



OPEN ACCESS

EDITED BY

Guglielmo M. Trovato,
European Medical Association (EMA), Belgium

REVIEWED BY

Suhas Srinivasan,
Stanford University, United States
Teresa Abbattista,
Senigallia Hospital, Italy

*CORRESPONDENCE

Jeena Joseph
✉ jeenajoseph005@gmail.com

RECEIVED 08 June 2025

ACCEPTED 08 July 2025

PUBLISHED 23 July 2025

CITATION

Joseph J (2025) Algorithmic bias in public health AI: a silent threat to equity in low-resource settings.
Front. Public Health 13:1643180.
doi: 10.3389/fpubh.2025.1643180

COPYRIGHT

© 2025 Joseph. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Algorithmic bias in public health AI: a silent threat to equity in low-resource settings

Jeena Joseph*

Department of Computer Applications, Marian College Kuttikkanam Autonomous, Kuttikkanam, Kerala, India

KEYWORDS

algorithmic bias, artificial intelligence in public health, health equity, low-resource settings, digital health disparities, inclusive AI

Public health systems have long touted the integration of artificial intelligence (AI) as the game-changing innovation with the potential to transform the delivery of health services, the diagnosis of diseases, and the protection of people in the community. Ranging from sophisticated disease surveillance to computer-aided diagnosis applications, AI is being used in myriad contexts within public health, with the promise to make things more efficient, personalized, and accessible. Underlying all this promise, however, is one much under-investigated concern: the threat posed by algorithmic bias. In contexts where margins for error are thin and exclusion is likely to have high stakes, biased AI systems can widen the existing health gaps instead of bridging them (1). Hidden in code, imperceptible in action, this is an urgent threat to consider.

AI systems are only as effective as the data used to train them and the assumptions under which they are created (2). Many public health AI models, however, draw from data sets in populations which are unrepresentative of those in the low- and middle-income countries (LMICs). The resulting data inequity means the algorithms then do not capture the cultural, linguistic, genetic, or environmental variety in the underserved populations (3). Consequently, health AI systems will systematically underdiagnose, misclassify, or outright ignore patterns in the non-conforming population. This is especially risky in settings where the healthcare infrastructure is already under stress, such as where digital health interventions are viewed as cost-saving measures (4).

Sources of health AI bias are multi-pronged. They begin with the data collection stage, where the majority of data sets come from city hospitals, research centers, or wealthy countries. Such data sets systematically exclude rural patients, ethnic minorities, indigenous people, or socially marginalized groups (2). During the labeling phase, clinical annotations can introduce bias when medical thresholds and definitions are drawn from the dominant population without accounting for cultural or biological variation (5, 6). Algorithmic in its own right, models optimized to be as accurate as possible might disregard considerations of fairness, resulting in consistently disparate performance across subgroups (7, 8). Finally, at the deployment phase, systems can act capriciously if brought into use in settings unlike those in which it has been developed and evaluated (2). Cumulatively, the problems lead to a series of silent reaffirmations of structural disparities.

To more rigorously frame the sources of AI bias, it is important to refer to well-established typologies in algorithmic bias literature. Public health AI typically suffers from historic bias, by which prior injustices—inequities in access to care or discriminatory health policy, say—are embedded within datasets for learning (9, 10). Representation bias is present when samples from urban, wealthy, or connected groups lead to the ignoring of samples from rural, indigenous, or disenfranchised groups (2). In addition, measurement bias is present when health endpoints are approximated with the help of proxy

variables—hospital attendance or smartphone usage, say—strikingly different between socioeconomic or even cultural environments (11). These biases are again compounded by aggregation bias, by which models assume homogeneity between heterogeneous groups, and deployment bias, by which tools developed within high-resource environments are simply implemented without modification into low-resource environments (12). Recognition of these forms of bias allows more structured comprehension of how algorithmic harm takes place and where mitigation efforts must be targeted.

Real-world instances of algorithmic bias in public health reinforce the urgency of addressing these typologies. A well-documented case of historic bias is evident in a widely used U.S. healthcare risk prediction algorithm that systematically underestimated the health needs of Black patients by using prior healthcare expenditure as a proxy—unintentionally replicating patterns of historical underutilization of care (13). Representation bias has surfaced in sepsis prediction models developed in high-income settings that showed significantly reduced accuracy among Hispanic patients due to unbalanced training data (14). In the realm of measurement bias, India's digital health initiatives have often relied on smartphone usage for patient engagement, effectively excluding large segments of women, older adult individuals, and rural populations who lack digital access (15). Deployment bias was starkly illustrated during the COVID-19 pandemic through the Aarogya Setu contact tracing app, which failed to reach populations without smartphones, particularly in rural and low-income communities, raising concerns about uneven public health protection (8). These examples underscore that algorithmic bias is not merely theoretical; it has material consequences that can exacerbate the very inequalities public health systems aim to mitigate.

At the heart of algorithmic bias lies the issue of structural data exclusion. Many AI systems draw their inputs from data sources that systematically omit rural populations, marginalized castes, indigenous groups, or those without digital access. This exclusion is not accidental—it is a byproduct of how data is collected, labeled, and interpreted (16). Clinical annotations are often standardized using thresholds derived from dominant populations, ignoring genetic, environmental, or cultural differences that influence health outcomes (17). When these biased models are used in low-resource contexts, they risk making inaccurate diagnoses or failing to detect public health emergencies altogether (18). This is not a technical shortcoming but a structural flaw, one that reflects who is seen, heard, and counted in the design of digital health tools.

These effects of algorithmic prejudice in public health do not remain theoretical. They appear as material harm. In Brazil, AI models trained with city-level data did not capture rural disease epidemics from environmental and socioeconomic features missing from the training data (19). In India, mobile applications using smartphone availability for self-reporting or tele-consultation exclude significant parts of the population who have no familiarity with digital technologies or simply no means to own or access mobile phones (20, 21). Such exclusion is neither a technical error nor simply an oversight; it is the further extension of historic inequalities now replicated through ostensibly neutral technology. The illusory

assumption that technology has intrinsic objectivity obscures the need for contextual intelligence in the crafting and deployment of systems.

A central risk in today's AI-in-health interventions is the myth of neutrality. It is assumed frequently that AI systems lack biases inherent in human beings due to their use of data and statistically grounded reasoning. AI models, however, are created, trained, and evaluated by people, and cannot help but reflect the developers' assumptions, tastes, and blind spots. This is particularly problematic where AI development is concentrated in the Global North, and deployment in LMICs has radically dissimilar social and health contexts (22). The outcome is one digital colonialism where technologies created in one context get transferred to another with little or no adaptation or local contribution.

This myth of neutrality obscures the fact that AI systems are human artifacts—shaped by the data they are trained on, the objectives they are given, and the assumptions of their developers. Public health AI often reflects the institutional biases of the healthcare systems and research ecosystems from which it originates (11). For instance, the development of AI tools predominantly in high-income countries leads to the encoding of social, cultural, and biological norms that are misaligned with the realities of low-resource settings (23). When these systems are deployed in vastly different environments, they bring with them embedded assumptions that do not travel well. As a result, the notion that AI offers a 'scientific' or 'impartial' solution to global health problems becomes not only misleading but dangerous—especially when it prevents critical reflection or accountability for systemic failures in underrepresented communities (24).

Public health systems in low-resource settings tend to be poorly equipped to detect or respond to the impacts of biased AI. They lack effective regulatory frameworks to govern AI, have weak or unenforced data protection laws, or lack the technical skills to conduct audits of AI systems (25). This is a perfect storm in which suboptimal algorithms can be injected into national health policy with inadequate oversight. Digital exclusion further exacerbates these dynamics. Underprivileged populations who already have inadequate access to healthcare services are also likely to be marginalized from digital interventions, further entrenching inequalities (23).

Tackling algorithmic bias in public health requires more than technical interventions—it demands a reengineering of values into systems (26). Equity cannot be retrofitted; it must be a foundational design principle. This means ensuring that development teams are multidisciplinary and inclusive of voices from the Global South, marginalized communities, and local health ecosystems (27). Participatory design, where affected populations co-create and critique AI tools, should be standard practice rather than an afterthought (28). Moreover, the use of fairness audits, synthetic data for underrepresented cases, and multilingual NLP models can help mitigate systemic blind spots (29). Countries like India, with their complex demographic and linguistic diversity, offer both cautionary tales and promising models. If India pioneers equity-focused AI governance, it could

serve as a replicable framework for other low- and middle-income countries facing similar challenges. To meet these challenges, we need a paradigm shift in the conceptualization, development, and deployment of AI for public health. More than technical sophistication or predictive precision, developers and policymakers need to support an equity-focused orientation. This work would start with inclusive data practices with the express goal to capture the full diversity of the population. Strategies in data collection need to be redesigned to encompass rural areas, under-represented languages, and marginalized groups (30). Local health workers as well as community-based organizations can be critical allies in the creation of representative as well as relevant.

Bias reduction should also be integrated into the development and assessment of AI systems. AI tools should be subject to fairness audits in front of deployment to examine their accuracy against different demographic and socio-economic groups (31). Imbalances in the accuracy of predictions, false positives, and false negatives should be recorded and rectified. Crucially, such evaluations should not be looked at as singular processes but as continuing routines in the life cycle of the AI system (26). Transparency and explainability should be valued as well. Public health officials as well as community stakeholders should have insight into the workings of these systems, the data used, and the assumptions.

Health technologists need to collaborate with social scientists, public health practitioners, ethicists, and impacted communities to ensure AI systems remain contextually situated. Such collaboration should not only occur in the development stage but also in deployment, monitoring, and evaluation. Participatory methods, in which the individuals at the heart of the issue participate in defining the tools impacting them, can bridge the disconnect between technical innovation and everyday experience (32).

One new path to enhancing equity in public health AI is to utilize synthetic data generation to bridge the gaps in under-representative populations—as long as it is carried out ethically (33). Generative models such as GANs can, for instance, mimic data from rare diseases or rural settings where authentic data is in short supply. This cannot entirely supplant genuine field data, but it can supplement it at early stages to circumvent algorithmic blind spots. Also, incorporating local languages and indigenous knowledge into NLP-based health applications has the potential to expand coverage in multilingual cultures (34). Decentralized AI architectures, in which models get trained locally using varied datasets in place of central data repositories, have as-yet untapped potential. Such methods could limit prejudice and amplify cultural emphasis.

India is an interesting case study in the intersection of AI, public health, and equity. With its demographics, geographic variation, and hybrid public-private healthcare delivery system, India has adopted AI at breakneck velocity in applications from disease surveillance to diagnosis to telemedicine (35). The velocity has, to date, not been paired with a strong framework to determine the threat of algorithmic bias. The Aarogya Setu app used in the COVID-19 response to track contacts serves to highlight the strength as well as the limitations of digital public health

technologies (36). It reached the millions, but only those with smartphones, with severe privacy as well as data governance worries. AI models to predict diabetes as well as cardiovascular risk have had promise, but whether their findings can generalize to the varied subgroups in India is uncertain given the variation in regions in terms of dietary intake, lifestyle, as well as availability of care.

India's experience highlights the necessity to develop country-specific strategies for AI in consideration of socio-cultural, linguistic, and infrastructural diversity (35). Developing AI systems in India with ethics and awareness about biases needs technological investment as well as in local research, grassroots engagement, and policy reform. If India is successful in developing an inclusive AI environment, it can be replicated in similar challenges in other LMICs.

No single stakeholder is responsible for addressing algorithmic bias in public health AI. It is the collective responsibility of governments, developers, funders, researchers, and communities (37). International reformers, international funders, and global health institutions need to make bias assessments and equity measures mandatory as part of project proposals and program evaluations (16). Tech companies and research institutions need to put open-source models and data sets with global diversity as their priority. International regulators need to frame explicit guidelines for the deployment of ethical AI with mechanisms to ensure community redress and accountability (38).

In the end, the goal is not to eliminate AI but to rethink it as an inclusionary technology, rather than an exclusion technology. This means designing systems which do not simply reproduce what is effective in privileged situations but also develop in accordance with the needs of those who have long been unserved. It means designing feedback mechanisms in which users can challenge, critique, and shape AI systems. It means committing to the value that equity needs to be engineered—and not just assumed.

In the years to come, artificial intelligence will further leave its mark in world health. As it does so, the hidden biases embedded in algorithms will determine who is diagnosed, who is treated, and who is left behind. The stakes could hardly be higher. Without conscious effort to detect, minimize, and block algorithmic bias, we risk automating injustice and entrenching inequality in the very technologies we hope to enhance. The time to act is now, before such biases harden into the default reasoning of digital health.

By making fairness, transparency, and inclusivity the foundation in the development and deployment of public health AI, we can turn the potential threat into an empowering force toward health equity. It will take more than good intentions to do it, however; it will take critical analysis, system reform, as well as the willingness to break the seductive myth about neutrality.

Author contributions

JJ: Writing – review & editing, Conceptualization, Writing – original draft.

Funding

The author declares that no financial support was received for the research and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Stypińska J, Franke A. AI revolution in healthcare and medicine and the (re-)emergence of inequalities and disadvantages for ageing population. *Front Sociol.* (2023) 7:1038854. doi: 10.3389/fsoc.2022.1038854
2. Cross JL, Choma MA, Onofrey JA. Bias in medical AI: Implications for clinical decision-making. *PLOS Digit Health.* (2024) 3:e0000651. doi: 10.1371/journal.pdig.0000651
3. Fletcher RR, Nakeshimana A, Olubeko O. Addressing fairness, bias, and appropriate use of artificial intelligence and machine learning in global health. *Front Artif Intell.* (2021) 3:561802. doi: 10.3389/frai.2020.561802
4. Istasy P, Lee WS, Iansavitchene A, Upshur R, Sadikovic B, Lazo-Langner A, et al. The impact of artificial intelligence on health equity in oncology: a scoping review. *Blood.* (2021) 138:4934. doi: 10.1182/blood-2021-149264
5. Perets O, Stagno E, Yehuda EB, McNichol M, Celi LA, Rappoport N, et al. *Inherent Bias in Electronic Health Records: A Scoping Review of Sources of Bias.* (2024) Available from: <http://medrxiv.org/lookup/doi/10.1101/2024.04.09.24305594> (Accessed June 8, 2025).
6. Straw I, Callison-Burch C. Artificial Intelligence in mental health and the biases of language based models. *PLoS ONE.* (2020) 15:e0240376. doi: 10.1371/journal.pone.0240376
7. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med.* (2021) 27:2176–82. doi: 10.1038/s41591-021-01595-0
8. Khoshnava Azar A, Draghi B, Rotalinti Y, Myles P, Tucker A. The impact of bias on drift detection in AI health software. In: Juarez JM, Marcos M, Stiglic G, Tucker A, editors. *Artificial Intelligence in Medicine.* Cham: Springer Nature Switzerland (2023). p. 313–22. (Lecture Notes in Computer Science; vol. 13897). Available from: https://link.springer.com/10.1007/978-3-031-34344-5_37 doi: 10.1007/978-3-031-34344-5_37 (Accessed 2025 Jun 8).
9. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med.* (2018) 169:866–72. doi: 10.7326/M18-1990
10. Brogan J. The Next Era of Biomedical Research: Prioritizing Health Equity in The Age of Digital Medicine. *Voices Bioeth.* vol. 7. (2021). Available from: <https://journals.library.columbia.edu/index.php/bioethics/article/view/8854> (Accessed June 22, 2025).
11. Gichoya JW, Thomas K, Celi LA, Safdar N, Banerjee I, Banja JD, et al. AI pitfalls and what not to do: mitigating bias in AI. *Br J Radiol.* (2023) 96:20230023. doi: 10.1259/bjr.20230023
12. Tejani AS, Ng YS Xi Y, Rayan JC. Understanding and mitigating bias in imaging artificial intelligence. *RadioGraphics.* (2024) 44:e230067. doi: 10.1148/rg.230067
13. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* (2019) 366:447–53. doi: 10.1126/science.aax2342
14. Cronjé HT, Katsiferis A, Elsenburg LK, Andersen TO, Rod NH, Nguyen TL, et al. Assessing racial bias in type 2 diabetes risk prediction algorithms. *PLOS Glob Public Health.* (2023) 3:e0001556. doi: 10.1371/journal.pgph.0001556
15. Estiri H, Strasser ZH, Rashidian S, Klann JG, Waghlikar KB, McCoy TH, et al. An objective framework for evaluating unrecognized bias in medical AI models predicting COVID-19 outcomes. *J Am Med Inform Assoc.* (2022) 29:1334–41. doi: 10.1093/jamia/ocac070

Generative AI statement

The author declares that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

16. Aquino YSJ, Carter SM, Houssami N, Braunack-Mayer A, Win KT, Degeling C, et al. Practical, epistemic and normative implications of algorithmic bias in healthcare artificial intelligence: a qualitative study of multidisciplinary expert perspectives. *J Med Ethics.* (2025) 51:420–8. doi: 10.1136/jme-2022-108850
17. Chase AC. Ethics of AI. *Voices Bioeth.* Vol. 6 (2020).
18. Cardenas S, Vallejo-Cardenas SF. Continuing the conversation on how structural racial and ethnic inequalities affect AI biases. In: *2019 IEEE International Symposium on Technology and Society (ISTAS).* Medford, MA, USA: IEEE (2019). p. 1–7. Available from: <https://ieeexplore.ieee.org/document/8937853/> (Accessed June 22, 2025).
19. Bellanda VCF, Medeiros AS, Ferraz DA. Transforming Brazilian healthcare with AI: progress and future perspectives. *Discov Health Syst.* (2025) 4:47. doi: 10.1007/s44250-025-00227-5
20. Feinberg I, Menon J, Smith R, Rajeev JG, Kumar RK, Banerjee A. Potential for mobile health (mHealth) prevention of cardiovascular diseases in Kerala: a population-based survey. *Indian Heart J.* (2017) 69:182–99. doi: 10.1016/j.ihj.2016.11.004
21. Rajkumar E, Gopi A, Joshi A, Thomas AE, Arunima NM, Ramya GS, et al. Applications, benefits and challenges of telehealth in India during COVID-19 pandemic and beyond: a systematic review. *BMC Health Serv Res.* (2023) 23:7. doi: 10.1186/s12913-022-08970-8
22. Weissglass DE. Contextual bias, the democratization of healthcare, and medical artificial intelligence in low- and middle-income countries. *Bioethics.* (2022) 36:201–9. doi: 10.1111/bioe.12927
23. Yu L, Zhai X. Use of artificial intelligence to address health disparities in low- and middle-income countries: a thematic analysis of ethical issues. *Public Health.* (2024) 234:77–83. doi: 10.1016/j.puhe.2024.05.029
24. O'Connor S, Liu H. Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities. *AI Soc.* (2024) 39:2045–57. doi: 10.1007/s00146-023-01675-4
25. Dangi RR, Sharma A, Vageriya V. Transforming healthcare in low-resource settings with artificial intelligence: recent developments and outcomes. *Public Health Nurs.* (2025) 42:1017–30. doi: 10.1111/phn.13500
26. Cary MP, Bessias S, McCall J, Pencina MJ, Grady SD, Lytle K, et al. Empowering nurses to champion Health equity & BE FAIR: Bias elimination for fair and responsible AI in healthcare. *J Nurs Scholarsh.* (2025) 57:130–9. doi: 10.1111/jnu.13007
27. Dua M, Singh JP, Shehu A. Health equity in AI development and policy: an AI-enabled study of international, national and intra-national AI infrastructures. *Proc AAAI Symp Ser.* (2024) 4:275–83. doi: 10.1609/aaais.v4i1.31802
28. Jadhav N, Marathe S, Yakundi D, Patil H. Participatory audit and planning of flexible funds under national health mission in Maharashtra, India. In: *Oral Presentations.* BMJ Publishing Group (2016). p. A17–8. Available from: <https://gh.bmj.com/lookup/doi/10.1136/bmjgh-2016-EPHAbstracts.22> (Accessed June 22, 2025).
29. Qin H, Kong J, Ding W, Ahluwalia R, Morr CE, Engin Z, et al. Towards trustworthy artificial intelligence for equitable global health. *arXiv.* (2023) Available from: <https://arxiv.org/abs/2309.05088> (Accessed June 22, 2025).
30. Yogarajan V, Dobbie G, Leitch S, Keegan TT, Bensemann J, Witbrock M, et al. Data and model bias in artificial intelligence for healthcare applications in New Zealand. *Front Comput Sci.* (2022) 4:1070493. doi: 10.3389/fcomp.2022.1070493
31. Kim JY, Hasan A, Kellogg K, Ratliff W, Murray S, Suresh H, et al. *Development and Preliminary Testing of Health Equity Across the AI Lifecycle (HEAAL): A Framework for Healthcare Delivery Organizations to Mitigate the Risk of AI Solutions Worsening*

Health Inequities (2023). Available from: <http://medrxiv.org/lookup/doi/10.1101/2023.10.16.23297076> (Accessed June 8, 2025).

32. Nadarzynski T, Knights N, Husbands D, Graham CA, Llewellyn CD, Buchanan T, et al. Achieving health equity through conversational AI: A roadmap for design and implementation of inclusive chatbots in healthcare. *PLOS Digit Health*. (2024) 3:e0000492. doi: 10.1371/journal.pdig.0000492

33. Pasculli G, Virgolin M, Myles P, Vidovszky A, Fisher C, Biasin E, et al. Synthetic data in healthcare and drug development: definitions, regulatory frameworks, issues. *CPT Pharmacomet Syst Pharmacol*. (2025) 14:840–52. doi: 10.1002/psp4.70021

34. Tyagi N, Bhushan B. Natural Language Processing (NLP) Based Innovations for Smart Healthcare Applications in Healthcare 4.0. In: Ahad MA, Casalino G, Bhushan B, editors. *Enabling Technologies for Effective Planning and Management in Sustainable Smart Cities*. Cham: Springer International Publishing (2023). p. 123–50.

Available from: https://link.springer.com/10.1007/978-3-031-22922-0_5 (Accessed June 8, 2025).

35. Gore MN, Olawade DB. Harnessing AI for public health: India's roadmap. *Front Public Health*. (2024) 12:1417568. doi: 10.3389/fpubh.2024.1417568

36. Jhunjhunwala A. Role of Telecom Network to Manage COVID-19 in India: Aarogya Setu. *Trans Indian Natl Acad Eng*. (2020) 5:157–61. doi: 10.1007/s41403-020-00109-7

37. DeCamp M, Lindvall C. Mitigating bias in AI at the point of care. *Science*. (2023) 381:150–2. doi: 10.1126/science.adh2713

38. Alderman JE, Palmer J, Laws E, McCradden MD, Ordish J, Ghassemi M, et al. Tackling algorithmic bias and promoting transparency in health datasets: the STANDING Together Consensus Recommendations. *NEJM AI*, Vol. 2. (2025) Available from: <https://ai.nejm.org/doi/10.1056/AIp2401088> (Accessed June 8, 2025).