

Viewpoint

Empathy and Equity: Key Considerations for Large Language Model Adoption in Health Care

Erica Koranteng^{1*}, MBChB, MBE; Arya Rao^{1*}, BA; Efren Flores¹, MD; Michael Lev¹, MD; Adam Landman¹, MD, MIS, MHS, MS; Keith Dreyer¹, PhD, DO; Marc Succi², MD

¹Harvard Medical School, Boston, MA, United States

²Massachusetts General Hospital, Boston, United States

* these authors contributed equally

Corresponding Author:

Marc Succi, MD

Massachusetts General Hospital

55 Fruit St

Boston, 02114

United States

Phone: 1 617 935 9144

Email: msucci@mgh.harvard.edu

Abstract

The growing presence of large language models (LLMs) in health care applications holds significant promise for innovative advancements in patient care. However, concerns about ethical implications and potential biases have been raised by various stakeholders. Here, we evaluate the ethics of LLMs in medicine along 2 key axes: empathy and equity. We outline the importance of these factors in novel models of care and develop frameworks for addressing these alongside LLM deployment.

(*JMIR Med Educ* 2023;9:e51199) doi: [10.2196/51199](https://doi.org/10.2196/51199)

KEYWORDS

ChatGPT; AI; artificial intelligence; large language models; LLMs; ethics; empathy; equity; bias; language model; health care application; patient care; care; development; framework; model; ethical implication

Introduction

The rapid proliferation of applications that leverage the ability of large language models (LLMs) to use large amounts of complex information to find relevant patterns and apply them to novel use cases promises great innovation in health care and many other sectors. Many health care applications, such as clinical decision support, patient education, electronic health records (EHRs), and workflow optimization, have been proposed [1]. Despite the immense potential advantages of this technology, various key stakeholders have raised concerns regarding its ethical implications and potential perpetuation of existing biases and structural barriers [2-6]. Furthermore, its growing usage in the health care setting also raises the concern of transparency or disclosure about its use and role in patient management. Ethically incorporating LLMs into health care delivery requires honest dialogue about the principles we aim to uphold in patient care and a comprehensive analysis of the various ways in which LLMs could bolster or impair these.

Studies have demonstrated the utility of LLMs as a clinical decision support tool in various settings, including in triage, diagnostics, and treatment [7-11]. While LLMs show great promise in improving the efficiency of clinical workflows, they lack one key facet of physician-patient encounters: empathy. Though LLMs can be trained to use empathetic language [12] and have been able to use empathetic language in patient interactions [13], this concept of artificial empathy is easily distinguishable from real empathy from a patient's perspective, and real empathy matters to patients [14]. The concept of artificial empathy, which aims to imbue artificial intelligence (AI) with human-like empathy, ought not to be considered interchangeable with human empathy. Efforts made to design artificial empathy, while commendable, should aim to be complementary to human empathy in order to avoid further isolating patients in their time of need by destroying the therapeutic alliance between patients and physicians [15]. Loneliness is one of the key public health crises of our time, and conflating technology with human-to-human interaction will only exacerbate this [16]. Empathetic care for patients should be one of the core mandates of the health care sector, and true

empathy requires human connection. Therefore, while LLMs show great promise in clinical workflows, they should augment, rather than replace, physician-led care (Table 1).

In addition to empathy, equity is crucial in novel models of care. The current most popular LLMs, including ChatGPT, Bard, Med-PaLM, and others, are trained on vast sources of data, including wide swaths of the internet. These sources are rife with inherent biases and lack transparency regarding the contents of the training data sets. They also lack specific evaluation of model biases, which may be harbingers of ethical dilemmas via the rapid incorporation of LLMs into clinical spaces. While there is little consensus regarding the degree of bias in current LLMs, in most embedding models, which have similar underlying architecture, there is evidence of racial, gender, and age bias [17]. LLMs have been demonstrated to associate negative terms with given names that are popular among the African American as well as with the masculine poles of most gender axes [17]. Until systematic evaluation of LLMs is performed in clinical use cases to understand and mitigate biases against vulnerable demographics, careful risk-benefit calculations and a regulatory framework should be implemented by relevant governing bodies before LLMs are permitted in clinical care. This framework must ensure that these models are improving health care delivery and outcomes for all. Importantly, the US Food and Drug Administration lacks a robust authorization pathway for software as a medical device; this in itself is challenging, and given the rapid development of LLMs, would benefit from expeditious guidelines [18] (see Table 2 for proactive measures to ensure the equitable incorporation of LLMs into health care). Following a previously published ethical framework for integrating innovative domains into medicine, we suggest an LLM framework guided by Blythe et al [19] grounded in principled primary motivations as detailed in Tables 1 and 2.

Despite these ethical risks, the potential benefits of incorporating LLMs into health care are numerous. LLMs are adept at quickly synthesizing large amounts of complex data, which can form the basis for numerous applications in the health care sector, including the management and interpretation of EHRs and clinical notes, adjuncts for patient visits (eg, encounter transcription and patient translation), billing for medical services, patient education, and more [20,21]. Thus, the key

ethical question at hand is as follows: do the benefits outweigh the risks?

From a utilitarian perspective, we must consider this question to not only enhance decision-making but also take advantage of opportunities to mitigate potential harms. Proposals for the incorporation of a systematized, frequently reevaluated method of bias evaluation into clinical applications of LLMs [3], the addition of human verification steps at both the input and output stages for LLM-guided generation of clinical texts [22], and the implementation of self-questioning—a novel prompting strategy that encourages prompt iteration to improve accuracy in a medical context—are all steps in the correct direction. Comprehensive frameworks that include the use of diverse training data sources and continuous evaluation of bias, such as those proposed by the World Economic Forum and the Coalition for Health AI, can provide useful guardrails as new proposals for ethical validation and have been tested [23,24]. Furthermore, ensuring that physicians are actively involved in the development and evaluation of LLMs for health care is essential in keeping with a physician-led approach. Strategies such as these are key in navigating the ethics of empathy and equity in the development of novel clinical technologies.

It is essential to approach the ethical conundrums of LLM adoption in clinical care with a balanced perspective. LLMs that were built on data with inherent systemic biases must be implemented strategically into health care through a justice-oriented innovation lens to advance health equity. To keep pace with the accelerated adoption of LLMs in the clinic, ethical evaluations should be conducted together with an evaluation of use case efficacy to ensure both efficient and ethical health care. A complete assessment of the risks and benefits associated with this technology—an admittedly challenging task—may remain elusive if not tested in real-world settings. Clinical use cases of LLMs are already being tested; delaying collaboration among all stakeholders, including health care professionals, ethicists, AI researchers, and (crucially) patients, will only delay the discovery of potential harms. Real-world pilots, therefore, should be deployed alongside regular monitoring, oversight, and feedback from all parties. As we collectively seek to make full use of this exciting new technology, we must keep empathy and equity at the forefront of our minds.

Table 1. Approaches to the incorporation of large language models (LLMs) in clinical care.

Approach	Primary motivation	Impact on empathy and health equity
LLM-led clinical care or patient-facing LLMs	Advancement-driven: incorporation of new and sophisticated technologies mainly aimed at improving efficiency	<ul style="list-style-type: none"> • Perpetuates and exacerbates inequities and biases on which it was built, making it detrimental to achieving health equity • Replaces human empathy with artificial empathy, which threatens patient dignity
Physician-led LLM incorporation in clinical care	Holistic, equitable, and empathetic health care delivery	<ul style="list-style-type: none"> • Early recognition of ways in which models perpetuate inequity and appropriate measures to prevent this • Opportunity to actively leverage LLMs to mitigate existing inequities • Use of LLMs as tools in a physician's toolkit allows more time to engage in empathetic dialogue with patients

Table 2. Potential proactive measures for promoting equitable incorporation of large language models (LLMs) into clinical care.

Stakeholder	Examples of proactive measures
Regulatory bodies	<ul style="list-style-type: none"> Development of robust regulations for software as a medical device that ensure appropriate strategies for (1) continuous evaluation of evolving technology and (2) assessment of use cases that have significant impact in health care given the broad capabilities of LLMs
Professional societies	<ul style="list-style-type: none"> Development and continuous updates of guidelines for equitable use of LLMs in health care Allocation of grant funding toward projects that aim to use LLMs to ameliorate inequities
Journals	<ul style="list-style-type: none"> Prioritizing publications that focus on (1) novel methods of leveraging LLMs for equitable care delivery and (2) comparisons of use cases of LLMs for equitable care delivery
Software developers and industry	<ul style="list-style-type: none"> Collaboration with health care workers on model improvement strategies that improve health equity

Acknowledgments

This project was supported in part by an award from the National Institute of General Medical Sciences (T32GM144273). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

Conflicts of Interest

EF is co-chair of the Radiological Society of North America (RSNA) Health Equity Committee; associate editor and editorial board member of the Journal of the American College of Radiology (JACR); has received speaker honoraria for academic Grand Rounds, from WebMD and from GO2 for Lung Cancer foundation; GO2 Foundation Travel support; grant funding from NCI K08 1K08CA270430-01A1. ML is a consultant for GE Healthcare and for Takeda, Roche, and SeaGen Pharma. AL is a consultant for the Abbott Medical Device Cybersecurity Council.

References

- Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell.* 2023 May 4;6:1169595 [FREE Full text] [doi: [10.3389/frai.2023.1169595](https://doi.org/10.3389/frai.2023.1169595)] [Medline: [37215063](https://pubmed.ncbi.nlm.nih.gov/37215063/)]
- Rozado D. Wide range screening of algorithmic bias in word embedding models using large sentiment lexicons reveals underreported bias types. *PLoS One.* 2020 Apr 21;15(4):e0231189 [FREE Full text] [doi: [10.1371/journal.pone.0231189](https://doi.org/10.1371/journal.pone.0231189)] [Medline: [32315320](https://pubmed.ncbi.nlm.nih.gov/32315320/)]
- Garrido-Muñoz I, Montejó-Ráez A, Martínez-Santiago F, Ureña-López LA. A survey on bias in deep NLP. *Appl Sci.* 2021 Apr 02;11(7):3184 [doi: [10.3390/app11073184](https://doi.org/10.3390/app11073184)]
- Liu R, Jia C, Wei J, Xu G, Vosoughi S. Quantifying and alleviating political bias in language models. *Artificial Intelligence.* 2022 Mar;304:103654 [doi: [10.1016/j.artint.2021.103654](https://doi.org/10.1016/j.artint.2021.103654)]
- Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. *Lancet Digital Health.* 2023 Jun;5(6):e333-e335 [doi: [10.1016/s2589-7500\(23\)00083-3](https://doi.org/10.1016/s2589-7500(23)00083-3)]
- Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digit Med.* 2023 Jul 06;6(1):120 [FREE Full text] [doi: [10.1038/s41746-023-00873-0](https://doi.org/10.1038/s41746-023-00873-0)] [Medline: [37414860](https://pubmed.ncbi.nlm.nih.gov/37414860/)]
- Rao A, Kim J, Kamineni M, Pang M, Lie W, Dreyer KJ, et al. Evaluating GPT as an adjunct for radiologic decision making: GPT-4 versus GPT-3.5 in a breast imaging pilot. *J Am Coll Radiol.* 2023 Oct;20(10):990-997 [doi: [10.1016/j.jacr.2023.05.003](https://doi.org/10.1016/j.jacr.2023.05.003)] [Medline: [37356806](https://pubmed.ncbi.nlm.nih.gov/37356806/)]
- Rao A, Kim J, Kamineni M, Pang M, Lie W, Succi M. Evaluating ChatGPT as an adjunct for radiologic decision-making. *medRxiv.* Preprint posted online February 7, 2023 [FREE Full text] [doi: [10.1101/2023.02.02.23285399](https://doi.org/10.1101/2023.02.02.23285399)] [Medline: [36798292](https://pubmed.ncbi.nlm.nih.gov/36798292/)]
- Rao A, Pang M, Kim J, Kamineni M, Lie W, Prasad A, et al. Assessing the utility of ChatGPT throughout the entire clinical workflow. *medRxiv.* Preprint posted online February 26, 2023 [FREE Full text] [doi: [10.1101/2023.02.21.23285886](https://doi.org/10.1101/2023.02.21.23285886)] [Medline: [36865204](https://pubmed.ncbi.nlm.nih.gov/36865204/)]
- Varney ET, Lee CI. The potential for using ChatGPT to improve imaging appropriateness. *J Am Coll Radiol.* 2023 Oct;20(10):988-989 [doi: [10.1016/j.jacr.2023.06.005](https://doi.org/10.1016/j.jacr.2023.06.005)] [Medline: [37400048](https://pubmed.ncbi.nlm.nih.gov/37400048/)]
- Chonde DB, Pourvaziri A, Williams J, McGowan J, Moskos M, Alvarez C, et al. RadTranslate: an artificial intelligence-powered intervention for urgent imaging to enhance care equity for patients with limited English proficiency during the COVID-19 pandemic. *J Am Coll Radiol.* 2021 Jul;18(7):1000-1008 [FREE Full text] [doi: [10.1016/j.jacr.2021.01.013](https://doi.org/10.1016/j.jacr.2021.01.013)] [Medline: [33609456](https://pubmed.ncbi.nlm.nih.gov/33609456/)]

12. Sharma A, Lin IW, Miner AS, Atkins DC, Althoff T. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nat Mach Intell*. 2023 Jan 23;5(1):46-57 [doi: [10.1038/s42256-022-00593-2](https://doi.org/10.1038/s42256-022-00593-2)]
13. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med*. 2023 Jun 01;183(6):589-596 [doi: [10.1001/jamainternmed.2023.1838](https://doi.org/10.1001/jamainternmed.2023.1838)] [Medline: [37115527](https://pubmed.ncbi.nlm.nih.gov/37115527/)]
14. Guidi C, Traversa C. Empathy in patient care: from 'Clinical Empathy' to 'Empathic Concern'. *Med Health Care Philos*. 2021 Dec 01;24(4):573-585 [FREE Full text] [doi: [10.1007/s11019-021-10033-4](https://doi.org/10.1007/s11019-021-10033-4)] [Medline: [34196934](https://pubmed.ncbi.nlm.nih.gov/34196934/)]
15. Smoktunowicz E, Barak A, Andersson G, Banos RM, Berger T, Botella C, et al. Consensus statement on the problem of terminology in psychological interventions using the internet or digital components. *Internet Interv*. 2020 Sep;21:100331 [FREE Full text] [doi: [10.1016/j.invent.2020.100331](https://doi.org/10.1016/j.invent.2020.100331)] [Medline: [32577404](https://pubmed.ncbi.nlm.nih.gov/32577404/)]
16. Jaffe S. US Surgeon General: loneliness is a public health crisis. *The Lancet*. 2023 May;401(10388):1560 [doi: [10.1016/s0140-6736\(23\)00957-1](https://doi.org/10.1016/s0140-6736(23)00957-1)]
17. Nadeem M, Bethke A, Reddy S. StereoSet: Measuring stereotypical bias in pretrained language models. 2021 Presented at: 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing; 2021; Online p. 5356-5371 [doi: [10.18653/v1/2021.acl-long.416](https://doi.org/10.18653/v1/2021.acl-long.416)]
18. Dortche K, McCarthy G, Banbury S, Yannatos I. Promoting health equity through improved regulation of artificial intelligence medical devices. *JSPG*. 2023 Jan 23;21(03) [doi: [10.38126/JSPG210302](https://doi.org/10.38126/JSPG210302)]
19. Blythe JA, Flores EJ, Succi MD. Justice and innovation in radiology. *J Am Coll Radiol*. 2023 Jul;20(7):667-670 [doi: [10.1016/j.jacr.2023.05.005](https://doi.org/10.1016/j.jacr.2023.05.005)] [Medline: [37315912](https://pubmed.ncbi.nlm.nih.gov/37315912/)]
20. Jiang LY, Liu XC, Nejatian NP, Nasir-Moin M, Wang D, Abidin A, et al. Health system-scale language models are all-purpose prediction engines. *Nature*. 2023 Jul 07;619(7969):357-362 [FREE Full text] [doi: [10.1038/s41586-023-06160-y](https://doi.org/10.1038/s41586-023-06160-y)] [Medline: [37286606](https://pubmed.ncbi.nlm.nih.gov/37286606/)]
21. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. *Lancet Digital Health*. 2023 Apr;5(4):e179-e181 [doi: [10.1016/s2589-7500\(23\)00048-1](https://doi.org/10.1016/s2589-7500(23)00048-1)]
22. Singh S, Djalilian A, Ali MJ. ChatGPT and ophthalmology: exploring its potential with discharge summaries and operative notes. *Semin Ophthalmol*. 2023 Jul 03;38(5):503-507 [doi: [10.1080/08820538.2023.2209166](https://doi.org/10.1080/08820538.2023.2209166)] [Medline: [37133418](https://pubmed.ncbi.nlm.nih.gov/37133418/)]
23. A Blueprint for Equity and Inclusion in Artificial Intelligence 2022. World Economic Forum. URL: <https://www.weforum.org/whitepapers/a-blueprint-for-equity-and-inclusion-in-artificial-intelligence/> [accessed 2023-11-14]
24. Blueprint for Trustworthy AI Implementation Guidance and Assurance for Healthcare 2023. Coalition for Health AI. URL: https://www.coalitionforhealthai.org/papers/blueprint-for-trustworthy-ai_V1.0.pdf [accessed 2023-11-14]

Abbreviations

- AI:** artificial intelligence
EHR: electronic health record
LLM: large language model

Edited by K Venkatesh; submitted 24.07.23; peer-reviewed by SY Tan, B Bizzo, YD Cheng, L Zhu; comments to author 28.09.23; revised version received 01.10.23; accepted 14.10.23; published 28.12.23

Please cite as:

Koranteng E, Rao A, Flores E, Lev M, Landman A, Dreyer K, Succi M
Empathy and Equity: Key Considerations for Large Language Model Adoption in Health Care
JMIR Med Educ 2023;9:e51199
URL: <https://mededu.jmir.org/2023/1/e51199>
doi: [10.2196/51199](https://doi.org/10.2196/51199)
PMID: [38153778](https://pubmed.ncbi.nlm.nih.gov/38153778/)

©Erica Koranteng, Arya Rao, Efren Flores, Michael Lev, Adam Landman, Keith Dreyer, Marc Succi. Originally published in JMIR Medical Education (<https://mededu.jmir.org/>), 28.12.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on <https://mededu.jmir.org/>, as well as this copyright and license information must be included.