analyze patient data across multiple dimensions to achieve a 93% diagnostic accuracy rate on complex medical cases, handling |

| | | |
|---|---|---|
| | to create holistic patient profiles that inform personalised care strategies. | in days what would take human doctors years (MedTech World, 2025). |
| Joining service analytics & service planning | AI connects operational and clinical data to optimise resource allocation while improving clinical decision-making. | Hayat Technologies in Malaysia developed Selangkah Screening, which uses AI to predict COVID-19 hotspots based on existing health data and extended this capability to predict dengue outbreaks, linking public health surveillance with clinical response planning [6] |
| Joining research & practice | AI bridges the gap between clinical research and everyday practice by continuously learning from real-world data and integrating new evidence into care recommendations. | DeepMind and Google Health developed an AI system that analyzes electronic health records to predict acute kidney injury 48 hours before occurrence, learning from over 700,000 patient records to identify subtle patterns clinicians might miss, demonstrating how AI can transform routine clinical data into actionable insights that fit within existing hospital workflows [7] |
| Joining Big Tech & Healthcare | AI enables new partnerships between technology companies and healthcare organizations, bringing advanced computational capabilities to healthcare challenges | AstraZeneca partnered with Malaysia's Institut Kanser Negara to implement AI technology for lung screening, combining pharmaceutical expertise with advanced AI capabilities to improve early detection of lung conditions [8] |

Table 2. Opportunities for reimagining healthcare created by AI.

| | Ethical concern | Explanation | Example of impact on patient safety or system trustworthiness |
|---|---|---|---|
| **Epistemic concerns** | Inconclusive evidence | Algorithmic outcomes (e.g., classification) are probabilistic and not infallible. They are rarely sufficient to posit the existence of a causal relationship. In the case of generative AI, outputs are also stochastic and may be hallucinated. | EKG readers in smartwatches may "diagnose" a patient as suffering from arrhythmia when it may be due to a fault with the watch not being able to accurately read that user's heartbeat (for example, due to the color of their skin) or the norm is inappropriately calibrated for that individual. |
| | Inscrutable evidence | Recipients of an algorithmic decision rarely have complete oversight of the data used to train or test an algorithm, or the data points used to reach a specific decision. | A clinical decision support system deployed in a hospital may make a treatment recommendation, but it may not be clear on what basis it has made that decision, raising the risk that it has used data that are inappropriate for the individual in question or that there is a bug inthe system leading to issues with over- or under-prescribing. |
| | Misguided evidence | Algorithmic outcomes can only be as reliable | Watson for Oncology was widely used in China for diagnosis via image recognition but was pri- |

| | | | |
|---|---|---|---|
| | | (or neutral) as the data they are based on. | marily trained on a Western dataset, leading to issues with concordance and poorer results for Chinese patients than their Western counterparts. |
| **Normative concerns** | Unfair outcomes | An action can be found to have more of an impact (positive or negative) on one group of people. | An algorithm learns to prioritise patients it predicts to have better outcomes for a particular disease. This turns out to have a discriminatory effect on people within the Black and minority ethnic communities. |
| | Transformative effects | Algorithmic activities, like profiling, re-conceptualise reality in unexpected ways. | An individual using a personal health app has limited oversight over what passive data it is collecting and how that is being transformed into a recommendation to improve, limiting their ability to challenge any recommendations made and a loss of personal autonomy and data privacy. |
| **Traceability concerns** | | The harm caused by algorithmic activity is hard to debug (to detect the harm and find its cause), and it is hard to identify who should be held responsible for the harm caused. | If a decision made by clinical decision support software leads to a negative outcome for the individual, it is unclear who to assign the responsibility and/or liability to and, therefore, to prevent it from happening again. |

Table 3. High-level summary of ethical risks posed by the introduction of AI into healthcare adapted from Morley and Floridi (2024).[9]

| **Ethical concern** | **Related Ethical Principles** | **Description of Relationship** |
|---|---|---|
| Inconclusive Evidence | Reliability<br>Safety<br>Non-Maleficence<br>Integrity | Probabilistic and potentially hallucinated outputs from AI systems challenge the principle of reliability, which requires consistent and dependable performance. They also raise safety concerns, potentially violating non-maleficence when incorrect outputs lead to harmful clinical decisions. Scientific integrity is compromised when evidence lacks sufficient rigor to support clinical claims. |
| Inscrutable Evidence | Transparency<br>Explainability<br>Verifiability<br>Auditability | The “black box” nature of many AI algorithms directly conflicts with transparency and explainability principles. When recipients cannot understand how decisions are reached, verifiability is impossible, preventing users from confirming the accuracy or appropriateness of recommendations. Auditability becomes impractical when the data points and reasoning behind decisions remain hidden. |
| Misguided Evidence | Fairness<br>Equity<br>Representativeness<br>Inclusivity | Biased training data leads to outputs that systematically disadvantage certain populations, violating fairness principles. Equity is compromised when AI performs differently across populations. Representativeness and inclusivity require diverse data that accurately reflects all potential users, which is often lacking in healthcare AI. |

| | | |
|---|---|---|
| Unfair Outcomes | Justice<br>Non-discrimination<br>Equality<br>Beneficence | Disparate impacts across demographic groups directly violate justice and non-discrimination principles. AI systems fail to uphold equality when they provide better care to specific populations. Beneficence requires that AI improve healthcare for all patients, not just those well-represented in training data. |
| Transformative Effects | Autonomy<br>Dignity<br>Human agency<br>Self-determination | AI-driven reconceptualization of health undermines patient autonomy when individuals cannot meaningfully challenge algorithmic recommendations. Human dignity requires respect for individuals beyond their data profiles. Patient agency and self-determination are compromised when AI systems shape care pathways without transparent justification or patient input. |
| Traceability issues | Accountability<br>Responsibility<br>Governance<br>Recourse | Unclear responsibility for AI-driven outcomes fundamentally conflicts with accountability principles. Responsible innovation requires clear attribution of liability. Effective governance becomes impossible without traceability. Patient recourse—the ability to seek redress for harm—is compromised when responsibility cannot be assigned. |

Table 4. Ethical risks highlighted in Table 3 mapped to common ethical principles.

| **Academic Example: FUTURE-AI** | **Stakeholder Example: CHAI** | **Medical Association Example: American National Academy of Medicine** | **National Government Example: Pan-Canadian AI for Health (AI4H) Guiding Principles** | **International Organisation Example: World Health Organization (WHO) Ethics and Governance of AI Guidance** |
|---|---|---|---|---|
| The FUTURE-AI Consortium, comprised of 117 interdisciplinary experts from 50 countries, was founded in 2021. Over two years, the consensus-driven FUTURE-AI guideline was developed based on six guiding principles:[10]<br><br>1. **F**airness<br>2. **U**niversality<br>3. **T**raceability<br>4. **U**sability<br>5. **R**obustness<br>6. **E**xplainability | The Coalition for Health AI (CHAI) was created to bring together diverse stakeholders to listen, learn, and collaborate to drive the development, evaluation, and appropriate use of AI in healthcare. In 2024, CHAI produced the Responsible AI Guide around a set of core principles (CHAI, 2025):<br><br>1. Usefulness, usability, and efficacy<br>2. Fairness and equity<br>3. Safety and reliability | In 2024, the American National Academy of Medicine produced the draft AI Code of Conduct framework, intended to be used by developers, researchers, health systems, payers, patients, and federal agencies, and comprised of a harmonised set of principles and a distilled set of simple rules:[12]<br><br>1. Engaged (understanding, expressing, and prioritizing the needs, preferences, and goals of people) | The Pan-Canadian AI for Health (AI4H) Guiding Principles, updated in early 2025, outline shared values intended to guide Federal, Provincial, and Territorial (FPT) governments' efforts to support the responsible and ethical adoption and use of AI technologies across Canada's health systems:[13]<br><br>1. Person-centricity<br>2. Equity, diversity, and inclusion<br>3. Privacy and security<br>4. Safety and oversight | The WHO produced guidance on the ethics and governance of AI in 2021, based on six principles (WHO, 2021):<br><br>1. Protecting human autonomy<br>2. Promoting human well-being and safety and the public interest<br>3. Ensuring transparency, Explainability, and intelligibility<br>4. Fostering responsibility and accountability<br>5. Ensuring inclusiveness and equity<br>6. Promoting AI that is responsive and sustainable |

| | | | | |
|---|---|---|---|---|
| | 4. Transparency, Intelligibility, and Accountability<br>5. Security and Privacy | 2. Safe<br>3. Effective<br>4. Equitable<br>5. Efficient<br>6. Accessible<br>7. Transparent<br>8. Accountable<br>9. Secure<br>10. Adaptive | 5. Accountability and responsibility<br>6. Transparency and understandability<br>7. AI literacy<br>8. Robust data and data practices<br>9. Indigenous-led governance and data sovereignty | |

Table 5. Example principle-based ethics frameworks for AI in healthcare.

| **Principle 1** | **Principle 2** | **Tension** |
|---|---|---|
| Explainability | Accuracy | Transparent models sometimes sacrifice statistical performance[15] |
| Explainability | Autonomy | Excessive explanation can, paradoxically, increase automation bias[16] |
| Privacy | Equity | Simultaneously achieving group-level benefits (requiring access to broad datasets) and protecting individual patient privacy can be difficult |
| Accuracy | Accessibility | Determining appropriate thresholds for statistical performance in different contexts can be difficult, i.e., must AI exceed human performance in regions where skills of human care are already scarce |
| Accuracy | Autonomy (intellectual property protection) | Open-source models may outperform closed-source models yet pose a greater threat to autonomy operationalised via intellectual property. This is a particularly pressing problem for generative AI[17] |

Table 6. Examples of frequently occurring ethical tensions.

| **Infrastructure Type** | **Key Examples** | **Purpose** |
|---|---|---|
| Electronic Health Records | Singapore has an ongoing generative AI project that aims to automate record updating across the healthcare system before the end of 2025.[18]<br><br>As part of Saudi Vision 2030, Saudi Arabia introduced the Unified Health File, which aims to unify medical records across the healthcare system (Healthcare World, 2024; الصحة, 2025).<br><br>Following the publication of National Health Stack (2018) and National Digital Health Blueprint (2019), India has introduced Healthlocker – an electronic national health registry and cloud-based data storage system | Improving data availability by digitizing patient records and reducing reliance on paper-based systems. |

| | | |
|---|---|---|
| | designed to serve as a single source of health data for the country.[21] | |
| Data Access Environments | In the UK, the national network of secure data environments, the Health Data Research UK data gateway, and the Federated Data Platform all aim to facilitate access to electronic health record data, and administrative data, for research and analytics purposes.<br><br>India is implementing a federated Personal Health Records (PHR) framework to enable secure and accessible digital health records for citizens, facilitating data sharing with healthcare providers and supporting medical research.[22] | Improving data accessibility by supporting privacy-preserving research access to healthcare data. |
| Interoperability Initiatives | Brazil's National Health Data Network (RNDS)[23] introduced in 2019, is an interoperability platform designed to promote information exchange between services across healthcare networks.<br><br>The Saudi Health Information Exchange (SeHE) is part of the National Platform for Health and Insurance Exchange Services (NPHIES). It aims to enable the electronic sharing of health data across medical facilities, health organisations, and government agencies.[20]<br><br>The European Health Data Space[24] Regulations aim to establish a common framework for the use and exchange of electronic health data across the EU. | Improving data quality by enabling seamless data exchange between different healthcare systems – facilitating the creation of larger, more representative, datasets. |
| Analytics Platforms | Phase 2 of Uruguay's national electronic health record system (HCEN) aims to create an analytics platform, that will be capable of facilitating advanced data analytics to support decision-making for public health policies, and evaluating the quality of services provided to users within the healthcare system.[25]<br><br>India's National Data and Analytics Platform (NADP) facilitates access to government data, including datasets related to health and family welfare.[26] | Improving real-time data availability by providing specialised environments for AI implementation into frontline care. |

Table 7. Example investments in data infrastructure made by countries to overcome barriers to AI benefits-realization related to the availability, accessibility, and quality of healthcare data.

| Initiative Type | Key Examples | Purposes |
|---|---|---|
| Multifaceted educational program | Digital, Artificial Intelligence and Robotics (DART-Ed): Introduced by Health Education England following the 2019 *"Topol Review: Preparing the Workforce for a Digital Future"*,[27] with the intention<br><br>Colombian Association of Radiology's AI Committee: Comprising an interdisciplinary group of specialists in ethics, bioethics, AI, and machine learning, the committee's primary goal is "to promote the ethical, safe, and effective use of AI in health care" through national educational initiatives, including a national congress on radiology AI and machine learning, AI foundations for imaging technologists, and subspecialty-specific practical conferences.[28] | Developing the knowledge, skills, professional attributes, and behaviors the healthcare workforce needs to use, design, implement, and use AI. |
| Professional development event | Lithuanian University of Health Sciences (LSMU): Hosted the country's first international scientific conference on "Artificial Intelligence in Healthcare" in November 2024. The event, which included theoretical presentations and practical workshops on applying AI, brought together medical doctors, pharmacists, and other healthcare professionals.[29] | Providing healthcare providers with a means of learning key AI capabilities and healthcare managers with a means of learning what AI tools are available to help their organizations with key efficiency-related problems. |

Table 8. Examples of educational initiatives developed by countries to ensure clinicians have the skills, knowledge, and confidence to assess AI outputs critically.

| Document Name | Description | Key References |
|---|---|---|
| Generating Evidence for Artificial Intelligence-Based Medical Devices: A Framework for Training, Validation and Evaluation | Produced by the WHO in 2021, this framework provides an overview of considerations used in evaluating clinical evidence regarding AI, aiming to help formulate a consensus for guiding validation, evidence generation, and reporting across the total product lifecycle within a global health context. | 30 |
| Focus Group on Artificial Intelligence for Health (FG-AI4H) Del 7.4: Clinical Evaluation of AI for Health | Produced by members of the WHO/ International Telecommunication Union (ITU) Focus Group on AI for Health the framework for clinical evaluation divides evaluation into four phases: evaluation of model purpose and suitability; algorithmic validation; clinical validation; and ongoing monitoring. | 31,32 |

Table 9. Example of guidance setting out the standardised steps involved in evaluating the safety and effectiveness of AI tools.

| Evaluation methodology | Description | Key References |
|---|---|---|
| Clinical simulation | A controlled method of recreating healthcare scenarios to assess AI tools without patient risk. Clinicians engage with simulated cases ranging from basic lab tasks to high-fidelity environments, allowing researchers to measure how AI affects clinical decisions and actions. | 33 |
| Agent-based modeling | A computational simulation method creating dynamic clinical environments where autonomous AI agents interact with virtual stakeholders. Simulates patient-physician interactions and hospital processes to evaluate how LLMs handle tools, data, and decision-making without patient risk. | 34 |
| AI Structured Clinical Examinations (AI-SCE) | An evaluation framework adapting clinical education assessment principles to measure LLM performance in healthcare scenarios. It assesses both final outputs and intermediate reasoning steps through standardised clinical challenges derived from complex real-world cases. | 34 |
| Translational Trials | A prospective, non-interventional evaluation method where AI systems operate in their intended clinical environment without affecting patient care. The AI runs parallel to standard practice, assessing model performance, workflow integration issues, and implementation challenges in real-world settings. | 35 |
| Causal Off-Policy Evaluation (OPE) | A methodology that enables developers to retrospectively simulate the effect of an AI intervention without actually deploying it in clinical practice. | 36 |

Table 10. Examples of innovative methodologies for evaluating the safety and effectiveness of AI tools.

| Initiative | Description | Key References |
|---|---|---|
| Good machine learning practice for medical device development: guiding principles | Produced by the International Medical Device Regulators Forum, the 10 guiding principles for Good Machine Learning Practice (GMLP) are a call to action to international standards organizations, international regulators, and other collaborative bodies to further advance GMLP:<br><br>1. The intended use/intended purpose of the device is well understood, and multidisciplinary expertise is leveraged throughout the total product life cycle<br>2. Good software engineering, medical device design, and security practices are implemented throughout the total product life cycle<br>3. Clinical evaluation includes the use of datasets that are representative of the intended patient population<br>4. Training datasets are independent of test sets<br>5. Selected reference standards are fit-for-purpose | 37 |

| | | |
|---|---|---|
| WHO Regulatory Considerations on AI for Health | 6. Model choice and design are tailored to the available data and intended use/ intended purpose of the device | 38 |
| Resolution RDC No. 657/2022, regulating Software as a Medical Device (SaMD) (Brazil) | 7. The AI device is assessed with a focus on human-AI interactions in the intended use environment, including the performance of the human-AI team, rather than just the device in isolation | 28 |
| Decree of the Government of the Russian Federation No. 1906280 of 24.11.2020 | 8. Testing demonstrates device performance during clinically relevant conditions | 39 |
| Medical Device Regulations (EU) | 9. Users are provided with clear, essential information | 40 |
| FDA Action Plan for AI/ML-Based Software (US) | 10. Deployed models are monitored for performance, and re-training risks are managed | 41-43 |
| MHRA Software and AI as a Medical Device Change Programme (UK) | Published in September 2021 and updated in June 2023, establishes a progressive framework for regulating AI as medical devices. The roadmap establishes eleven work packages across three strategic themes (enhanced pre-market requirements, enhanced post-market oversight, and creating a supportive environment) to address the unique characteristics of AI medical devices | 44 |
| Saudi Food & Drug Authority Guidance on Artificial Intelligence (AI) and Machine Learning (ML) technologies based Medical Devices (the MDS-G010) | Introduced in November 2022 to clarify the requirements for obtaining Medical Device Marketing Authorization (MDMA) for AI/ML-based medical devices to place them on the market, and creates classification criteria and control methods for AI-based medical device | 45 |
| DASH for SaMD (Japan) | Launched by the Japanese Pharmaceutical Safety and Environmental Health Bureau and the Ministry of Health, Labour and Welfare in 2020, the DX(digital transformation) action strategies in healthcare for software as a medical device program is composed of four workstreams which, amongst other things, are designed to: publish guidance regarding the characteristics of SaMD; and establish a post-approval change management protocol. | 46,47 |
| Guidelines for the Classification and Definition of AI Medical Software Products and Guidelines for the Registration and Review of Medical Device Software (China) | Defines risk-based classifications for AI-based medical device software | 48 |

Table 11. Examples of initiatives designed to expand and update software as a medical device regulations to ensure they are fit for AI, including generative AI.

| Initiative | Country | Description |
|---|---|---|
| AI and Data Act (AIDA) | Canada | Introduced in June 2022. Proposes a risk-based framework for regulating AI systems, including in healthcare. |
| Federal Law Regulating AI | Mexico | Introduced in 2024. Aims to establish a legal framework for developing, deploying, and using AI technologies across various sectors, including healthcare. |
| AI Framework Bill | Brazil | Proposed in May 2023. Intended to establish a horizontal regulatory framework for AI, adapting a risk-based approach similar to the EU AI Act and implementing a dedicated regulatory authority. |
| Bill No 16821-19 | Chile | Introduced in May 2024. Adopts a risk-based classification similar to the EU's AI Act and imposes obligations on high-risk applications. Like the AI Act, these obligations relate to data governance, transparency, and human oversight. |
| Bill 3003-d-2024 | Argentina | Introduced in June 2024. Aims to establish a legal framework for the responsible use of AI. |
| Law 059/23 | Colombia | Introduced in November 2023. Establishes guidelines for the development, use, and regulation of AI. |
| Organic law for the regulation and promotion of AI | Ecuador | Introduced in July 2024. Establishes a risk-based framework for the safe and ethical use of AI |
| National AI Roadmap | Malaysia | Introduced in 2023. This led to the creation of a national AI office in November 2024. The office aims to oversee policy formulation and regulatory frameworks for AI. |
| Artificial Intelligence Regulation Law | Bahrain | Introduced in April 2024. Includes 38 articles on privacy, unlicensed activities, and the risks of AI systems |
| Policy on AI regulation and ethics | Israel | Introduced in December 2023. Adopts a risk-based, sector-specific approach aligned with the EU AI ACT and emphasises human-centric AI, equality, transparency, accountability, and system reliability, |

Table 12. Examples of international initiatives to develop horizontal AI regulation drawn from.[48]

| Infrastructure component | Description |
|---|---|
| Hardware and software computing infrastructure | Frontend and backend hardware required to help develop and deploy AI, e.g., data access environments (backend) and electronic health record systems within which AI-enabled clinical decision support can be embedded (frontend) |
| Clinical content | The structured and unstructured information (data) recorded by the healthcare system, such as lists of symptoms, diagnoses, prescriptions, and images, and used to train and inform AI tools |
| Human-Computer Interface | The interface clinicians or patients use to enter information into an AI tool. |

| | |
|---|---|
| People | AI developers, clinicians, patients, and other stakeholders |
| Workflow and communication | Decisions regarding, for example, when AI alerts work during a clinical consultation, how it is acted upon, how it is interpreted, and how the results are communicated to patients. |
| Internal (to individual healthcare organisations) policies, procedures, and cultures | For example, the protocols governing the development of AI range from data collection to ongoing monitoring and organisational and system values. |
| External rules, regulations, and pressures | Including data protection regulation, medical device regulation, consumer protection regulation, and discrimination regulation. |
| System measurement and monitoring | Including post-market surveillance mechanisms, and overarching indicators of success |

Table 13. Infrastructure components used to inform the development of necessary but not sufficient requirements for healthcare AI benefits realisation, adapted from.[49]

| **Intervention** | **Purpose** |
|---|---|
| Develop speciality-specific EHR templates | EHRs are designed with and for clinicians, with usability and interoperability in mind<br>EHR design permits unrestricted clinical information recording<br>EHR design makes the capture of high-quality data a natural by-product of clinical work<br>Clinicians only have to learn how to use one form of interface |
| Prioritise Model-to-Data (MTD) [50] data access mechanisms | MTD mechanisms facilitate the development of AI without granting developers direct access to clinical data<br>Healthcare systems rely more heavily on secure data environments (or trusted research environments)<br>The need to move data away from its source location is minimised, and data can be extracted more readily from centrally controlled, well-curated, harmonised, AI-ready data assets<br>Data access is more auditable, and governance rules are technically implemented (not paper-based) and more auditable<br>Cross-organisational data access is streamlined<br>Healthcare systems rely less on insecure deidentification and dissemination mechanisms of data sharing |
| Mandate compliance with interoperability standards | Clinicians do not have to log into multiple systems<br>Data exchange protocols are standardised<br>AI can be more seamlessly integrated with EHR workflows<br>Aggregation of AI 'results' is easier, facilitating system-wide impact analysis<br>Clinicians can interact with one, well-designed, user-friendly interface embedded within an EHR |
| Establish international protocols to mitigate the risk of data colonialism and ensure data sovereignty | Local communities maintain ownership and control over their health data<br>Benefits from data usage flow back to source communities<br>Avoids exploitative practices through transparent data governance<br>Ensures that IA development serves local needs rather than extracting value<br>Builds trust and sustainable relationships with data-providing communities |

| | |
|---|---|
| Mandate modular design and local customisability at both AI and supporting infrastructure level | Individual healthcare organisations can select – in a plug-and-play fashion – the exact AI 'modules' they wish to use<br>Help prevent clinicians from becoming overwhelmed by or overly dependent on AI and can exert a degree of independent control over what they want to use AI for and when<br>Different components of the overarching AI 'system' (e.g., interface, knowledge base, EHR integration) can be 'swapped in and out' without enforced obsolescence |
| Integrate AI with audit and feedback mechanisms | AI advice given to clinicians can be contextualised to the specific clinician<br>Clinicians can provide feedback on the usability and utility of individual AI<br>Facilitate the development of a two-way partnership between AI and clinician so that the clinician remains firmly in the loop and the interactions between the two are more conversational than algorithmically paternalistic. |
| Invest in the development of transfer learning and domain adaptation techniques | Enables AI models trained on data-rich contexts to be effectively adapted to data-sparse settings<br>Reduces the amount of local data needed for effective AI implementation<br>Facilitates AI deployment in underrepresented regions and for underserved populations<br>Promotes more efficient use of available data resources<br>Helps address data inequities while respecting data sovereignty |
| Explore alternative data sources for underrepresented regions | Utilises available data sources like satellite imagery, street view data, or public social media in data-sparse settings<br>Supplements traditional clinical data with contextual information<br>Enables AI development in regions with limited healthcare data infrastructure<br>Creates innovative approaches to health monitoring and intervention<br>Bridges data gaps while working to build more comprehensive health data systems |

Table 14. Example policy, legislative, regulatory, or strategic interventions for operationalising the core requirement regarding robust data exchange.

| **Intervention** | **Purpose** |
|---|---|
| Value data work and invest in the creation of AI-ready datasets | Healthcare organisations can signpost AI developers to known data assets that have been carefully curated to meet specific clinical, or system needs.<br>Healthcare organisations can tell AI developers upfront what the specific limitations of specific data assets are for specific uses.<br>Healthcare organisations are aware of the underpinning assumptions of any AI-ready data assets, and know of any processing conducted to overcome issues such as missingness that might impact the accuracy of the dataset for the specific purpose.<br>Healthcare organisations can ensure key data assets are not used for purposes that would be inappropriate, i.e., for uses for which the data would not be suitable. |

| | |
|---|---|
| Make interpretability a standardised (ideally mandatory) requirement of AI | Clinicians can verify during the consultation whether the recommendation provided by AI is based on information relevant to the specific information need<br>AI does not create an informational asymmetry between patients, clinicians, and itself |
| Protect the clinician's right to override AI advice | Clinicians only act on the advice of AI if it meets a clinical need in a way that they feel is reasonable, accurate, and in keeping with the best interests of their patient<br>Trust builds in clinician-patient interactions and is maintained over time<br>AI use supports the 'true' meaning of evidence-based medicine, i.e., evidence that is contextualised, interpreted, and combined with lived experience<br>The complexity and variability of healthcare is never underestimated nor presumed to be reducible to simple if-this-then-that rules |
| Incorporate data literacy into medical education and continuing professional development | Clinicians can understand statistical concepts, algorithmic limitations, and appropriate AI outputs<br>Clinicians feel confident in evaluating AI outputs and exerting their authority |
| Position clinicians as the primary drivers of AI demand | All AI development is tied to a very clear, specific, clinical or broader healthcare system need that clinicians believe will be met (beneficially) through the implementation of AI<br>All AI is seen as relevant and essential by clinicians<br>Clinicians do not perceive AI as timewasting<br>All AI is designed to provide information that clinicians deem to be vivid and salient for a specific purpose<br>The risk of alert fatigue and information chaos is reduced |
| Create clear 'warning' labels for AI in use | Clinicians are aware if AI is being used for a patient that would count as an 'out-of-sample' patient for that specific use case |
| Create design templates for human-AI collaboration | Shifts evaluation from "human versus AI" to "human plus AI versus human alone"<br>Focuses on process design in addition to algorithm performance<br>Optimises the interaction between human judgment and AI recommendations<br>Accommodates varying levels of human expertise and experiences in different contexts |
| Conduct root cause analysis to help address distrust | Goes beyond surface-level trust-building to understand underlying reasons for distrust<br>Addresses historical and systemic factors affecting technology acceptance<br>Creates meaningful accountability for past harms or mistakes<br>Develops targeted interventions based on specific community concerns<br>Enables more authentic and sustainable trust relationships |

Table 15. Example policy, legislative, regulatory, or strategic interventions for operationalising the core requirements 'epistemic certainty' and 'autonomous staff and patients'.

| Intervention | Purpose |
|---|---|
| Mandate AI is built by teams with diverse skills and lived experiences | AI is designed to be capable of recognising different cultural and personal values and healthcare priorities<br>AI meets the needs of a range of patients<br>The risk of AI amplifying healthcare inequalities is mitigated by social as well as technical means |
| Mandate the involvement of patients and the public in all stages of the AI development, deployment, and use pipeline | Patients and the public are treated as 'true interlocutors' in the identification of suitable clinical and healthcare system needs for AI to solve, rather than a problem to overcome.<br>The social license to develop and implement AI is maintained<br>Patient autonomy is protected from the beginning of the AI development process<br>AI developers are made aware of their moral duty to patients and are not so distanced that the impact of their product is invisible. |
| Prioritise the collection, curation, standardisation, and integration of 'qualitative' health data (e.g., patient-reported outcomes) | Patient outcomes are considered as important as quantitative or measurable outcomes in the monitoring of AI effectiveness<br>'Soft' aspects of clinical needs, including the need to meet these needs in an empathetic and contextualised fashion, are given due attention in the development of AI<br>Advice produced by AI is predicated on 'capability', not *just* efficacy, i.e., what the real options are for the specific patient, for treating their conditions, or improving their health, based on their circumstances, values, and preferences<br>AI is positioned as a digital companion to patients rather than a source of algorithmic paternalism. |
| Prioritise AI development for conditions with high unmet needs | Shifts focus from well-served conditions to areas of greater need<br>Ensures AI contributes to addressing healthcare gaps rather than reinforcing them<br>Directs resources toward the highest-impact applications<br>Creates incentives for developers to address neglected conditions<br>Maximises the equity-enhancing potential of healthcare AI |
| Integrate AI initiatives with broader health system strengthening | Ensures AI deployment is coordinated with physical infrastructure improvements<br>Recognises that AI effectiveness depends on the broader healthcare context<br>Creates synergies between digital and physical healthcare resources<br>Avoids technology-first approaches disconnected from healthcare realities |
| Develop coordinated policies connecting AI to social determinants of health | Recognises that AI effectiveness depends on broader social and economic factors<br>Coordinates technology policies with social safety net provisions<br>Addresses underlying healthcare inequities that technology alone cannot solve<br>Creates holistic approaches to health improvement beyond digital solutions<br>Maximises the impact of AI by ensuring basic health needs are addressed |

Table 16. Example policy, legislative, regulatory, or strategic interventions for operationalising the core requirement 'actively protected values'.

| Intervention | Purpose |
|---|---|
| Broaden and standardise the definition of AI as a medical device, so as often as possible, AI used in clinical care is classified as Software as a Medical Device, registered, and regularly reported publicly | As often as possible, AI is subject to at least high-level regulatory scrutiny before being used in patient care<br>As often as possible, AI is legally bound to an intended use<br>As often as possible AI developers are legally required to generate high-quality, independently validated and verified evidence of efficacy as often as possible.<br>As often as possible, AI is regularly monitored once it is deployed to ensure that it is continuing to meet its information needs safely, effectively, and for all patients<br>As often as possible responsibility for errors in AI outputs that result in patient harm is clarified, and the pathway to rectification is clear |
| Re-assess the risk classification of all Software as a Medical Device | AI, including generative AI, is not considered to be low risk simply because it is 'information only' and not a physical intervention |
| Adapt Health Technology Assessment (HTA) processes to be suitable for AI and ensure AI is subject to HTA before being procured by individual healthcare organisations (where applicable) | AI is, from the outset, linked clearly to a stated outcome, its contribution to which can be quantified via a transparent, independent review of its evidence of efficacy<br>AI is evaluated from a standardised, and contextualised cost-effectiveness perspective<br>Funding is made available for developing appropriate methods for AI HTA<br>All AI is subject to open public debate before being rolled out at a national or international level |
| Ensure processes for testing robustness, reliability, accuracy, efficacy, and safety are centralised, streamlined, and transparent. | All AI is subject to the same process (subject to standardised variation for model type) of internal validation, temporal validation, external validation, and local calibration<br>All AI is held to the same performance standard and meets its specific clinical or healthcare system needs safely, robustly, and reliably<br>AI developers cannot unfairly benefit from a process of 'validation and evaluation shopping' where they choose the testing centre that has the most favourable (i.e., easy) validation and evaluation process<br>AI developers cannot fake or exaggerate claims of efficacy, safety, or robustness<br>All patients and clinicians can access a verifiable record of the performance of any AI that is used in their care<br>AI developers cannot solely rely on reporting basic statistical performance measures, such as AUROC, as indicators of clinical efficacy<br>Individuals responsible for commissioning and procuring AI within their organisation do not feel solely responsible (and thus apprehensive) about assessing whether AI being sold to them is genuinely safe, effective, and robust |
| Establish a body for ongoing monitoring of AI in use | (At minimum) A national register of all AI in use is maintained<br>Mechanisms are put in place for monitoring ongoing changes in performance<br>Mechanisms are put in place to alert AI developers and regulators to any emerging safety issues |
| Simplify rules regarding data ownership, controllership, use case and consent | All decisions regarding what healthcare data can be used for, who by, and how it can be accessed are standardised AI developers cannot 'game' the system, for example, gaining access to data under false pretences |

| | |
|---|---|
| | Various Information Governance frameworks are more interpretable, understandable, and transparent to all, improving patient and public trust, and mitigating the likelihood that AI developers and healthcare organisations may make well-meaning but ill-informed decisions regarding data access and use. |
| Revise and enhance the liability, discrimination, and consumer protection components of process accountability | All clinicians are aware of the edges of their responsibility, as are AI developers<br>All patients are aware of what they should expect from clinicians using AI and how to seek compensation if harm occurs<br>AI users are encouraged to remain aware of the possibility of automation bias, and the importance of continuing to act as knowledgeable intermediaries even when using AI<br>AI users are discouraged from using AI for unvalidated purposes<br>No AI developers, users, or even beneficiaries are under the impression that AI is 'just' software incapable of causing harm |
| Mandate that all decisions along the AI development pipeline be clearly and publicly documented | Any patient, member of the public, clinician, or other stakeholder can verify any claims made by AI developers or others regarding the efficacy, safety, reliability, or even utility of any specific instance of AI<br>The rationale for developing and deploying any AI is publicly accessible to encourage transparency regarding decisions to implement<br>All AI is clearly and transparently tied to a use case for which there is a clear, robust, and publicly accessible evidence-base<br>All patients can be reassured of its safety, efficacy, and cost-effectiveness, and, therefore, of the healthcare system's commitment to 'do no harm'<br>The entire pipeline of AI development is auditable, and bad actors can be identified and sanctions |
| Develop standardised evaluation frameworks for AI studies and implementations | The use of tools like APPRAISE-AI [51]that provide quantitative evaluation across domains, including clinical relevance, data quality, methodological conduct, robustness, reporting quality and reproducibility, is encouraged<br>Establishes consistent quality benchmarks across different AI implementations<br>Identifies common shortcomings (e.g., data sources, sample sizes, bias assessment, error analysis)<br>Facilitates comparison across different contexts and implementations |
| Develop collaborative international frameworks for interoperable AI regulation | Reduces regulatory fragmentation across borders<br>Facilitates faster and more consistent approval processes<br>Enables sharing of post-market surveillance data<br>Prevents regulatory arbitrage while respecting the local healthcare context<br>Encourages adoption of best practices globally |
| Invest in centres of excellence for regulatory science | Develops innovative mechanisms for monitoring "on-market adaptive AI' approved through predetermined change plans<br>Develops innovative mechanisms for supporting breakthrough and fast-track programs for technologies addressing unmet needs.<br>Establishes mechanisms such as "airlock classifications" for temporary market entry with enhanced monitoring to enable more real-time testing |

| | |
|---|---|
| Establish frameworks for validating, piloting, and scaling AI tools in real-world settings | Creates clear pathways from research to implementation<br>Addresses the "AI graveyard" problem where promising solutions fail to scale<br>Provides structured approaches for translating research into practice<br>Enables early identification of implementation barriers<br>Facilitates knowledge sharing about successful implementation strategies |
| Embed sociotechnical evaluation in all relevant evaluation and regulatory requirements | Ensures relevant evaluations:<br>Assess how AI interacts with existing social and professional practices<br>Identifies potential unintended consequences before deployment<br>Compliment technical performance metrics with social impact metrics |
| Design AI evaluations that consider "human plus algorithm" rather than "human versus algorithm" | Recognise that a systems approach, rather than a product-focused view, is critical for AI, where performance can vary dramatically between testing environments and clinical settings, primarily due to human factors. |

Table 17. Example policy, legislative, regulatory, or strategic interventions for operationalising the core requirements of 'validated outcomes' and 'meaningful accountability.'

| **Intervention** | **Purpose** |
|---|---|
| Mandate environmental impact assessments for healthcare AI | Ensures developers measure and report on the energy consumption and carbon footprint of their systems<br>Creates awareness of environmental costs alongside clinical benefits<br>Establishes baselines for improvement<br>Enables comparison between different approaches<br>Encourages optimising for energy efficiency alongside accuracy |
| Establish energy efficiency standards for healthcare AI | Limits unnecessary computational resource usage<br>Promotes model compression and optimisation techniques<br>Encourages reuse of pre-trained models where appropriate<br>Incentivises energy-efficient hardware for deployment<br>Balances performance needs with sustainability considerations |
| Require lifecycle planning for healthcare AI hardware | Ensures responsible sourcing of materials for AI infrastructure<br>Plans for appropriate hardware reuse, recycling, and disposal<br>Minimises e-waste from rapid technology turnover<br>Extends the life of computational resources through maintenance and upgrading<br>Reduces the overall environmental footprint of AI healthcare systems |
| Develop frameworks for prioritising high-impact use cases | Focuses computational resources on applications with the most significant health impact<br>Discourages deployments with minimal or incremental benefits<br>Ensures health gains justify environmental costs<br>Promotes resource sharing for common AI tasks across healthcare systems<br>Avoids duplicative or redundant AI development |

| | |
|---|---|
| Incorporate sustainability metrics into healthcare AI procurement frameworks | Makes environmental impact a factor in purchasing decisions<br>Creates market incentives for sustainable AI development<br>Enables healthcare systems to align technology choices with sustainability goals<br>Encourages transparency in reporting environmental impacts<br>Supports organizational climate commitments |
| Develop mobile-first and offline-capable AI applications | Ensures accessibility for patients in remote or connectivity-challenged areas<br>Accommodates varying levels of digital infrastructure<br>Makes AI tools usable on widely available consumer devices<br>Reduces barriers to adoption in resource-constrained settings<br>Bridges the digital divide in healthcare access<br>Helps to mitigate environmental impact |

Table 18. Example policy, legislative, regulatory, or strategic interventions for operationalising the core requirement of environmental sustainability.

## References


1. Karpathakis, K., Morley, J. & Floridi, L. A Justifiable Investment in AI for Healthcare: Aligning Ambition with Reality. *Minds & Machines* 34, 38 (2024).
2. Sgaier, S. & Dominici, F. Using AI to Understand What Causes Diseases. *Harvard Business Review* https://hbr.org/2019/11/using-ai-to-understand-what-causes-diseases (2019).
3. Natarajan, S., Jain, A., Krishnan, R., Rogye, A. & Sivaprasad, S. Diagnostic Accuracy of Community-Based Diabetic Retinopathy Screening With an Offline Artificial Intelligence System on a Smartphone. *JAMA Ophthalmol* 137, 1182 (2019).
4. Adedinsewo, D. A. *et al.* Contextual challenges in implementing artificial intelligence for healthcare in low-resource environments: insights from the SPEC-AI Nigeria trial. *Front. Cardiovasc. Med.* 12, 1516088 (2025).
5. China Announces the World's First AI Hospital, Marking Asia's Leadership in Healthcare Innovation. *MedTech World* https://med-tech.world/news/china-worlds-first-ai-hospital-milestone-in-healthcare-innovation/.
6. Yee, H.-J., Boo, I., Tan, I. K. T., Tan, J. S. & Zakariah, H. Determining latent features and forecasting of COVID-19 hospitalisations in Malaysia using a national patient assessment data platform: a study of machine learning modelling against expert system. Preprint at https://doi.org/10.1101/2023.01.17.22281858 (2023).
7. Tomašev, N. *et al.* A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 572, 116–119 (2019).
8. AstraZeneca. AstraZeneca continues to drive healthcare innovation with AI X-Ray for lung cancer diagnosis. https://www.astrazeneca.my/media/press-releases/2021/astrazeneca-continues-to-drive-healthcare-innovation-with-ai-x-ray-for-lung-cancer-diagnosis.html (2021).
9. Morley, J. & Floridi, L. The Ethics of AI in Health Care: An Updated Mapping Review. Preprint at https://doi.org/10.2139/ssrn.4987317 (2024).
10. Lekadir, K. *et al.* FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ* e081554 (2025) doi:10.1136/bmj-2024-081554.
11. CHAI. Responsible AI Guide. *CHAI - Coalition for Health AI* https://chai.org/responsible-ai-guide/ (2025).
12. Adams, L. *et al.* Artificial Intelligence in Health, Health Care, and Biomedical Science: An AI Code of Conduct Principles and Commitments Discussion Draft. *National Academy of Medicine* https://nam.edu/perspectives/artificial-intelligence-in-health-health-care-and-biomedical-science-an-ai-code-of-conduct-principles-and-commitments-discussion-draft/#:~:text=The%20draft%20AI%20Code%20of,health%20systems%2C%20payers%2C%20patients%2C (2024).
13. Canada, H. Pan-Canadian AI for Health (AI4H) Guiding Principles. https://www.canada.ca/en/health-canada/corporate/transparency/health-agreements/pan-canadian-ai-guiding-principles.html (2025).

14. *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Executive Summary*. (World Health Organization, Geneva, 2021).
15. Cohen, I. G., Evgeniou, T., Gerke, S. & Minssen, T. The European artificial intelligence strategy: implications and challenges for digital health. *The Lancet Digital Health* 2, e376–e379 (2020).
16. Prinster, D. *et al.* Care to Explain? AI Explanation Types Differentially Impact Chest Radiograph Diagnostic Performance and Physician Trust in AI. *Radiology* 313, e233261 (2024).
17. Dennstädt, F., Hastings, J., Putora, P. M., Schmerder, M. & Cihoric, N. Implementing large language models in healthcare while balancing control, collaboration, costs and security. *npj Digit. Med.* 8, 143 (2025).
18. Ang, A. Singapore invests $150M for public health genAI adoption. https://www.healthcareitnews.com/news/asia/singapore-invests-150m-public-health-genai-adoption (2024).
19. World, H. HEALTHCARE AND AI IN THE KINGDOM OF SAUDI ARABIA. *Healthcare World* https://healthcareworld.com/healthcare-and-ai-in-the-kingdom-of-saudi-arabia/ (2024).
20. الصحة, ف. ب. و. Ministry Of Health Saudi Arabia. *Ministry Of Health Saudi Arabia* https://www.moh.gov.sa/en/Pages/Default.aspx (2025).
21. Parry, C. & Aneja, U. Artificial Intelligence for Healthcare: Insights from India | 3. AI in Healthcare in India: Applications, Challenges and Risks. https://www.chathamhouse.org/2020/07/artificial-intelligence-healthcare-insights-india/3-ai-healthcare-india-applications (2023).
22. NITI Aayog. *National Health Stack*. https://abdm.gov.in:8081/uploads/NHS_Strategy_and_Approach_1_89e2dd8f87.pdf#:~:text=a%20Federated%20personal%20health%20records%20(phr)%20Framework:,for%20advancing%20our%20understanding%20of%20human%20health;&text=This%20includes%20a%20Coverage%20and%20Claims%20Platform%2C,Natioaal%20Health%20Analytics%20Framework%20(NHAF)%20amongst%20others. (2018).
23. Barbalho, I. M. P. *et al.* Electronic health records in Brazil: Prospects and technological challenges. *Front. Public Health* 10, 963841 (2022).
24. European Commission. European Health Data Space Regulation (EHDS) - European Commission. https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en (2025).
25. United Nations Educational, S. and C. O. O. des N. U. pour l'éducation, la science et la culture Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura Организация Объединенных Наций по вопросам образования, науки и культуры منظمة الأمم المتحدة للتربية والعلم والثقافة 联合国教育、科学及文化组织 United Nations Educational, Scientific and Cultural Organization *et al.* Uruguay: artificial intelligence readiness assessment report. (2024).
26. NADP. Home - NDAP. https://ndap.niti.gov.in/ (2025).
27. Topol, E. & Health Education Endland. The Topol Review — NHS Health Education England. *The Topol Review — NHS Health Education England* https://topol.hee.nhs.uk/ (2019).

28. Kitamura, F. C. *et al.* Forging Connections in Latin America to Advance AI in Radiology. *Radiology: Artificial Intelligence* 4, e220125 (2022).
29. Smaidziunaite, M. Lithuania's First Conference on AI for Healthcare: Insights from Renowned International and Lithuanian Experts. *LSMU* https://lsmu.lt/en/lithuanias-first-conference-on-ai-for-healthcare-insights-from-renowned-international-and-lithuanian-experts/ (2024).
30. *Generating Evidence for Artificial Intelligence-Based Medical Devices: A Framework for Training, Validation and Evaluation*. (World Health Organization, Geneva, 2021).
31. Wiegand, T. *et al. Whitepaper for the ITU/WHO Focus Group on Artificial Intelligence for Health*. https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/FG-AI4H_Whitepaper.pdf (2020).
32. ITU. *Focus Group on Artificial Intelligence for Health (FG-AI4H) Del 7.4: Clinical Evaluation of AI for Health*. https://www.itu.int/dms_pub/itu-t/opb/fg/T-FG-AI4H-2023-3-PDF-E.pdf (2023).
33. Lyell, D., Lustig, A., Denyer, K., Vedantam, S. & Magrabi, F. Using Clinical Simulation to Evaluate AI-Enabled Decision Support. in *Studies in Health Technology and Informatics* (eds. Bichel-Findlay, J., Otero, P., Scott, P. & Huesing, E.) (IOS Press, 2024). doi:10.3233/SHTI230975.
34. Mehandru, N. *et al.* Evaluating large language models as agents in the clinic. *npj Digit. Med.* 7, 84 (2024).
35. McCradden, M. D. *et al.* CANAIRI: the Collaboration for Translational Artificial Intelligence Trials in healthcare. *Nat Med* 31, 9–11 (2025).
36. Joshi, S. *et al.* AI as an intervention: improving clinical outcomes relies on a causal approach to AI development and validation. *Journal of the American Medical Informatics Association* 32, 589–594 (2025).
37. IMDRF. Good machine learning practice for medical device development: Guiding principles | International Medical Device Regulators Forum. https://www.imdrf.org/documents/good-machine-learning-practice-medical-device-development-guiding-principles (2025).
38. WHO. *Regulatory Considerations on Artificial Intelligence for Health*. https://iris.who.int/bitstream/handle/10665/373421/9789240078871-eng.pdf. (2023).
39. Maria, A. & Ablameyko, S. AI image-based diagnosis systems: how to implement them? *JDH* 12–21 (2023) doi:10.55976/jdh.22023113912-21.
40. Zapata, K. A. C., Ward, T., Loughran, R. & McCaffery, F. A Review of the Artificial Intelligence Act Proposal and the Medical Device Regulation. in *2023 31st Irish Conference on Artificial Intelligence and Cognitive Science (AICS)* 1–6 (IEEE, Letterkenny, Ireland, 2023). doi:10.1109/AICS60730.2023.10470832.
41. Aquino, Y. S. J. *et al.* Defining change: Exploring expert views about the regulatory challenges in adaptive artificial intelligence for healthcare. *Health Policy and Technology* 13, 100892 (2024).
42. FDA. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. *FDA* (2024).
43. Warraich, H. J., Tazbaz, T. & Califf, R. M. FDA Perspective on the Regulation of Artificial Intelligence in Health Care and Biomedicine. *JAMA* 333, 241 (2025).

44. MHRA. Software and AI as a Medical Device Change Programme - Roadmap. *GOV.UK* https://www.gov.uk/government/publications/software-and-ai-as-a-medical-device-change-programme/software-and-ai-as-a-medical-device-change-programme-roadmap (2023).
45. Solaiman, B. Regulating AI-Based Medical Devices in Saudi Arabia: New Legal Paradigms in an Evolving Global Legal Order. *ABR* 16, 373–389 (2024).
46. MHLW. *DX(Digital Transformation) Action Strategies in Healthcare for Software as a Medical Device*. https://exporttojapan.co.uk/wp-content/uploads/2021/10/MHLW-Mr-Fukuta-Presentation-for-E2J.pdf (2020).
47. Tochibayashi, N. & Kutty, N. Three AI tools revolutionising healthcare in Japan. *World Economic Forum* https://www.weforum.org/stories/2023/12/three-ai-tools-setting-the-stage-for-a-tech-revolution-by-japans-entrepreneurial-doctors/ (2023).
48. Busch, F. *et al*. AI regulation in healthcare around the world: what is the status quo? Preprint at https://doi.org/10.1101/2025.01.25.25321061 (2025).
49. Sittig, D. F. & Singh, H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. *Quality and Safety in Health Care* 19, i68–i74 (2010).
50. Bergquist, T. *et al*. Piloting a model-to-data approach to enable predictive analytics in health care through patient mortality prediction. *Journal of the American Medical Informatics Association* 27, 1393–1400 (2020).
51. Kwong, J. C. C. *et al*. APPRAISE-AI Tool for Quantitative Evaluation of AI Studies for Clinical Decision Support. *JAMA Netw Open* 6, e2335377 (2023).

# CLOSING THE AI BENEFITS GAP: SYSTEMS DESIGN FOR POPULATION HEALTH EQUITY

Jessica Morley[1] – Luciano Floridi[1,2]

Abstract – Despite widespread enthusiasm, artificial intelligence (AI) has largely failed to deliver meaningful benefits in real-world healthcare settings. This commentary examines the infrastructural, ethical, technical, and regulatory barriers preventing AI from improving healthcare outcomes globally. We argue that current efforts to implement AI in healthcare are doubly flawed: they lack cohesive systems-level strategies and predominantly focus on personalised medicine rather than population health. Drawing on insights from the 2024 Global Health in the Age of AI Symposium, we propose a framework of information infrastructure requirements and call for a fundamental reorientation of AI's purpose through stakeholder-driven development of population health functional requirements. We can realise its potential to create more effective, efficient, and equitable healthcare systems worldwide only by addressing both implementation barriers and AI's current individualising focus.

*Keywords*: Artificial intelligence, global health, health equity, implementation science, population health.

## Unfulfilled promises

Artificial Intelligence (AI), heralded as the silver bullet that will finally cure the global healthcare system of its ills, including clinician burnout, structural inequities in access, and high error rates, is currently failing to live up to its potential.

In theory, AI's ability to decouple problem-solving capability from human intelligence[1] creates significant opportunities to reimagine healthcare delivery and make it more effective, efficient, and equitable. However, in reality, these promised benefits rarely ma-

---

**1.** Digital Ethics Center, Yale University, 85 Trumbull Street, CT, 06510, United States; **2.** Department of Legal Studies, University of Bologna, Italy.

terialise. Examples of AI working effectively in real-world healthcare settings outside research environments remain limited, with less than 2% of AI models reaching beyond the prototyping phase.[2] Where the move from the lab to the clinic does occur, evidence of impact on patient outcomes is limited, and solutions concentrate on narrowly focused medical domains and well-resourced regions. Consequently, the widespread adoption of AI risks widening rather than narrowing global health disparities. AI is becoming part of the problem.

## Hurdles to benefits realisation

The 'benefits gap' in AI-health has no single cause. Rather, as discussed at the 2024 Global Health in the Age of AI (GHAI) Symposium hosted by the Cini Foundation and Yale Digital Ethics Center, it is due to a range of ethical, technical, social, and regulatory factors that combine to impede AI's successful development, deployment, and use.[3]

First, AI's positive transformative potential is accompanied by significant ethical risks. AI can facilitate the dissemination of up-to-date specialist knowledge to underserved regions, enable preventative care, and optimise decision-making in real-time. However, it also poses considerable threats to patient safety, clinical authority, and healthcare's overall reliability and trustworthiness.[4] Although multiple principle-based ethical frameworks are available to help healthcare organisations manage these ethical risks proportionally, their overall utility is limited: high-level principles prove difficult to operationalise; insufficient attention is given to instances where principles conflict with each other; and contextual and cultural blind spots emerge from the narrow parameters of most frameworks. Ethical uncertainty thus acts as a dampener to adoption willingness.

Second, much of the world's healthcare systems suffer from significant data infrastructure limitations. Key infrastructural components, such as electronic health records, are often missing, outdated, or proprietary and hence inaccessible. Datasets are plagued with issues of messiness, missingness, and consequential representation biases. Complex, fragmented, and sometimes contradictory national and international information governance frameworks are typically more hindrance than help, blocking health data research and doing little to aid interoperability or privacy-preservation. Combined, these issues contribute to 'data poverty'[5] for underrepresented populations and undermine the usability and utility of AI.

Third, patient, public, and clinician trust is essential for realising AI benefits. To be trustworthy, AI must be demonstrably safe and effective, contribute to public benefit, protect healthcare's core relational values[6] (e.g., empathy and compassion), and be designed *with* not just *for* end users. Meeting these criteria is challenging. To start, there is no agreed gold standard of evidence of effectiveness for AI, nor a clear definition of public benefit; AI's aim to augment or replace aspects of human reasoning and decision-making might entirely transform the therapeutic relationship; and factors such as technical complexity and rapid development cycles often prevent effective public, patient, and professional involvement in development. These challenges pose problems because, without trust, even technically faultless AI systems face adoption resistance.

Finally, AI introduces unprecedented challenges across regulatory domains – including data protection, privacy, cybersecurity, intellectual property, liability, and medical device regulation.[7] Under-resourced and sometimes under-skilled regulators are ill-equipped to rise to the scale of this challenge. Therefore, they are struggling to keep pace with innovation. Not only does this leave patients and clinicians exposed, but it also creates legal ambiguity, which typically drives technology scepticism and abandonment rather than adoption.[8]

## Closing the benefits gap

Unless the multiple hurdles to benefits realisation are cleared in a strategic, systematic, and coordinated way, the hopes of ailing healthcare systems worldwide are likely to be dashed as enthusiasm for, and investment in, AI's potential drops and a new healthcare-specific AI winter rolls in. Fortunately, awareness of the need to fix AI's foundations before widespread adoption across the global healthcare system can be achieved seems to be growing. At both national and international scales, several ongoing initiatives are designed to tackle the outlined issues of data availability, accessibility, and quality; evidence of effectiveness; trust, and adoption willingness; and regulatory uncertainty (see Table 1).

| Benefits impediment | Example Initiatives |
|---|---|
| Data availability, accessibility, and quality | Singapore's generative AI project for automated record updating[9]<br>Saudi Arabia's Unified Health File[10]<br>Brazil's National Health Data Network[11]<br>European Health Data Space Regulation[12] |
| Evidence of effectiveness | WHO's Framework for Training, Validation and Evaluation of AI-based medical devices[13]<br>STARD-AI,[14] DECIDE-AI,[15] and CONSORT-AI[16] reporting guidelines<br>International Medical Device Regulators Forum's Good Machine Learning Practice principles[17] |
| Trust and adoption willingness | FUTURE-AI Consortium's six guiding principles[18]<br>Coalition for Health AI (CHAI) Responsible AI Guide[19]<br>UK's DART-Ed educational programme for healthcare workers[20]-ED<br>Pan-Canadian AI4H Guiding principles[21] |
| Regulatory uncertainty | FDA's Action Plan for AI/ML-Based Software[22]<br>UK MHRA's Software and AI as a Medical Device Change Programme[23]<br>Brazil's Resolution RDC No. 657/2022 for Software as a Medical Device[22]<br>WHO's Regulatory Considerations on AI for Health[24] |

Table 1. Examples of initiatives addressing impediments to AI benefits realisation in healthcare.[3]

The apparent willingness of organisations to invest in infrastructure, interoperability, education, methodological innovation, standards, medical device regulation, and more suggests that whilst the current barriers to AI benefits-realisation are formidable, they are not insurmountable. However, as the GHAI symposium revealed, the fragmented nature of these initiatives significantly undermines their effectiveness. Individual projects may directly address specific barriers, but no cohesive strategy exists to tackle the benefits gap as a whole. Consequently, several challenges persist: developers struggle to navigate an overwhelming landscape of guidance, policy and standards documents; data and interoperability initiatives rarely address end-to-end infrastructure needs; international regulations diverge in structure, enforcement, and specifications, risking regulatory arbitrage; and a lack of attention to sustainability in all initiatives introduces the risk that the ecological impact of AI might undermine its positive impact on health. Overall, current impediment-mitigating strategies overlook a fundamental truth: AI is a systems-level technology[25] that can transform every aspect of healthcare. Adoption and implementation strategies cannot treat it like an upgrade to an existing product that can simply be plugged into the global healthcare system. Closure of the benefits gap instead requires systems design.[26]

## Requirements for success

Treating AI like another product to be plugged into the global healthcare system encourages policymakers to fall back on 'command and control'[27] approaches to planned technology implementation. The inherently reductionist nature of these approaches almost guarantees failure-inducing policy fragmentation because they assume benefit-impediments exist within bounded rationality solvable through linear solutions[28] – for example, assuming that purchasing an EHR will automatically result in the creation of clean, structured datasets. This assumption overlooks the complex interrelationships within healthcare systems. Systems design, a poietic (i.e. "making") science that helps policymakers understand the present, develop ideal system models, and use policy to realise these models,[26] offers an alternative: actively seeking to make complexity manageable rather than denying its existence.[29] The process begins by identifying the system's components, connections, constraints, and stakeholders' needs, asking open questions about 'what can be?' rather than prescribing specific solutions. Then follows an iterative inferential process where these insights are translated conceptually into comprehensive requirements for implementation success.[26] Thus, for AI in healthcare, effective systems design means identifying the information infrastructure requirements (the interacting technical, informational, procedural, sociocultural, ethical, and regulatory components) that enable AI systems to perform as intended in the global healthcare system from ideation to post-implementation.

Following this systems design approach after the GHAI symposium[3] revealed five core requirements (see Table 2) for closing the AI benefits gap in healthcare: (1) infrastructure components related to hardware, software, and clinical content designed around robust data exchange; (2) human-computer interaction and workflow components that foster epistemic certainty and autonomy; (3) internal policies and organisational culture built on actively protected healthcare values; (4) external policies, regulations, and management structures enforcing validated outcomes and meaningful accountability; and (5) environmental sustainability as a core consideration throughout. Although ensuring these interconnected requirements are met in ways that are simultaneously contextually specific and internationally interoperable will not be easy, it will be worth the attempt. Together, they create a framework for managing AI's complexity, establishing the necessary foundation for realising AI's potential benefits while minimising its risks.

| **System components** | **Core requirement** | **Benefit-impediments addressed** | **Example policy interventions** |
|---|---|---|---|
| Infrastructure components related to hardware, software, and clinical content | **Robust data exchange** to facilitate the trustworthy flow of data between patients (source), clinicians, developers, regulators, and system managers for the entire AI pipeline. | • Data infrastructure limitations<br>• Missing, outdated, or proprietary EHRs<br>• Data poverty for underrepresented populations<br>• Privacy infringing data sharing<br>• Lack of interoperability | • Develop specialty-specific UX-friendly open EHR templates<br>• Value data work and invest in AI-ready datasets<br>• Fund the development of model-to-data access mechanisms and their evaluation<br>• Mandate compliance with interoperability standards<br>• Establish protocols to enable international collaboration to protect data sovereignty<br>• Prevent vendor lock-in<br>• Invest in transfer learning techniques |

| | | | |
|---|---|---|---|
| Human-computer interaction and workflow components | **Epistemic certainty and autonomous staff and patients** to help ensure clinicians and patients can interpret AI outputs, and feel confident that outputs are valid, valuable, and constructed to improve their capacity to make shared decisions that will ultimately lead to better outcomes | • Threats to clinician authority<br>• Transformation of the therapeutic relationship<br>• Adoption reticence due to distrust | • Incorporate data literacy into medical education<br>• Protect clinicians' right to override AI advice<br>• Design AI outputs to be actionable<br>• Ensure AI limitations are discoverable<br>• Develop AI in response to identified clinical needs, not just data availability |
| Internal policies and organisational culture | **Actively protected values** to help ensure equity, patient-centricity, and trustworthiness are built into the entire AI lifecycle | • Widening of global health disparities<br>• Difficulty operationalising ethical principles<br>• Transformation of the therapeutic relationship<br>• Insufficient attention to context and culture | • Mandate diverse AI development teams<br>• Require patient, public, and professional involvement in all stages of the AI lifecycle<br>• Incorporate qualitative data in AI systems as well as quantitative<br>• Integrate AI with broader health system strengthening<br>• Prioritise conditions with high unmet needs |
| External policies, regulations, and management structures | **Validated outcomes and meaningful accountability** to ensure AI is supported by appropriately high-quality evidence, there are clear liability guidelines, proportionate governance of the entire pipeline, and regulatory interoperability | • Lack of gold standard evidence for AI<br>• Regulatory challenges across domains<br>• Legal ambiguity driving technology scepticism<br>• Regulatory arbitrage | • Standardise AI as medical device definitions<br>• Adapt Health Technology Assessment Protocols for AI<br>• Mandate documentation of development decisions<br>• Embed sociotechnical evaluation in regulations |

| | | | |
|---|---|---|---|
| | | | • Design human plus algorithm evaluations<br>• Simplify information governance frameworks |
| Environmental sustainability | **Ecological responsibility** to ensure climate considerations are built into AI development, and there is an appropriate balance between technological advancement and environmental protections so that the ecological impact of AI does not undermine any health gains derived from its implementation | • Concerns regarding climate change and the impact that this will have on healthcare<br>• Duplicative and wasteful development efforts | • Mandate AI be accompanied by environmental impact assessments<br>• Establish energy-efficient standards for AI<br>• Require lifecycle planning for AI hardware<br>• Develop frameworks for prioritising high-impact use cases |

Table 2. Information infrastructure requirements necessary for supporting the benefit-realising implementation of AI into the global healthcare system.[3]

## Outcomes follow function, not form

The information infrastructure requirements provide only a necessary foundation for benefits-realisation, not a sufficient one. Systems must fulfil specific functions to be effective, and when inappropriate functions are assigned, even well-designed infrastructure cannot guarantee desired outcomes. Currently, the purpose of AI in healthcare is primarily defined by policymakers and technology developers rather than the global healthcare community itself, emphasising personalised and predictive medicine that targets individuals rather than populations. This purpose rests on several unfounded assumptions and has led to over-investment in the development of largely ineffective, and sometimes directly harmful, risk stratification algorithms that (a) shift responsibility for gaining and maintaining health from systems to patients under the guise of 'empowerment';[30] (b) exacerbate global health inequities through the "global inverse data quality law" – where data quality and algorithmic performance are inversely correlated with population health needs; and (c) contribute to a "politics of avoidance" – the tendency for policymakers to favour technological interventions as workarounds for addressing more contentious social determinants of health.[31] Thus, while AI implementation continues to prioritise individual-level interventions over population health approaches, its potential benefits will remain largely unrealised.

To truly close the AI benefits gap, a fundamental reorientation is needed. What must come next is the collective development of population-health functional requirements for AI systems, defined through inclusive dialogue among diverse stakeholders from across the global healthcare community and focused on how AI can meaningfully contribute to population-level flourishing. Such an approach requires clarifying what constitutes population-level flourishing in measurable terms, identifying the institutional and technical barriers that maintain AI's individualised orientation, and developing concrete strategies to redirect AI development toward applications that address social determinants of health. Only through this collective redefinition of AI's purpose can we establish functional requirements that ensure these technologies fulfil their promise of making healthcare more effective, efficient, and equitable for all populations, not just privileged individuals in resource-rich settings.

## Conclusion

As enthusiasm and investment in AI's healthcare potential grows, so does the risk of disillusionment and a healthcare-specific AI winter if benefits remain unrealised. Addressing this requires tackling both implementation challenges and AI's fundamental misalignment of purpose. Only by designing both the infrastructural underpinnings and the strategic direction can AI transform from a technology that exacerbates inequities into a catalyst for systemic change that promotes flourishing for all populations.

## Acknowledgements

This research is the outcome of the three-day symposium, entitled "Global Health int eh Age of AI", organised by the Fondazione Giorgio Cini, in collaboration with Yale's Digital Ethics Center, on 7-9 November 2024, on the Island of San Giorgio Maggiore, Venice. The authors are very grateful for the support they received form the Foundation. They declare no conflicts of interest.

## References

1. Floridi, L. AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis. *Philos. Technol.* 38, 30, s13347-025-00858–9 (2025).
2. Van De Sande, D. *et al.* To warrant clinical adoption AI models require a multi-faceted implementation evaluation. *npj Digit. Med.* 7, 58 (2024).
3. Morley, J. *et al.* Global Health in the Age of AI: Charting a Course for Ethical Implementation and Societal Benefit. SSRN Scholarly Paper at https://papers.ssrn.com/abstract=5217060 (2025).
4. Morley, J. *et al.* The ethics of AI in health care: A mapping review. *Social Science & Medicine* 260, 113172 (2020).
5. Ibrahim, H., Liu, X., Zariffa, N., Morris, A. D. & Denniston, A. K. Health data poverty: an assailable barrier to equitable digital health care. *The Lancet Digital Health* 3, e260–e265 (2021).
6. Kerasidou, A. Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. *Bull. World Health Organ.* 98, 245–250 (2020).
7. Gerke, S., Minssen, T. & Cohen, G. Ethical and legal challenges of artificial intelligence-driven healthcare. in *Artificial Intelligence in Healthcare* 295–336 (Elsevier, 2020). doi:10.1016/B978-0-12-818438-7.00012-5.
8. Greenhalgh, T. *et al.* The NASSS-CAT Tools for Understanding, Guiding, Monitoring, and Researching Technology Implementation Projects in Health and Social Care: Protocol for an Evaluation Study in Real-World Settings. *JMIR Res Protoc* 9, e16861 (2020).
9. Ang, A. Singapore invests $150M for public health genAI adoption. https://www.healthcareitnews.com/news/asia/singapore-invests-150m-public-health-genai-adoption (2024).
10. World, H. HEALTHCARE AND AI IN THE KINGDOM OF SAUDI ARABIA. *Healthcare World* https://healthcareworld.com/healthcare-and-ai-in-the-kingdom-of-saudi-arabia/ (2024).
11. Barbalho, I. M. P. *et al.* Electronic health records in Brazil: Prospects and technological challenges. *Front. Public Health* 10, 963841 (2022).

12. European Commission. European Health Data Space Regulation (EHDS) – European Commission. https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en (2025).

13. WHO. *Generating Evidence for Artificial Intelligence-Based Medical Devices: A Framework for Training, Validation and Evaluation*. (World Health Organization, Geneva, 2021).

14. Sounderajah, V. *et al*. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. *BMJ Open* 11, e047709 (2021).

15. Vasey, B. *et al*. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ* 377, e070904 (2022).

16. Liu, X. *et al*. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *The Lancet Digital Health* 2, e537–e548 (2020).

17. IMDRF. Good machine learning practice for medical device development: Guiding principles | International Medical Device Regulators Forum. https://www.imdrf.org/documents/good-machine-learning-practice-medical-device-development-guiding-principles (2025).

18. Lekadir, K. *et al*. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ* e081554 (2025) doi:10.1136/bmj-2024-081554.

19. CHAI. Responsible AI Guide. *CHAI – Coalition for Health AI* https://chai.org/responsible-ai-guide/ (2025).

20. About the DART-Ed programme | Workforce, training and education | NHS England. *NHS England | Workforce, training and education | Digital Transformation* https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/about-the-dart-ed-programme.

21. Canada, H. Pan-Canadian AI for Health (AI4H) Guiding Principles. https://www.canada.ca/en/health-canada/corporate/transparency/health-agreements/pan-canadian-ai-guiding-principles.html (2025).

22. Busch, F. *et al*. AI regulation in healthcare around the world: what is the status quo? Preprint at https://doi.org/10.1101/2025.01.25.25321061 (2025).

23. MHRA. Software and AI as a Medical Device Change Programme – Roadmap. *GOV.UK* https://www.gov.uk/government/publications/software-and-ai-as-a-medical-device-change-programme/software-and-ai-as-a-medical-device-change-programme-roadmap (2023).

24. WHO. *Regulatory Considerations on Artificial Intelligence for Health*. https://iris.who.int/bitstream/handle/10665/373421/9789240078871-eng.pdf. (2023).

25. McCradden, M. D. *et al*. CANAIRI: the Collaboration for Translational Artificial Intelligence Trials in Healthcare. *Nat Med* 31, 9–11 (2025).

26. Floridi, L. The Logic of Design as a Conceptual Logic of Information. *Minds & Machines* 27, 495–519 (2017).

27. Chapman, J. *System Failure: Why Governments Must Learn to Think Differently*. (Demos, London, 2004).

28. Atun, R. Health systems, systems thinking and innovation. *Health Policy and Planning* 27, iv4–iv8 (2012).
29. Meadows, D. H. & Wright, D. *Thinking in Systems: A Primer*. (Chelsea Green Pub, White River Junction, Vt, 2008).
30. Morley, J. & Floridi, L. The Limits of Empowerment: How to Reframe the Role of mHealth Tools in the Healthcare Ecosystem. *Sci Eng Ethics* 26, 1159–1183 (2020).
31. Shipton, L. & Vitale, L. Artificial intelligence and the politics of avoidance in global health. *Social Science and Medicine* 359, (2024).

PRINTED FOR LEO S. OLSCHKI EDITORE
BY ABC TIPOGRAFIA • CALENZANO (FI)
IN MARCH 2026

fondazione GIORGIO CINI ONLUS

# QUADERNI DI SAN GIORGIO

Between 1955 and 1984, the Foundation published thirty-eight volumes of the 'Quaderni di San Giorgio' which brought together contributions in the form of essays or articles from the most authoritative names on the international cultural scene, those who would come to the island of San Giorgio to take part in conferences and study days, in other words, to implement what has always been defined here as 'the dialogue between cultures'.
Even today, with their elegant and sober design, graceful typeface, ivory paper and wide margins for notes, the Quaderni represent an essential source of inspiration and knowledge for all those who wish to explore their contents, sometimes attracted by truly unusual and incredibly topical titles. The themes covered included social, cultural, religious, economic and anthropological issues, revolving around interdisciplinary education with both scientific and humanistic approaches, ranging from mathematics to the environment and from art to the Gospel. They found their place and nourishment in a fruitful dialectical counterpoint rarely oriented towards pure historical investigation, but expressly aimed at seeking keys to understanding the contemporary world of the day.
The series was published by Fondazione Giorgio Cini Foundation with Sansoni Editore.

1. *Il valore del fine del mondo*, scritti di G. Bozzetti, F. Carnelutti, L. Fantappiè, P. Rondoni, 1955
2. *Arte figurativa e arte astratta*, introduzione di Francesco Carnelutti, 1955
3. *I dialoghi del Vangelo e la Civiltà*, conversazioni di Francesco Carnelutti, 1956
4. *Interpretazione di Capograssi*, conversazioni di Francesco Carnelutti, 1956
5. *Processo dell'Islam alla civiltà occidentale*, di Guido Piovene, 1957
6. *L'avventura dell'individuo*, di Francesco Carnelutti, 1957
7. *Casualità e finalità*, a cura di Marino Gentile, 1959
8. *Arte e scienza*, a cura di Augusto Guzzo, 1959
9. *Figure del Vangelo*, colloqui di Francesco Carnelutti, 1958
10. *Il granello di senapa (Parabole del Vangelo)*, colloqui di Francesco Carnelutti, 1959
11. *Adolescenza traviata*, a cura di Mario Pomilio, 1961
12. *Il discorso della libertà*, colloqui di Francesco Carnelutti, 1960
13. *Cinema e civiltà*, a cura di Paolo Gadda Conti, 1960
14. *Nuove vie per la scienza*, di Luigi Fantappié, 1961
15. *L'esperienza della preghiera*, di Paolo Brezzi, 1961
16. *Il poema di Gesù*, di Francesco Carnelutti, 1961
17. *Cinema e giustizia*, a cura di Paolo Gadda Conti, 1961
18. *Cinema e sesso*, a cura di Paolo Gadda Conti, 1962
19. *Primi problemi della riforma del processo penale*, a cura di Giuseppe De Luca, 1962
20. *Maria e le altre donne*, colloqui di Francesco Carnelutti, 1962
21. *Valore della musica*, a cura di Marino Gentile, 1965
22. *Cinema e libertà*, a cura di Paolo Gadda Conti, 1963
23. *Arte e cultura contemporanee*, a cura di Piero Nardi, 1964
24. *Cinema e società*, a cura di Paolo Gadda Conti, 1965

25. *Spazio e cielo*, a cura di Vittorio Mathieu, 1966
26. *La Magistratura nello stato democratico*, a cura dell'Unione Giuristi Cattolici Italiani, 1965
27. *Teilhard de Chardin e il pensiero cattolico*, a cura di C. Cuénot, 1966
28. *Lo spirito dell'Umanesimo africano*, a cura di Stefano Rosso-Mazzinghi, 1967
29. *Arte e cultura nella civiltà contemporanea*, a cura di Piero Nardi, 1966
30. *Settimana di studi giuridici in onore di Francesco Carnelutti*, a cura di Nicola Mangini, 1967
31. *Il fenomeno «città» nella vita e nella cultura d'oggi*, a cura di Piero Nardi, 1971 [31-32]
32. *Il fenomeno «città» nella vita e nella cultura d'oggi*, a cura di Piero Nardi, 1971 [31-32]
33. *La critica, forma caratteristica della civiltà moderna*, vol. I, a cura di Vittore Branca, 1970
34. *L'uomo e il suo ambiente*, a cura di Stefano Rosso-Mazzinghi, 1973
35. *Da Antonio Canova alla Convenzione dell'Aja. Evoluzione della protezione delle opere d'arte in caso di conflitto armato*, a cura di Stefano Rosso-Mazzinghi, 1975

Nuova serie

36. *Tolstoj oggi*, a cura di Sante Graciotti e Vittorio Strada, 1980
37. *Dostoevskij nella coscienza d'oggi*, a cura di Sante Graciotti, 1981
38. *I linguaggi del sogno*, a cura di Vittore Branca, Carlo Ossola e Salomon Resnik, 1984

OLSCHKI EDITORE

39. *Global Health in the Age of AI*, curated by Luciano Floridi. 2026, vi-88 pp.

Luciano Floridi is the founding director of the Digital Ethics Center and a professor in the Cognitive Science program at Yale University. He was a professor at Oxford University and is a full professor of Sociology of Culture and Communication at the University of Bologna. He has published more than 300 works on the philosophy of information, digital ethics, the ethics of artificial intelligence, and the philosophy of technology. He is the author of numerous publications, including: *The 4th Revolution* (2017, winner of the Walter J. Ong Award), *Thinking about the infosphere* (2020), *The green and the blue* (2020) and *The Ethics of Artificial Intelligence* (2022). In 2022, the President of the Italian Republic bestowed him with the title of *Cavaliere di Gran Croce dell'Ordine al Merito*.

The analyses that emerged during the Symposium, which was held at the Fondazione Giorgio Cini in November 2024, show that without a coordinated, multi-level effort, AI risks exacerbating rather than reducing global health inequalities. This has prompted the idea of formulating recommendations that propose a systemic approach to coordinate technical-scientific, social, ethical, and regulatory aspects. These recommendations are illustrated in this 39th *Quaderno di San Giorgio*, and they represent both the outcomes of the symposium as well as a starting point for new studies and solutions, providing a collection of coordinates that may serve to sketch potential future scenarios and actions that could be undertaken today. One fundamental question remains open: no "tech solutionism" can compensate for the social and structural inequalities that affect access to healthcare and health outcomes globally; constant and sustained commitment is needed from all stakeholders.

Versione cartacea - Printed version
ISBN 978 88 222 9539 2

Versione digitale - Digital PDF
ISBN 978 88 222 8693 2

DOI: 10.82026/9788822286932