

<https://doi.org/10.1038/s41746-025-01695-y>

# A framework for considering the use of generative AI for health

Isabella Joy de Vere Hunt<sup>1,2</sup>, Kang-Xing Jin<sup>1,3</sup> & Eleni Linos<sup>1</sup> ✉

This perspective provides a critical framework for evaluating the use of Generative AI (GenAI) in healthcare, informed by two roundtable events and a series of semi-structured expert interviews. By outlining key principles and risks—including mapping applications to GenAI strengths, ensuring rigorous evaluation frameworks, and fostering transparency and collaboration—it will support healthcare professionals and policymakers in navigating the ethical and impactful integration of GenAI into healthcare.

*Generative artificial intelligence (GenAI)* describes computational techniques capable of generating seemingly new, meaningful content such as text, images, or audio from training data<sup>1</sup>. There has been a rapid expansion in GenAI use heralded by the launch in November 2022 of *Chat Generative Pre-trained Transformer (GPT)*, a non-domain-specific large language model (LLM) trained on a large corpus of text data. LLMs are a type of GenAI system trained on large amounts of text data, that understands and generates human-like language<sup>2</sup>. Beyond LLMs (such as GPT-4 or Claude), other examples of GenAI systems include image generators, which utilize another form of GenAI model termed generative adversarial networks (GANs) (e.g. Midjourney); code generation tools (e.g. Copilot); and audio generation tools (e.g. resemble.ai)<sup>3</sup>. GANs are being leveraged in medical diagnosis (often in conjunction with LLMs), for example, to identify patterns related to disease in medical imaging<sup>3,4</sup>. However, this article principally focuses on LLMs, given their predominance in emerging medical use cases in the wake of the launch of ChatGPT (see Table 1 for healthcare-specific examples).

Despite the rise in GenAI use, with an array of emerging use cases in healthcare being considered and implemented, there is no universal agreement on how GenAI should be used in healthcare<sup>5</sup>. We summarize key principles for all potential users and evaluators of GenAI tools in healthcare, incorporating learnings from two roundtable events and in-depth qualitative interviews with experts in Generative AI and digital health, encompassing perspectives from academia, funding bodies, implementers, and health system leaders. These events and interviews were conducted as part of the development of a White Paper (see Supplementary Data 1) funded by the Advancing Health Online Initiative, seeking to address on a rapid timescale the lack of guidance for funders regarding where to invest in GenAI to maximize health impact in low- and middle-income countries<sup>6</sup>. The intended audience of this White Paper is individuals responsible for investment decisions and the implementation of GenAI in health settings. However, academics, clinicians, and public health practitioners must also

consider how GenAI will impact our field, and how we can best leverage GenAI for positive health impact, whilst mitigating potential harm. This is of pressing relevance given the wide interest in adopting GenAI. Implementing such technologies in health contexts requires specific considerations, with the risk of potential harms—but also potential benefits—amplified, compared to some other sectors. We have therefore distilled our key findings into four overarching principles and four key categories of risk to guide the use of GenAI for health, with a focus on ensuring equitable global implementation.

## Framework development

We defined our key stakeholder groups as follows: *Health Implementer*: part of organization responsible for creating or implementing GenAI health tools; *Health Funder*: part of organization responsible for funding the implementation of GenAI health tools; *Academic*: part of organization that conducts research on digital or global health; *Health System Expert*: part of organization that delivers healthcare services; and *Tech Facilitator*: part of technology platforms that support technical implementation of GenAI health tools. Although we were able to incorporate secondary analysis of patient and public feedback collected by health implementers as part of five in-depth case studies, we did not engage directly with patient end-users for the writing of this piece, meaning that their perspectives on this framework are not reflected. Authors EL and IDVH are both practicing clinicians, able to provide insight from the perspective of healthcare provider end-users.

The first roundtable event was a workshop held at Stanford University on October 30, 2024, attended by 12 academics, 3 health implementers, 1 health funder, and 4 health system experts. Attendees were selectively invited to achieve a diverse range of expert perspectives, and international attendance was facilitated by coordination with the Stanford Center for Digital Health Annual Symposium. The purpose of the workshop was to facilitate open discussion centered on the following questions: (1) *How is GenAI being used to improve health in LMICs (in terms of use cases and*

<sup>1</sup>Stanford Center for Digital Health, Stanford University, Stanford, CA, USA. <sup>2</sup>Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK. <sup>3</sup>Advancing Health Online Initiative, Rahway, NJ, USA. ✉e-mail: [linos@stanford.edu](mailto:linos@stanford.edu)

Table 1 | Key LLM task domains with healthcare examples

LLM Task	Definition	Healthcare-specific example
Summarization	Condensing content into shorter summaries	Summarizing long medical guidelines into succinct summaries for immediate ‘in-clinic’ use.
Classification	Assigning labels or categories to content	Categorizing online patient messages into categories such as medical versus administrative queries, to facilitate more efficient handling of queries.
Extraction	Identifying and retrieving information from a larger body of content	Identifying and extracting salient data points such as diagnoses, medications, and test results from patients’ medical records.
Translation	Converting content from one form to another, across languages, formats, or styles	Rewriting content to match different tones or styles, e.g. transforming clinical documents into patient-facing material.
Conversation	Engaging in dynamic, context-aware exchanges.	A health chatbot providing real-time, personalized responses to user questions.

health areas addressed)? (2) How should we measure health impact and cost-effectiveness? (3) What are the key risks, and how can we mitigate them? The workshop was audio-recorded and transcribed verbatim. Author IDVH undertook an inductive thematic analysis of the full workshop transcript in the qualitative analysis software NVIVO 12, using a mind-mapping and constant comparison approach<sup>7,8</sup>. All codes and preliminary themes were reviewed by a second researcher for accuracy and iteratively refined.

Preliminary themes from the Stanford Workshop were utilized to inform a semi-structured interview guide for a series of 1:1 interviews, each lasting approximately 45 minutes and conducted over Zoom by IDVH from November–December 2024. The interview guide explored the current use, health impact, cost-effectiveness, implementation challenges, and ethical and governance principles regarding the use of GenAI for health in LMICs (see Supplementary Note 2). Purposeful sampling ensured a breadth of key stakeholder perspectives, with a total of 11 health implementers, 3 health funders, 2 tech facilitators, and 7 health system experts represented in the initial round of interviews. Interviews were audio-recorded and transcribed verbatim, and IDVH undertook a rapid inductive thematic analysis of the full interview transcripts to incorporate the additional interview findings into the preliminary themes identified at the Stanford workshop, and all additional codes and themes were reviewed by a second researcher. These preliminary themes were presented at a roundtable event hosted in Nairobi on December 3, 2024, prior to the Global Digital Health Forum (GDHF), attended by 5 academics, 7 health implementers, 1 health funder, and 5 health system experts. We gathered focused feedback on the initial themes, achieving group consensus, and refining concepts regarding best practices for evaluating health impact and potential harms. Following this roundtable, a further 3 interviewees (1 health implementer, 1 health funder, and 1 health system expert) were identified by purposeful sampling at GDHF to maximize expert input and ensure saturation of key themes. To support validation of key concepts, some 1:1 interviewees also attended either the Stanford workshop or Nairobi roundtable, with a total of 54 individuals represented across the two roundtable events and semi-structured interviews (14 academics, 17 health implementers, 7 health funders, 2 tech facilitators, and 13 health system experts). Final consensus on key principles and risks was reached at a meeting with researchers from the Stanford Center for Digital Health and the Center for Advanced Study in the Behavioral Sciences.

This commentary complements existing literature on digital health governance and AI ethics by providing concrete principles and risk mitigation strategies, emphasizing equitable implementation in LMICs. The recent FUTURE-AI framework provides valuable and comprehensive guidance for the development and deployment of trustworthy AI tools in healthcare<sup>9</sup>. However, Lekadir et al. state that ‘the deployment and adoption of AI technologies remain limited in real-world clinical practice’, which our extensive scoping work has identified is no longer the case (see example case studies in Supplementary Data 1). Further, their expert consortium represents an entirely academic perspective. This paper adds novel insights by integrating current perspectives from implementing organizations, on-the-ground health-system experts, and funding body representatives, thereby grounding its recommendations in timely real-world considerations.

Key Principles

Map potential applications to GenAI strengths

The most successful LLM implementations are for domain-specific use cases that map well onto LLM strengths and weaknesses. Partnership with community organizations is crucial to ensuring the needs and concerns of target end-users are heard and integrated throughout the design and implementation process. In terms of strengths, LLMs have been shown to perform significantly better than previous AI approaches in specific task areas. Technologies are advancing rapidly, but some currently validated task domains with healthcare-specific examples are outlined in Table 1. We expand on key weaknesses and mitigation approaches below.

Define and implement evaluation frameworks

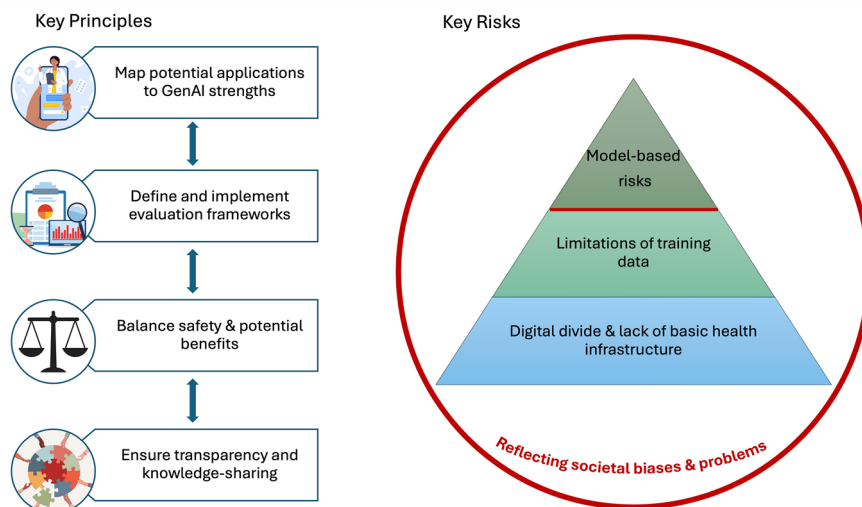
Although existing health-specific outcome measures are broadly applicable across digital health interventions, with morbidity and mortality remaining key, we lack established standards for measurement and benchmarking specific to GenAI tools. The intended health impact of an intervention and corresponding outcome metrics should be clearly defined from the design phase. The Medical Research Council and the National Institute for Health Research’s 2021 Framework for Developing & Evaluating Complex Interventions recommends evaluators work with stakeholders to assess which outcomes are most important<sup>10</sup>. For tools using LLMs to generate responses to queries (whether from a healthcare worker or user), metrics of interest are likely to include correctness and completeness of LLM responses and time- and cost-savings compared to previous non-GenAI methods, as well as qualitative factors such as understandability, empathy, and appropriateness of tone and style. In addition to asking whether an intervention achieves its intended outcome, evaluators should consider broader contextual questions surrounding impact, including interactions with the existing healthcare context<sup>10</sup>. Gold-standard evaluation methods such as randomized controlled trials can take years to provide actionable results, meaning we also need better ways to measure success to inform time-critical implementation and funding decisions in interim periods. Unlike a pharmaceutical compound, GenAI outputs are constantly evolving, so evaluation and review processes must be continuous to ensure tools remain accurate.

Balance safety & potential benefits

Applying the central ethical principle of ‘do no harm’ when considering possible GenAI interventions is intuitive, and underpins the WHO guidance on *Ethics & Governance of Artificial Intelligence for Health*<sup>11</sup>, and ‘Responsible AI’ policies more broadly. Although we echo and emphasize the paramount importance of optimizing safety, it is also important to consider the opportunity costs of not considering and testing GenAI tools, given their potential to improve access to care, and the current need for innovative solutions to meet critically under-addressed healthcare needs: the projected shortage of health workers in Africa is expected to be 6.1 million by 2030<sup>12</sup>. Despite the pressing need to ensure safe implementation of potential GenAI tools, there is not a clear regulatory landscape internationally. The EU Artificial Intelligence Act is the first comprehensive regulation on AI by a major regulator globally, coming into force on August 1, 2024<sup>13</sup>. AI governance frameworks are also rapidly emerging in LMICs: a growing number of countries in Africa, South Asia, and South America have

**Fig. 1 | A framework summarizing the key principles and risks of using generative AI for health.**

We present four key principles to guide the use of GenAI for health: map potential applications to GenAI strengths; define and implement evaluation frameworks; balance safety and potential benefits; and ensure transparency and knowledge-sharing. We also identify four key categories of risk, conceptualized as a pyramid model: model-based risks; limitations of training data; digital divide and lack of basic health infrastructure; and reflecting societal biases and problems.



released national AI strategies, with others in development, but these often focus on strategic development and ethical guidelines rather than enforceable legal requirements<sup>14–17</sup>. Indeed, regulators in many LMICs lack the resources to govern digital health tools effectively. Ideally, approaches to apply GenAI to health, both voluntary and regulatory, will foster experimentation and implementation in lower-risk applications to maximize benefits, while putting in appropriate controls in higher-risk settings.

### Ensure transparency & knowledge-sharing

The importance of collaboration, transparency, and knowledge-sharing cannot be underestimated; failures, while inevitable, are also valuable learning opportunities. Funders can support in creating the right incentives for organizations to collaborate and facilitate the sharing of benchmarks for the evaluation of emerging GenAI tools. We emphasize the importance of transparent reporting, including details of how AI systems were designed and trained, alongside disclosures of conflicts of interest and publication of negative findings in a field where commercial and public priorities are often tightly intertwined. The CONSORT-AI extension provides a useful tool for reporting clinical trials for AI interventions<sup>18</sup>. Many organizations are currently piloting applications of GenAI—establishing forums where they can share interim results and real-time learnings will also be valuable. The successful implementation of interventions able to effect system-level change will also require extensive government buy-in and cross-sector collaboration.

### Key risks

Even when embracing the above key principles, risks associated with the use of GenAI tools for health must be considered throughout the design, implementation, and evaluation phases. We conceptualize these risks as a pyramid model, with intrinsic model-based risks layered upon largely extrinsic risks stemming from limitations of training data, and the digital and local infrastructure of healthcare systems in which they are integrated (Fig. 1). These risks in turn, are embedded within existing societal biases and problems which do not stem from the use of GenAI technologies per se, but which GenAI has the potential to amplify.

#### Model-based risks

These are risks we consider largely intrinsic to GenAI technologies themselves, the consequences of which must be carefully considered and mitigated in any healthcare setting.

**Inaccuracies.** First, the risk of inaccuracy, or ‘hallucinations’ (defined as mistakes in the generated content that seem plausible but are incorrect)<sup>19</sup>, is inextricably linked with a tool that is generating content. Just as we cannot eradicate human error and must ensure systems are in place in

healthcare settings to minimize this, the same is required for GenAI tools. The risk of errors and appropriate mitigation approaches must be evaluated in a context-dependent manner.

#### Mitigation strategies.

- Consider acceptable error rates for specific use cases, including the impact of errors and existing human error rates. For example, for a tool using GenAI to categorize incoming user queries by intent, false negatives (i.e., health queries misclassified as non-health) should be minimized at the cost of a higher false positive rate.
- Prompt engineering or tactics like Retrieval Augmented Generation<sup>20</sup> (constraining a model’s outputs to specific documents) can allow models to provide more domain-specific responses and reduce errors.
- Rigorous human oversight of any implemented tools: ‘human-in-the-loop’ is a practice increasingly emphasized in the field of AI and refers to the use of human intervention to control or change the outcome of a process<sup>21</sup>. This is a key feature of many current use cases where errors would detriment patient care (see case studies in Supplementary Data 1).
- Establish regular monitoring of error rates, mindful of the need for continual review processes given the ever-changing nature of GenAI algorithms.

**Cost and environmental impact.** Although costs are falling, LLMs are typically more expensive than other AI methods. Costs also differ for different languages, as LLMs operate on tokens: although in English most words are a token, in many African languages such as Swahili, several tokens are required for a single word. Languages with a higher token density are more expensive (and slower) for an LLM to process, with cost, latency, and equity implications. There are also environmental concerns regarding the increasing use of GenAI, as the development and operation of the technology consumes large amounts of electricity<sup>22,23</sup>.

#### Mitigation strategies.

- Only use GenAI tools when impact and cost-effectiveness modelling indicate they are likely to provide advantages over existing methods, guided by a responsible AI Return on Investment framework<sup>24</sup>.
- Direct resources where they are most needed: this might entail focusing on simple, lower-cost initiatives that have a high impact.
- Shared infrastructure for models trained with local languages can help reduce costs associated with the use of non-Western languages: UlizaLlama is the first free-to-use Swahili LLM, and is currently being expanded for use in other African languages<sup>25</sup>.

- Strategies to reduce the energy footprint of AI algorithms include more efficient training algorithms, reducing the size of neural network architectures, and optimizing hardware<sup>22</sup>.

**Data security and privacy.** Data security and privacy remain paramount as with all healthcare digitization, and GenAI providers have different data usage approaches and policies. Although data security and privacy issues transcend the ‘intrinsic’ model-based risks and also feed into concerns from a wider social context regarding potential misuse of personal data, there is a specific model-based concern regarding the potential of GenAI chatbots to elicit large amounts of personal information more readily than in current clinical settings. This raises concerns about storage practices and the risk of misuse, particularly in contexts where a given health-related behavior (such as same-sex sexual activity) is stigmatized or criminalized. GenAI models also have the potential to violate copyright laws depending on the data they are trained on<sup>26</sup>. Privacy-enhancing technologies (PETs)—including use of synthetic data, homomorphic encryption, differential privacy, federated learning and federated analysis, trusted execution environments and secure multi-party computation—can be implemented to protect privacy and cybersecurity, but introduce potential performance trade-offs for models whose accuracy relies on personal data, requiring nuanced implementation<sup>27</sup>. Although several regulators and government agencies cite the utility of PETs, there remains a lack of clear and consistent guidance: this will be important to encourage effective and appropriate implementation of PET approaches.

#### Mitigation strategies.

- Review the data safeguards and policies of any potential LLM provider, with consideration of the type of data that will be analyzed (e.g. is personal health information collected, or is data anonymized or aggregated at the health system level?) and the suitability of the model employed (e.g., open models for public data, private deployments for sensitive data).
- Ensure you have the rights to use training data<sup>23</sup>.
- Integrate automated scanning to detect and flag sensitive data before it reaches models.
- Consider trade-offs between data utility and data protection, and employ specific PETs where appropriate<sup>27</sup>.

#### Limitations of the training data

Even if a given GenAI model is optimized to mitigate the intrinsic risks detailed above, it is crucial to acknowledge that GenAI is only as good as the data it is trained on: currently, this is largely European languages with Western cultural contexts, with the corresponding limitations. Solving this problem requires sufficient training data in other languages, which may be challenging to acquire; we welcome the February 2025 announcement by the AI for Development Funders Collaborative of \$10 million towards the development of AI models that are inclusive of African languages<sup>28</sup>. The available data for a given language may also have inadequate coverage for particular health areas—for example, in places which lack gender parity, women’s health concerns may be overlooked or trivialized in existing language datasets.

#### Mitigation strategies.

- Invest in the development of diverse language data sets: for example, Lelapa AI is an organization developing small language models for low-resource African languages<sup>29</sup>.
- Voice models are being used to serve populations with low literacy; improving these models requires investment in voice-specific data.
- Invest in the development of benchmarks for the quality and accuracy of models for specific languages of interest.
- Test and train LLM-enabled tools on data relevant to the intended healthcare setting: for example, the AfriMed-QA dataset is an multi-institution, open-source dataset of 25,000 Africa-focused medical question-answer pairs created to represent the disease burden and medical practices common across Africa<sup>30</sup>.

#### Digital divide & lack of basic health infrastructure

Even fully optimized models and data sets remain redundant if their intended end users lack access to either the physical or digital infrastructure to enable their effective use. The ‘digital’ element is pressing, with the World Bank estimating in 2022 that only 36% of people living in Africa had access to broadband internet<sup>31</sup>. Many low-income settings lack the computing infrastructure needed to handle the complexity of training GenAI algorithms, resulting in reliance on external infrastructure—but this limits the control and ownership of GenAI initiatives in their local settings<sup>24</sup>. We must also consider potential lack of basic health infrastructure: for example, do the local clinics stock the HPV vaccine your chatbot has successfully motivated young women to seek? Even where digital and healthcare system infrastructure is present, a lack of education, familiarity with digital tools, and trust-building around their use can also result in implementation failures. It is also important to recognize how digital divide challenges interact with other societal inequalities. For instance, while the gap is narrowing, women in LMICs are still 15% less likely than men to use mobile internet, with a more pronounced disparity in Sub-Saharan Africa (32%) and South Asia (31%)<sup>32</sup>.

#### Mitigation strategies.

- Prioritize investment in basic healthcare infrastructure alongside digital interventions.
- Consider whether GenAI is the highest impact way of addressing your use case, taking into account existing basic healthcare and digital infrastructure.
- Assess organizational digital readiness before implementing any AI tools to prevent unnecessary wastage of resources and guide investment priorities, utilizing tools such as the Global Digital Health Monitor<sup>33</sup>.

#### Reflecting societal biases and problems

Finally, when considering risks, we propose a framework which distinguishes risks that are reflections of societal biases and problems—which GenAI could amplify and propagate—from intrinsic risks, with different approaches to mitigation. Biases can result from the training data as discussed under ‘*limitations of the training data*’, but even representative data sets are subject to the assumptions or biases of individuals developing AI tools. In contexts with entrenched cultural biases such as gender discrimination, the need for language- and context-appropriate LLM training data could result in local gender stereotypes being reinforced<sup>34</sup>. Additionally, we must remain mindful that bad actors also have access to this technology—without the limitations and safeguards associated with operating within Responsible AI frameworks.

#### Mitigation strategies.

- Developing algorithms that include checks against bias and discrimination<sup>24</sup>.
- Establish legal governance frameworks to ensure adherence to best practices in Responsible AI.
- Consider the STANDING Together recommendations to support transparency regarding limitations of health datasets and proactive evaluation of their effect across population groups<sup>35</sup>.
- Understand and monitor how adversarial actors are using GenAI to accomplish their goals, and establish appropriate mitigations.

#### Conclusions

To conclude, we synthesize four key principles and four key risks to guide the use of GenAI for health, which we urge all policymakers, clinicians and academics working in healthcare to critically consider when reviewing potential GenAI tools. We discuss intrinsic model-based risks and potential mitigation strategies, and how they interact with wider societal biases, with GenAI both a potential tool for tremendous health impact, but also a ‘magnifying mirror’ for systemic social biases. Although trust-building at the user, healthcare system and



governmental levels will be crucial to enabling successful implementation of GenAI tools. Equally, we should also seek to empower individuals to make their own decisions about using GenAI to support their health needs, and engage end-users throughout the design, implementation, evaluation, and governance processes, recognizing that existing tools, without tailoring for a healthcare context, are already in widespread use globally.

## Data availability

No datasets were generated or analysed during the current study.

Received: 14 February 2025; Accepted: 30 April 2025;

Published online: 21 May 2025

## References

- Feuerriegel, S., Hartmann, J., Janiesch, C. & Zschech, P. Generative AI. *Bus. Inf. Syst. Eng.* **66**, 111–126 (2024).
- Friedland, A. What Are Generative AI, Large Language Models, and Foundation Models? *Center for Security and Emerging Technology* <https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/> (2023).
- Reddy, S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implement. Sci.* **19**, 27 (2024).
- Suthar, A., Joshi, V. & Prajapati, R. A Review of Generative Adversarial-Based Networks of Machine Learning Artificial Intelligence in Healthcare. in 37–56 <https://doi.org/10.4018/978-1-7998-8786-7.ch003> (2022).
- Bouderhem, R. Shaping the future of AI in healthcare through ethics and governance. *Humanit. Soc. Sci. Commun.* **11**, 1–12 (2024).
- Generative AI for Health in Low & Middle Income Countries | Center for Digital Health. <https://cdh.stanford.edu/generative-ai-health-low-middle-income-countries>.
- Ziebland, S. & McPherson, A. Making sense of qualitative data analysis: an introduction with illustrations from DIPEX (personal experiences of health and illness). *Med. Educ.* **40**, 405–414 (2006).
- Braun, V. & Clarke, V. Using thematic analysis in psychology. *Qualit. Res. Psychol.* **3**, 77–101 (2006).
- Lekadir, K. et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ* **388**, e081554 (2025).
- Skivington, K. et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ* **374**, n2061 (2021).
- Ethics and governance of artificial intelligence for health. <https://www.who.int/publications/item/9789240029200>.
- Health workforce requirements for universal health coverage and the Sustainable Development Goals. <https://www.who.int/publications/item/9789241511407>.
- EU Artificial Intelligence Act | Up-to-date developments and analyses of the EU AI Act. <https://artificialintelligenceact.eu/>.
- Partnerships will ensure inclusivity for Nigeria's AI strategy. <https://www.luminategroup.com/posts/news/partnerships-nigeria-ai-strategy>.
- Kenya launches project to develop National AI Strategy in collaboration with German and EU partners | Digital Watch Observatory. <https://dig.watch/updates/kenya-launches-project-to-develop-national-ai-strategy-in-collaboration-with-german-and-eu-partners> (2024).
- INDIAai | Pillars. *IndiaAI* <https://indiaai.gov.in/>.
- Inteligência Artificial. *Ministério da Ciência, Tecnologia e Inovação* <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>.
- Liu, X., Rivera, S. C., Moher, D., Calvert, M. J. & Denniston, A. K. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. *BMJ* **370**, m3164 (2020).
- Ji, Z. et al. Survey of hallucination in natural language generation. *ACM Comput. Surv.* **55**, 248:1–248:38 (2023).
- Gao, Y. et al. Retrieval-augmented generation for large language models: a survey. Preprint at <https://doi.org/10.48550/arXiv.2312.10997> (2024).
- Meng, X.-L. Data science and engineering with human in the loop, behind the loop, and above the loop. *Harvard Data Sci. Rev.* **5**, <https://s3.amazonaws.com/assets.pubpub.org/evp00pc4n8q70w1qb00z6uvok7nxjzd8.pdf> (2023).
- Schwartz, R., Dodge, J., Smith, N. A., Etzioni, O. & Green, A. I. *Commun. ACM* **63**, 54–63 (2020).
- Feuerriegel, S., Hartmann, J., Janiesch, C. & Zschech, P. Generative AI. SSRN Scholarly Paper at <https://doi.org/10.2139/ssrn.4443189> (2023).
- AI White Paper. *Reach* <https://www.reachdigitalhealth.org/ai-white-paper>.
- Jacaranda launches first-in-kind Swahili Large Language Model - Jacaranda Health. <https://jacarandahealth.org/jacaranda-launches-first-in-kind-swahili-large-language-model/> (2023).
- Generative AI and Intellectual Property Rights | SpringerLink. [https://link.springer.com/chapter/10.1007/978-94-6265-523-2\\_17](https://link.springer.com/chapter/10.1007/978-94-6265-523-2_17).
- Centre for Information Policy Leadership. *Centre for Information Policy Leadership* [https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl\\_pets\\_and\\_ppts\\_in\\_ai\\_mar25.pdf](https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_pets_and_ppts_in_ai_mar25.pdf).
- February 2025, C. C. // 10. Exclusive: Donors commit \$10M to include African languages in AI models. *Devex* <https://www.devex.com/news/sponsored/exclusive-donors-commit-10m-to-include-african-languages-in-ai-models-109044> (2025).
- Shikwambane, N. & <https://elixirdsgn.com>, E. D., Developer: Ronnie Mbugua, Website: InkubaLM: A small language model for low-resource African languages. *Lelapa* <https://lelapa.ai/inkubalm-a-small-language-model-for-low-resource-african-languages/> (2024).
- AfriMed-QA. <https://afrimedqa.com/>.
- From Connectivity to Services: Digital Transformation in Africa. *World Bank* <https://projects.worldbank.org/en/results/2023/06/27/from-connectivity-to-services-digital-transformation-in-africa>.
- Gender Gap - Mobile for Development. <https://www.gsma.com/r/gender-gap/>.
- Global Digital Health Monitor. *Global Digital Health Monitor* <https://digitalhealthmonitor.org>.
- Hada, R. et al. Akal Badi ya bias: an exploratory study of gender bias in Hindi Language Technology. in (2024).
- Alderman, J. E. et al. Tackling algorithmic bias and promoting transparency in health datasets: the STANDING Together consensus recommendations. *Lancet Digit. Health* **7**, e64–e88 (2025).

## Acknowledgements

We incorporate learnings from two recent roundtable events (October and December 2024) and a series of semi-structured interviews (November – December 2024) with experts in Generative AI and digital health, encompassing perspectives from academia, funding bodies, implementers, and health system leaders, funded by the Advancing Health Online Initiative (AHO). The authors received no funding support for the writing of this article. We would like to thank Chris Struhar for role in conceptualizing the LLM task strengths framework, and James Parkhouse for his feedback on manuscript drafts.

## Author contributions

I.D.V.H. wrote the original draft of this viewpoint, incorporating the LLM task strengths framework conceptualized by K.X.J., with feedback from E.L. and K.X.J. All authors contributed equally to multiple rounds of re-drafting and edits and have read and approved the final manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at

<https://doi.org/10.1038/s41746-025-01695-y>.

**Correspondence** and requests for materials should be addressed to Eleni Linos.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025