PLOS Digital Health

Introducing the Team Card: enhancing governance for medical AI systems in the age of complexity

--Manuscript Draft--

Manuscript Number:	PDIG-D-24-00121
Article Type:	Research Article
Full Title:	Introducing the Team Card: enhancing governance for medical AI systems in the age of complexity
Short Title:	Team Cards (TCs) for enhanced governance of medical AI systems
Corresponding Author:	donnella S. comeau, MD, PhD Massachusetts General Hospital Imaging: Massachusetts General Hospital Department of Radiology Boston, MA UNITED STATES
Order of Authors:	Lesedi Mamodise Modise, MSc
	Mahsa Alborzi Avanaki
	Saleem Ameen
	Leo A. Celi
	Victor Chen
	Ashley Cordes
	Matthew Elmore
	Amelia Fiske
	Jack Gallifant
	Megan Hayes
	Alvin Marcelo
	Joao Matos
	Luis Nakayama
	Ezinwanne Ozoani
	Benjamin C. Silverman
	Donnella S. Comeau
Keywords:	AI; ML; algorithmic bias; data bias; bias; discrimination by algorithms; reflexivity; positionality; AI assurance; medical AI; Team Card; TC; ethics; governance
Abstract:	This paper proposes the Team Card (TC) as a disclosure protocol to communicate positionality attributes of the research and development teams behind Artificial Intelligence (AI) products that are deployed in clinical settings. As AI is increasingly adopted in Clinical Decision Support (CDS) software, there is growing potential for harmful bias to be perpetuated at scale. While the accuracy, robustness, and explainability of data-driven decisions remain essential to mitigating potential harm, it is increasingly important to situate the individuals behind AI-driven CDS software (medical AI systems) in order to advance effective bias mitigation throughout the AI lifecycle. TCs are intended to establish a direct link between medical AI systems and their creators, facilitating greater accountability in view of the harms that can result from the broad adoption of deficient AI in clinical settings. This visibility serves as a tool to operationalize the regulatory requirements for accountability, transparency, and for greater diversity in the broader research ecosystem. Inspired by positionality statements in research, Model Cards for model reporting as well as research reporting protocols such as Standards for Reporting Qualitative Research and Consolidated Criteria for Reporting Qualitative Research, TCs are structured profiles that allow developers of medical AI systems to self-report on potential areas of bias, recruiting

	reflexivity as a tool for bias mitigation.
Additional Information:	
Question	**Response**
Government Employee Are you or any of the contributing authors an employee of the United States government? Manuscripts authored by one or more US Government employees are not copyrighted, but are licensed under a CC0 Public Domain Dedication, which allows unlimited distribution and reuse of the article for any lawful purpose. This is a legal requirement for US Government employees. This will be typeset if the manuscript is accepted for publication.	No - No authors are employees of the U.S. government.
Financial Disclosure Enter a financial disclosure statement that describes the sources of funding for the work included in this submission. Review the submission guidelines for detailed requirements. View published research articles from *PLOS Digital Health* for specific examples. This statement is required for submission and **will appear in the published article** if the submission is accepted. Please make sure it is accurate. **Funded studies** Enter a statement with the following details: • Initials of the authors who received each award • Grant numbers awarded to each author • The full name of each funder • URL of each funder website • Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript?	The author(s) received no specific funding for this work.

Did you receive funding for this work?	
Competing Interests On behalf of all authors, disclose any competing interests that could be perceived to bias this work. This statement will be typeset if the manuscript is accepted for publication. Review the instructions link below and PLOS Digital Health's competing interests policy to determine what information must be disclosed at submission.	none
Data Availability Provide a **Data Availability Statement** in the box below. This statement should detail where the data used in this submission can be accessed. This statement will be typeset if the manuscript is accepted for publication. Before publication, authors are required to make all data underlying their findings fully available, without restriction. Review our PLOS Data Policy page for detailed information on this policy. Instructions for writing your Data Availability statement can be accessed via the Instructions link below.	None

Introducing the Team Card: enhancing governance for medical AI systems in the age of complexity

Lesedi Modise[1+]; Mahsa Alborzi Avanaki[2]; Saleem Ameen[3,4,5]; Leo A. Celi[5,6,7]; Victor Chen[1,8]; Ashley Cordes[9]; Matthew Elmore[10]; Amelia Fiske[11]; Jack Gallifant[5,12]; Megan Hayes[13]; Alvin Marcelo[14]; Joao Matos[5,15,16]; Luis Nakayama[5,17]; Ezinwanne Ozoani[18]; Benjamin C. Silverman[1,19,20]; Donnella S. Comeau[2,19*]

[1]Center for Bioethics, Harvard Medical School, Boston, Massachusetts, United States of America
[2]Department of Radiology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, United States of America
[3]Department of Biomedical Informatics, Harvard Medical School, Harvard University, Boston, Massachusetts, United States of America
[4]Tasmanian School of Medicine, College of Health and Medicine, University of Tasmania, Hobart, Tasmania, Australia
[5]Laboratory for Computational Physiology, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America
[6]Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, United States of America
[7]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America
[8]Faculty of Medicine, The Chinese University of Hong Kong, New Territories, Hong Kong SAR
[9]Department of Environmental Studies and Department of English, University of Oregon, Eugene, Oregon, United States of America
[10]Duke Health, AI Evaluation and Governance, Duke University, Durham, North Carolina, United States of America

[11]Institute of History and Ethics in Medicine, Department of Preclinical Medicine, TUM School of Medicine and Health, Technical University of Munich

[12]Department of Critical Care, Guy's and St. Thomas' NHS Trust, London, United Kingdom

[13]Department of Environmental Studies, University of Oregon, Eugene, Oregon, United States of America

[14]Medical Informatics Unit, College of Medicine, University of the Philippines Manila, Philippines

[15]Faculty of Engineering, University of Porto, Portugal

[16]Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal

[17]Department of Ophthalmology, Sao Paulo Federal University, Sao Paulo, Brazil

[18] Machine Learning and Ethics Research Engineer, Innovation n Ethics, Dublin, Ireland

[19]Department of Human Research Affairs, Mass General Brigham, Somerville, Massachusetts, United States of America

[20]Institute for Technology in Psychiatry, McLean Hospital, Belmont, Massachusetts, United States of America

[+] Authors are in alphabetical order between the first and senior authors

[*]Corresponding author

E-mail: dscomeau@mgb.org (DC)

Abstract

This paper proposes the Team Card (TC) as a disclosure protocol to communicate positionality attributes of the research and development teams behind Artificial Intelligence (AI) products that are deployed in clinical settings. As AI is increasingly adopted in Clinical Decision Support (CDS) software, there is growing potential for harmful bias to be perpetuated at scale. While the accuracy, robustness, and explainability of data-driven decisions remain essential to mitigating potential harm, it is increasingly important to situate the individuals behind AI-driven CDS software (medical AI systems) in order to advance effective bias mitigation throughout the AI lifecycle. TCs are intended to establish a direct link between medical AI systems and their creators, facilitating greater accountability in view of the harms that can result from the broad adoption of deficient AI in clinical settings. This visibility serves as a tool to operationalize the regulatory requirements for accountability, transparency, and for greater diversity in the broader research ecosystem. Inspired by positionality statements in research, Model Cards for model reporting as well as research reporting protocols such as Standards for Reporting Qualitative Research and Consolidated Criteria for Reporting Qualitative Research, TCs are structured profiles that allow developers of medical AI systems to self-report on potential areas of bias, recruiting reflexivity as a tool for bias mitigation.

Author summary

Harmful bias in clinical algorithms has engendered discrimination across a range of demographic variables, including race and gender. Discrimination by algorithm is typically attributed to bias in the data used to train medical AI systems. However, this does not account for the researchers themselves, who may be inherently and unknowingly biased. Thus, issues in researcher positionality – specifically, failures to acknowledge and disclose one's self and consequent influences on the research and development (R&D) processes of medical AI systems, have coincided with the proliferation of hazardous AI in clinical settings. We propose the TC as a disclosure protocol to communicate

positionality attributes of the R&D teams behind medical AI systems. It is inspired by positionality statements in research, Model Cards for model reporting and standard research reporting protocols. The TC's unique contribution to the field is its focus on addressing harmful bias in the development of healthcare technology, through an ethic of reflexivity among R&D teams. It is intended to support regulatory developments towards AI assurance policy in healthcare, by (i) establishing a transparent link between medical AI systems and their creators for accountability; and (ii) mediating an ethic of reflexivity for harmful bias mitigation in medical AI system development.

Introduction

Over 40% of hospitals in the United States have integrated Clinical Decision Support (CDS) software into a broad spectrum of critical functions including diagnostics, disease management, prescription services, and alarm systems, significantly enhancing healthcare delivery and patient care [1]. Despite the challenges presented by factors such as compatibility with existing clinical infrastructure, [2] data bias [3] and ethical concerns, [4, 5, 6] CDS software functions employing Machine Learning (ML) algorithms and other forms of Artificial Intelligence (AI) currently represent a rapidly expanding use case for AI in healthcare. These AI-driven CDS software functions (medical AI systems) are often described as non-knowledge based, because they generate recommendations through statistical analysis or pattern recognition within electronic health records and other data sources, instead of relying on established medical knowledge. [1] This raises concern for the explainability of clinical recommendations provided by medical AI systems, which can sometimes function as *black-boxes*, making decisions that are not easily understood. [4, 7] The increasing deployment of medical AI systems has thus prompted action to establish a basis of trust in these tools by improving their accuracy, robustness, and explainability. [8] To address these concerns in the United States, the Biden Administration issued an Executive Order in October 2023 on the trustworthy development and safe use of AI. [9] In healthcare, the order calls for the development of an AI assurance policy that

encompasses a safety program and relevant standards. This follows earlier action by both the Food and Drug Administration (FDA) and the Department of Health and Human Services (DHHS), in response to calls for accountability regulation to oversee the growing prevalence of algorithmic decision making in patient care.

In 2019 [10] and 2022, [11] the FDA issued draft and final guidance, respectively, classifying certain CDS software functions as medical devices that fall under its regulatory oversight. This guidance specifies that high risk CDS software functions should undergo a rigorous pre-market approval process that includes the review of supporting clinical data to ensure safety and effectiveness. However, many medical AI systems do not fall within the FDA's definition of a medical device and are thus not subject to oversight. [12, 13] Additionally, among the medical AI systems that do qualify as devices, most are either perceived to present low risk or to be substantially equivalent to previously authorized devices and thus obtain clearance through abridged processes that exempt clinical review. [14, 15]

In 2022, the DHHS proposed to update existing provisions of the Affordable Care Act that prohibit discrimination in covered health programs on the basis of race, sex, color, national origin, age, or disability, to also prohibit such discrimination by clinical algorithms. [16] However, some argue that the proposed framework places excessive demands on healthcare professionals, as it requires them to effectively evaluate every algorithm implemented in their practice. It is also argued that the framework fails to account for the degree of expertise required by healthcare professionals to assess ML algorithms for bias and unlawful discrimination. [12, 13] Finally, there is also concern that the standard implied by the framework may be unachievable; ML algorithms applied in healthcare settings are typically proprietary, and are thus not freely available for review. [17] Moreover, these algorithms may also be unexplainable, as is the case with black-box algorithms. [4, 7, 18] In response to these challenges, the American Medical Association has stressed the importance of establishing crosscutting responsibilities for developers and end users of medical AI systems, in operationalizing the proposed

regulation. [19] This highlights the need for greater accountability further upstream, among those who develop medical AI systems, regarding adherence to consensus-driven AI assurance standards. [20]

Current upstream efforts to avert harmful bias in medical AI systems have focused on better transparency through data and code sharing, a practice aiming to contextualize the performance characteristics of models driven by trained ML algorithms. [21] However, a researcher's positionality coincides with biases at every stage in the development of AI – not only in the acquisition of data and in the development of ML algorithms, but also in the assessment of the resulting medical AI system's performance. [22, 23] It thus becomes imperative for the individuals developing these systems to account for biases that may stem from assumptions or perspectives linked to researcher subjectivity throughout the AI development lifecycle.

To date, the most significant efforts to situate the individuals creating medical AI systems have been limited to scientometrics, which identify salient characteristics of research authors within scientific journals. While this approach provides some insight into descriptive features like the regionality of researchers over time, [24] there is no standardized method for communicating the situated knowledge and implicit bias of the individuals behind a given medical AI system. Moreover, scientometric methods do not assist R&D teams in cultivating greater awareness and critical reflection of their own positionality in relation to the AI that they develop.

This work proposes a disclosure protocol that we refer to as the Team Card (TC). The protocol is intended to engage an ethic that narrates contributor identity and positionality as essential components of AI development and thus serve as a tool to mitigate harmful bias in medical AI systems. In addition, TCs are intended to establish a direct link between medical AI systems and their creators, facilitating greater accountability in view of the harm that can result from the broad adoption of deficient AI in clinical settings. The TC protocol thus balances responsibility for AI malpractice among AI users in clinical settings and the developers of these systems. Finally, the protocol supports the informed assessment of contributor diversity in the development of medical AI systems, thereby

promoting the inclusion of a plurality of perspectives in the development of these systems and mitigating the potential for harmful bias to be encoded. Our intention is for TCs – much like research standards for qualitative data reporting or positionality statements – to serve as windows into the identities, backgrounds, and the roles of the individuals who create the AI that is adopted as CDS software.

Positionality in the R&D landscape of medical AI

Discussions of positionality in research derive from the premise that scientific knowledge is neither value-neutral nor objective. Rather, it is socially situated and is laden with both values and intent, which serve to reinforce the dominance of established views [25] and hierarchies of power. [26] *Dominant Science*, as observed by feminist and Indigenous scholars, assesses knowledge against its own self-image. In accepting – as scientific – those knowledges that align with established forms of power and eschewing those that do not, it engenders oppressive characteristics that warrant close examination, [27, 28] for they then serve to adjudicate who can engage in knowledge production. [29]

Positionality broadly refers to an individual's worldview – shaped by gender, ethnic identity, experiences, social milieu, cultural background, and other formative influences. The term has become increasingly central to qualitative research processes, serving to identify a method for quality control [30] by locating the researcher in the scientific process. Positionality statements in research reflect on the researcher's perspective regarding the design, implementation, and analysis of the research, illuminating how the unavoidable subjectivity of knowledge can influence outcomes at every stage of the research process. This is achieved by understanding the researcher's orientation toward the work through the lens of that researcher's personal values and within the context of a broader social milieu, thereby engaging an ethic of critical reflexivity as a necessary precursor to addressing harmful social biases in scientific knowledge. [25] Current efforts to situate knowledge in research are informed by established standards for qualitative reporting, such as Standards for Reporting Qualitative Research

(SRQR) and Consolidated Criteria for Reporting Qualitative Research (COREQ), which aim to transparently convey reflexivity and details about research teams. For example, Indigenous scholars have underscored the importance of locating themselves through explicit self-identification and cultural identification with protocols of introduction. [31] Similarly, others have described how by sharing their backgrounds and life experiences, they seek to build trust between the subjectivities of the researchers and of the researched [32] in a process of relational accountability. [33]

Developing these ideas, feminist and care-oriented approaches have proposed frameworks for reflexivity in the R&D of AI, which begin from the proposition that AI is socially situated and that the data with which ML models are trained are not neutral. [34] These frameworks support research methods [35] that focus on diversity and empowerment to mitigate the potential for representational harm in the AI development lifecycle. They do so by contending that a diversity of perspectives should shape the design and development of the technologies that define society, ensuring that these technologies are attuned to intersecting oppressions such as racism, [36] sexism and classism. [34]

Nevertheless, the current research landscape in AI lacks diversity. [37] A scientometric analysis of original research in medical AI system development indicates that a concentration of relevant academic papers emanates from distinct knowledge hubs in the US and in China. [38] This centrality of authorship establishes collaboration and co-authorship dynamics that serve to benefit scientific productivity, but also result in increasingly homogenous research [39] that reflects the positionality of some research groups to the exclusion of others. The resulting research is vulnerable to potential blind spots; its efficiency and proliferation derives from homogenous research teams that tend toward convergent team processes, quickly aligning on objectives and conclusions. [40] By contrast, divergent processes that juxtapose differing values and ideas are the hallmark of culturally diverse teams. [41] We argue that these divergent processes, born from a pluralism of perspectives, are important to the R&D of trustworthy medical AI systems that are free of harmful bias.

The importance of representation in the R&D of medical AI systems

The capacity for AI to perpetuate prejudice in healthcare is increasingly recognized. Bias in the development of clinical algorithms has engendered discrimination by algorithms across a range of demographic variables including race and ethnicity, age, disability, socio-economic status, English language proficiency, the presence of obesity, and gender. [42, 43, 44] Attention to health disparities in the US has largely focused on racial and ethnic disparities in patient outcomes. [45] These disparities have rooted across all stages of the clinical value chain, and are seen to affect the lives of millions of patients across the US. [46] The issue is most acute in medical diagnostics where, despite a contemporary understanding that race is not a dependable proxy for genetic difference, old beliefs to the contrary remain embedded in medicine. This is evident in the longstanding practice of *correcting* diagnostic algorithms for race. Ubiquitous diagnostic algorithms and clinical practice guidelines that adjust outputs for race include examples in cardiology, nephrology, obstetrics, oncology, endocrinology, pulmonology, and urology, all of which continue to entrench race-based medicine. [47] Explicit racial biases in healthcare – such as the aforementioned race corrections – mingle with implicit racial biases arising from learned attitudes, to produce the increasing number of AI failures observed along color lines in clinical settings. The growing corpus of scientific literature highlighting examples of racial discrimination in medical AI systems spans diagnostic procedures, therapeutic interventions and facility management systems. [46, 47, 48, 49, 50]

It should be noted that these disparate patient outcomes are typically attributed to a lack of diversity in the data used to train medical AI systems. The issue of data quality includes situations in which the data do not reflect the true epidemiology of a demographic, due for example to historic racial bias in diagnosis. [4, 51, 52] It also includes situations where unequal access to care has been encoded as statistical forms of bias, as well as situations where data sets do not contain enough demographic diversity. This last limitation leads to the decreased efficacy, reliability and generalizability of resulting medical AI systems in racially diverse settings. However, the true scope of the issue extends beyond

data quality; the R&D process of AI itself, is susceptible to the embedded and unconscious judgments of development teams, which has contributed to the observed emergence of medical AI systems with intrinsic racial bias and other harmful biases. [48]

The matter is made more complex where there are several digital determinants of health that will affect medical AI system performance. Practical examples of these determinants in the clinical setting might include device calibration, language concordance, digital literacy, and access to digital infrastructure such as electricity and stable internet connectivity. [53] It can thus be understood that harmful bias in medical AI system design is integrally related to forms of social bias. Discriminatory patient outcomes result when existing forms of social inequality are normalized by researchers and encoded into data. These biased outcomes further disadvantage communities that have already been structurally marginalized, such that forms of social bias and statistical bias in AI interact with each other in a recursive manner.

Emerging research on the autodidactic nature of certain medical AI systems presents further considerations: Wawira Gichoya et al. have demonstrated that when ML is applied to de-identified medical images such as radiographs, CT scans, and mammograms, it can predict a subject's self-identified racial identity with an accuracy of 80-99% across these imaging modalities.[54] The capability is readily acquired by standard ML models that are trained with diverse data sets and accuracy persists, even when controlling for potential racial proxies like body-mass index, disease distribution and breast density. These medical AI systems can accurately predict race from medical images that are corrupted, noisy, or cropped. Moreover, they base racial inferences on data that lies beyond the standard medical variables utilized by radiologists and other health care professionals. Thus, radiologists might be unable to recognize the variables used by ML models for racial predictions, which could frustrate efforts to monitor and control this behavior when it is undesirable. Additionally, since many models are built using de-identified data sets, the inherent bias of these models may not be readily identifiable. Set against a deep legacy of racism in US healthcare system, [55] the latent ability

of medical AI systems to make unprompted racial inferences is of serious concern. Furthermore, the same phenomenon has been observed in gender inferences, where an ML system reading de-identified retinal images was found to accurately identify a patient's self-reported sex, even though ophthalmologists could not. [6, 56]

These findings highlight the immense scope for medical AI systems to deepen disparate patient outcomes along racial and other social divides, particularly where ML models can make undetected inferences from a vast mosaic of potentially biased training data. It is thus crucial to establish a market paradigm that incentivizes the development of equity-focused medical AI systems; a paradigm that holds R&D teams accountable for the differential impact of their AI products. [57] Only such an approach will ensure that vulnerable patient populations are not further marginalized by rapid improvements in health innovation. [58]

TCs support diversity in the R&D of medical AI systems

The realization of trustworthy medical AI systems will require appropriate oversight, transparency, and data equity in the R&D of these systems. Indeed, a primary issue facing the upstream segment of the AI ecosystem is the matter of data representativeness, as the preponderance of training data currently originates from the United States and China. A 2022 analysis of PubMed publications reveals that c.40% of the datasets referenced in medical AI literature are sourced from the United States and c.14% from China. [59] A further concern is the issue of dominant groups in the R&D of AI; interwoven social and professional networks are prevalent in the medical research community, establishing niche dominance. [60] As might be expected, there is a distinct lack of diversity within these niches. Researchers with higher importance and centrality within their respective niches are seen to be less likely to be female or to come from low-and-middle-income countries (LMICs). [61]

The resulting centralization of knowledge production in high-income countries has created a power imbalance, where these regions have disproportionate influence on global standards and policies related to AI products. Consequently, potentially hazardous AI proliferates the downstream segment of the ecosystem, against which users have limited recourse. The emerging disconnect between AI developers and the contexts in which their products are deployed means that there may be no process of accountability by which to address situations in which AI products cause harm – particularly when products from high-income countries are exported to other regions.

Regulation, as exemplified by initiatives such as the EU AI Act, is needed to enforce accountability in the development of AI products, particularly for those systems with high-impact applications such as medical AI systems. International collaboration among governance structures is vital to intensify scrutiny on the data sets used to train the ML models powering medical AI systems, as these systems are adopted in diverse geographic and social contexts. [57] A globally collaborative approach fosters diversity and inclusion in medical AI system development and presents the opportunity to build capacity in LMICs, incentivizing data exchange and open science practices among research groups.

Although the AI landscape currently lacks coordinated regulation, [62] the ecosystem has sought to adopt various initiatives that support a standard of trustworthiness for AI. Successive waves of self-regulation have seen the adoption of Datasheets for Datasets, [63] which improve communication between dataset creators and users. The subsequent adoption of Model Cards [21] was a further step to standardize the disclosure of key performance characteristics for trained ML models. These cards detail operating parameters, such as intended use cases and performance evaluation criteria, which helps to reduce the deployment of trained ML models in scenarios for which they are not adequately designed. Algorithm Assurance [64] through model audits has also been proposed as an IT risk management protocol, and the growing number of audits signals a commitment to the development of equity focused AI. [22] However, more is required to realize a standard of AI trustworthiness.

The concept of the TC, as proposed in this paper, aligns closely with the goals identified by relevant regulatory frameworks. By narrating the situated knowledge of AI development teams, as well as their composition, expertise, and the ethical considerations provisioned, the TC disclosure protocol is intended to make visible the human element behind medical AI systems. This visibility serves as a tool to operationalize the regulatory requirements for transparency and accountability, ensuring that AI tools are not only technically proficient but also socially responsible and attuned to the needs of diverse contexts.

Core attributes of the TC

TCs are intended to provide relevant information on the authorship of medical AI systems. This includes information regarding the contributors' respective roles in developing the project, the expertise that they bring to the project, any institutional or industrial affiliations that they might hold, affiliations or special relationships pertaining to funding, and other information that speaks to the positionality of contributors. While TCs are not intended to be prescriptive, they should contain core information that enables stakeholders to gain familiarity with the team that has developed a given medical AI system.

We note that guidelines established by the SRQR and COREQ have been criticized for being overly rigid and for adopting a focus on checklist detail, rather than retaining focus on the holistic outcomes of relevant disclosures. [65] This protocol considers the shortcomings of comparable frameworks and instead invites teams to nominate the most relevant elements for disclosure in each situation. We hope that this flexibility will encourage all stakeholders to make these guidelines their own, adopting TCs as an avenue for authentic self-reflection, thereby mitigating the potential for harmful bias in the development of AI. Written statements, diagrams, illustrations, audio-visual and multi-modal content are all viable media for the expression of positionality in this disclosure protocol.

Elements of a TC have been proposed below. We maintain that teams, and individuals within teams, should retain control of their own level of disclosure, such that if there are aspects of their positionality that they prefer not to disclose, they should not be compelled to do so. Ultimately, the TC is not intended to serve as a regulatory device but rather as a transparency medium that is furnished in good faith. Thus, while we stress that teams are free to add or omit information as they find relevant, we equally stress that a culture of transparency is required to foster greater accountability, responsible and inclusive behavior, and ultimately the development of trustworthy AI.

Like the teams they describe, TCs are not intended to remain static over time. Our expectation is that they will be updated periodically to remain accurate in how they reflect the composition and structure of teams. We also recommend that prior contributors, who may not currently be regarded as active team members, be appropriately acknowledged. This will facilitate the fair recognition of researcher contributions over time, further supporting a culture of inclusion.

Finally, while we believe that TC disclosures can, and should, comprise information that would be made available (or would be discernible) in the course of typical R&D endeavor, we recognize that the disclosure of potentially sensitive personal information in a public document confers potential risks, such as privacy invasion. We thus encourage teams to consider these factors carefully when compiling TC disclosures.

Table 1: Core attributes of the TC

POSSIBLE DISCLOSURE PARAMETER	GOVERNANCE STANDARD	GOVERNANCE OBJECTIVES AND ACTION STEPS
Discussion of positionality	Transparency	● Provide summative reflections on the composition of the team, including perceived strengths, weaknesses, and anticipated implications, particularly with respect to harmful bias mitigation and inclusivity in AI system design
Role	Ethical oversight	● Include team functions that retain a focus on ethical considerations
	End-user advocacy	● Include perspectives that advocate for the end-user (e.g. patient advocates) so that resulting AI systems adequately meet the needs presented by the contexts in which they are deployed
	Regulatory compliance	● Maintain compliance with regulatory standards by including dedicated or consulted compliance functions
Institutional affiliation	Multi-disciplinary collaboration	● Consider cross-disciplinary collaborations to include varied perspectives and expertise
	Regulatory body engagement	● Maintain a dialogue with relevant regulatory bodies to enhance compliance functions
	Ethical review	● Design AI system development process that anticipates the formal review and approval of an appropriately constituted committee
Geographic location	Contextual testing and validation	● Test and validate AI tools in the geographic contexts in which they will be deployed
Race & ethnicity	Bias mitigation and inclusivity in design	*Inclusion is dynamic in that it recognizes dimensions of diversity that emerge from identities, including the intersection of socioeconomic status, ethnicity, cultural background, gender, ability, and sexual orientation but goes beyond to ensure belonging, respect, and success* ● Establish appropriate representation in teams to facilitate the identification of potential harmful biases ● Pursue inclusivity in AI system design by attuning to relevant demographic and sociocultural sensitivities ● Staff teams to encourage divergent perspectives, in addition to meeting the required technical skills
Gender identity & sexual orientation		
Age & other		

A presentation of our TC

We present the TC compiled by the authors of this manuscript to convey the positionality attributes resulting from the composition of our team. This illustration is a possible interpretation of the protocol, noting that other interpretations, through text, alternative visual, or mixed media representation would be viable.

Discussion of positionality

This area of research was prompted by the observations of physician-scientists on the team, who noted structural iniquities in healthcare delivery perpetuated by legacy systems with harmful bias. These observations raised concern for the potential deepening of disparate patient outcomes along social fault lines, with the layering of AI on existing healthcare infrastructure.

We note a diversity of competencies, which span clinical practice, AI R&D, social studies, bioethics, as well as healthcare services and technology investment as the core technical competencies of our team. We perceive this breadth of expertise and the multiperspectivity that it confers, as a key strength of the authors in proposing a protocol to support the realization of equitable AI in medicine. Additionally, this technical competency is balanced by a narrative that includes perspectives from a diversity of cultural backgrounds and ages and includes appropriate gender representation. Self-reported ethnicities and cultural identities of our team members include: African American, Arabian, East Asian, European, Igbo, Jewish American, KōKwel, Persian, Southeast Asian, Tswana, White American and White British. Our team is diverse in age, with members ranging from their 20s to their 50s. It includes individuals who identify as men and individuals who identify as women. Additionally, our team includes members of the LGBTQIA+ community.

The varied positionalities among members of the team include perspectives that attune to the lived experience of intersecting oppressions, and so offer an appreciation of the complex, and perhaps subtle, manner in which individuals are confronted with discrimination in healthcare. Additionally, the

authors bring both firsthand experience with the implementation of medical AI systems in clinical settings, as well as experience across the development lifecycle of these systems, to conversations about the need for greater accountability and diversity in AI R&D. As a key limitation, we note that the collective perspective of the team may be skewed to that of a socioeconomic group with high levels of education attainment, professional mobility and earning potential. Thus, the narrative presented in this manuscript could potentially be improved by the inclusion of other socioeconomic perspectives, with particular regard to perceptions towards AI and the medical establishment, as we believe that the authoring team may exhibit positive bias towards these points.

Roles and expertise

Authorship was engaged collaboratively by a team of 16 individuals, 12 of whom identify as clinicians, machine learning engineers, data scientists or an intersection of these descriptors. This allowed for self-reporting and advocacy from perspectives across the medical AI value chain; from the upstream perspective of AI product developers to the downstream perspective of healthcare professionals implementing medical AI systems. Remaining authors contributed expertise as anthropologists, social science researchers, bioethicists and healthcare technology investors, each providing an unique lens on the social determinants of equitable AI. Additionally, 2 members of the team contributed ethical review and oversight considerations informed by their engagements with Institutional Review Boards.

Institutional affiliations

Institutional affiliations of the authoring team include: Duke University, Harvard Medical School, Mass General Brigham, Massachusetts Institute of Technology, University of Oregon, Technical University of Munich School of Medicine and Health, The Beth Israel Deaconess Medical Center, University of the Philippines Manila. No funding affiliation is relevant to this manuscript.

Further disclosures

A visual representation of our TC can be found at the following link: Team Cards

The main figure representing relationships among team members utilizes a forced-directed graph to display the multifaceted relationships and backgrounds of the team, in a clear, visually discernible way. The visualization was rendered using React Flow (a node-based graph visualization library), and the layout algorithm was implemented with the support of D3.js (a toolkit for data-driven documents). The graph consists of two fundamental components: nodes that represent team members, and edges that represent the relationships between team members. Prior to rendering, nodes were labeled according to their group affiliation, so that clusters of closely connected individuals can be displayed within close proximity to one another. After labeling, a structured circular layout for each cluster was implemented, such that each node within a cluster is evenly spaced to form a radial distribution of equal segments that are proportional to the number of nodes within the given cluster.

This initial placement served as a starting point for the dynamic force simulation, which is implemented to iteratively refine node positions to an equilibrium state that represents the natural interconnections among team members. Using D3.js for the simulation, several forces are applied to each node and edge: (1) a repulsive charge force is used to ensure that nodes do not cluster too closely; (2) a spring-like link force is used to maintain the optimal distance between connected nodes, so that nodes too close to each other are separated, and nodes too far apart are pulled closer together; (3) a collision force is used to prevent overlap of nodes; (4) a boundary force guides the nodes to the center of the canvas to avoid edge crowding; (5) a custom grouping force is implemented to emphasize the grouping of team members. This force, based on group affiliation, pushes members of less connected groups towards the outer edges of the layout. Through the simulation, the position of each node is continuously and iteratively adjusted based on the interplay of these forces until a state of equilibrium is obtained.

In the accompanying visual representation of our TC, we include a panel summary of descriptors that inform our team's positionality, including the team's diversity in ethnicity, age, sexual orientation and gender, expertise, languages spoken, national identity, and institutional affiliation. In addition, our team's relational network is visually represented as a series of spatial arrangements, communicating the overall team structure and functional clusters. For instance, Dr. Leo Anthony Celi is centrally positioned to indicate his role in the project, with all other team members linked to him as perceived by the team. Additionally, team members are interconnected with color-coded bands that denote functional clusters within the group.

We note, as potential areas for improvement to the illustration provided, that the inclusion of photographs with corresponding pronouns and ethnic identifiers could improve the disclosure by providing a visual representation of gender and ethnic diversity. This level of disclosure has not been pursued by the team in recognition of the sensitive nature of these attributes, which have been / may be used to discriminate against members of the team. We have aggregated these descriptors for the purposes of this illustration but could provide more granular disclosure to oversight bodies as needed. Further improvements might include visuals that clarify the ethical oversight and policy frameworks governing the team's work as relevant structures develop. As a final limitation, we note that while efforts have been taken to balance dual commitments to individual privacy as well as to transparency and accountability, individual gender and ethnic attributes that are not directly interpretable from the illustration prepared, could be vulnerable to reidentification in the context of subsequent TC disclosures.

We hope that this discussion of the teams positionality and the accompanying relational network presented, promotes transparency and trust among stakeholders, by making aspects of the team's implicit perspective towards AI equity in medicine more transparent.

Considerations and limitations for operationalizing the TC and for ensuring diversity in the R&D of medical AI systems at scale

The proposed TC disclosure is unique to the AI R&D landscape in its focus on driving transparency through an ethic of reflexivity among researchers. The protocol has the potential to meaningfully improve transparency and accountability in the AI ecosystem by putting the people behind the development of medical AI systems in direct, transparent, and reflective conversation with the performance of those systems in the field.

We believe that a primary strength of the TC is that it can help to align the medical AI development process with ethical principles that underpin the regulation of research involving human subjects. [66] Moreover, by enhancing transparency regarding the potential biases embedded in AI products, the TC protocol allows stakeholders in the AI ecosystem to be more informed and discerning in their interactions with these products. From an operational perspective, this could serve to inculcate a pro-diversity ethos in the development of medical AI systems, allowing for greater mitigation of representational and allocative harm. [34]

The positionality of the individuals involved in the development of medical AI systems, including their funding and any potential biases or conflicts that may emerge from these relational parameters, must be reported transparently [60] if proposed assurance policies are to succeed in mitigating bias for non-discrimination. [12, 16] Thus the TC aligns with the recent executive order [9] in addressing a gap in the current regulation to operationalize greater accountability in the development of medical AI systems.

Finally, the TC is a practical means of recognizing the collective effort supporting the development of medical AI systems, promoting collaborator recognition in an inclusive and equitable manner. To foster a culture that values diversity in AI (understood broadly across demographic, institutional, geopolitical, and other relevant disclosure parameters), we hope that an intrinsic incentive will emerge to recognize the contributions of team members who might otherwise be overlooked. This is

of great importance in the current paradigm, which rewards niche dominance and individual seniority. [60]

To be sure, there are limitations pertaining to the implementation of the TC at scale. First, the protocol is not intended to have a prescribed format, but rather is intended to be a transparent and informative expression of reflexivity in the AI R&D process. While such flexibility has its advantages, it challenges the establishment of clear guidelines for TC reporting and related monitoring activity. As a mitigant, we observe that positionality statements in qualitative research serve as a good point of reference in understanding the objectives and, hence, the necessary elements of a TC disclosure. From this basis, teams should interpret the cards as a conduit for self-reflection, acknowledgment of potential bias, and evaluation of the impact that any bias may have on the project under consideration. While the most appropriate format of this expression should be left to teams' discretion, the content of the disclosure should respect a common understanding of the principles above.

Another key limitation of the TC is that it may be difficult to operationalize, in view of the voluntary and self-reported nature of the protocol. Additionally, asynchronous AI governance processes could prompt marked variation in the quality and the orientation of TC disclosures across regions. Here we believe that auxiliary initiatives to promote transparent, accurate, and useful reporting may include the implementation of validation protocols that are designed to identify and flag critical omissions by teams, in order to maintain the integrity and the usefulness of the proposed protocol. One such validation protocol could be an AI system development process that anticipates the formal review and approval of an appropriately constituted ethics committee or oversight body. Interactions with these overseeing bodies could then be supported by TC disclosures.

Further, the TC may raise data privacy concerns and prompt pushback from team members regarding the public reporting of their personal information, with respect to unintended consequences of this disclosure, such as potential tokenization and stigmatization. In response to this limitation, we maintain that while the TC is intended to promote greater transparency, it is nevertheless bound by

standard privacy protection regulations such as GDPR, [67] which should allow teams to retain autonomy in their personal disclosures. Additionally, we contend that much of the information required for the TC is already disclosed by teams, through various professional and social media. However this information is scattered across various platforms (e.g.: websites, publications, presentations) and media (e.g.: video, print, augmented reality) to little social benefit, whereas a generalizable protocol that centralizes all relevant disclosures by a given researcher, would greatly enhance efforts towards mitigating harmful bias in medical AI systems. The TC is thus an opportunity to consolidate existing metadata and to systematize the collection of further information, with the potential to streamline the acquisition and evaluation of researcher information by integrating the TC with standard identifiers such as ORCID iDs (ORCID is a non-profit organization that provides researchers with a unique digital identifier).

The current scale of institutional R&D of medical AI systems presents a practical impediment to the implementation of TCs; certain disclosure parameters may become cumbersome to track where projects involve multiple teams collaborating across various stages of the AI development lifecycle. Here we might suggest multiple TCs for a single medical AI system, prepared in condensed and modular fashion, for discrete stages in the AI lifecycle and paired with a clear record of the AI system's progression, linking and narrating all relevant TC disclosures.

Finally, a focus on inclusivity has the potential to crowd out other priorities, such as in instances where skill-based vetting of team members may be revised to satisfy diversity and inclusion targets that are encouraged by the TC disclosure protocol as a means of harmful bias mitigation. Existing social inequalities that perpetuate demographic biases in the labor market for AI R&D will indeed limit the extent to which diversity requirements can be operationalized. This reality will thus need to be weighed against increasingly pressing constraints imposed by an evolving regulatory landscape, towards AI assurance and the elimination of discrimination by algorithms.

Conclusion

Issues in the positionality of R&D teams – specifically, failures to acknowledge and disclose their selves and consequent influences on the R&D processes [68] of medical AI systems, have coincided with harmful bias that perpetuates discrimination by algorithms in healthcare. In addition to coordinated regulatory action to improve the safety and trustworthiness of medical AI systems, we believe that TCs could be a useful tool to mitigate harmful bias in relevant development processes.

The TC protocol is intended to build trust through transparency in the development of medical AI systems, as we believe that a standard of trustworthiness is crucial to the practical success of these systems. While the TC is a voluntary disclosure, rather than a regulatory device, its implementation requires a culture of transparency that will be supported by increasing regulatory action towards AI assurance and greater accountability in the ecosystem. The key strengths of the TC protocol in supporting regulatory developments will thus be in its establishing a direct link between medical AI systems and their creators and in mediating an ethic of reflexivity for harmful bias mitigation in medical AI system development.

491 **References**

[1] Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of Clinical Decision Support systems: benefits, risks, and strategies for success. NPJ Digit Med. 2020 Feb 6;3(1):1–10. doi: 10.1038/s41746-020-0221-y.

[2] Zhan A. Towards AI-assisted healthcare: system design and deployment for Machine Learning based Clinical Decision Support [PhD thesis] [Internet]. Baltimore (USA): The Johns Hopkins University; 2018 [cited 2024 Mar 24]. 135p. Available from: http://search.proquest.com.ezp-prod1.hul.harvard.edu/dissertations-theses/towards-ai-assisted-healthcare-system-design/docview/2212118810/se-2?accountid=11311

[3] House Committee on Ways & Means. Fact versus fiction: Clinical Decision Support tools and the (mis)use of race [Internet]. Congressional Publications. 2021[cited 2024 Mar24]; Available from: https://perma.cc/25NR-YBY7

[4] Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. PLoS Medicine. 2018 Nov 6;15(11):e1002689. doi: 10.1371/journal.pmed.1002689.

[5] Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible Machine Learning for health care. Nat Med. 2019 Aug 19;25(9):1337–40. doi: 10.1038/s41591-019-0548-6.

[6] Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in Machine Learning to advance health equity. Ann Intern Med. 2018 Dec 4;169(12):866–72. doi: 10.7326/M18-1990.

[7] Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the era of Artificial Intelligence. JAMA. 2018 Dec 4;320(21):2199–200. doi: 10.1001/jama.2018.17163.

[8] Liu H, Wang Y, Fan W, Liu X, Li Y, Jain S, et al. Trustworthy AI: a computational perspective. ACM Trans Intell Syst Technol. 2022 Nov 9;14(1):4:1-59. doi: 10.1145/3546872.

[9] House TW. The White House. FACT SHEET: President Biden issues Executive Order on safe, secure, and trustworthy Artificial Intelligence. 2023 [cited 2024 Mar 24]; Available from: https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

[10] Health C for D and R. Clinical Decision Support software: draft guidance for Industry and Food and Drug Administration staff [Internet]. FDA. 2019 [cited 2024 Mar 25]; Available from: https://www.regulations.gov/document/FDA-2017-D-6569-0041

[11] Health C for D and R. Clinical Decision Support software [Internet]. FDA. 2022 [cited 2024 Mar 25]; Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software

[12] Shachar C, Gerke S. Prevention of bias and discrimination in clinical practice algorithms. JAMA. 2023 Jan 5;329(4):283–4. doi: /10.1001/jama.2022.23867.

[13] Goodman KE, Morgan DJ, Hoffmann DE. Clinical algorithms, antidiscrimination laws, and medical device regulation. JAMA. 2023 Jan 5;329(4):285–6. doi: 10.1001/jama.2022.23870.

[14] Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med. 2020 Sep 11;3(1):118. doi: 10.1038/s41746-020-00324-0.

[15] Lee JT, Moffett AT, Maliha G, Faraji Z, Kanter GP, Weissman GE. Analysis of devices authorized by the FDA for Clinical Decision Support in critical care. JAMA Internal Medicine. 2023 Oct 9;183(12):1399–401. doi: 10.1001/jamainternmed.2023.5002.

[16] Federal Register [Internet]. Nondiscrimination in health programs and activities. 2022 [cited 2024 Mar 24]; Available from: https://www.federalregister.gov/documents/2022/08/04/2022-16217/nondiscrimination-in-health-programs-and-activities

[17] Giovanola B, Tiribelli S. Beyond bias and discrimination: redefining the AI ethics principle of fairness in healthcare Machine Learning algorithms. AI & Soc. 2022 May 21;38(2):549–63. doi: 10.1007/s00146-022-01455-6.

[18] Petch J, Di S, Nelson W. Opening the black box: the promise and limitations of explainable Machine Learning in cardiology. Canadian Journal of Cardiology. 2021 Sep 14;38(2):204–13. doi:10.1016/j.cjca.2021.09.004.

[19] Crigger E, Reinbold K, Hanson C, Kao A, Blake K, Irons M. Trustworthy augmented intelligence in health care. J Med Syst. 2022 Jan 12;46(2):12. doi: 10.1007/s10916-021-01790-z.

[20] Mello MM, Shah NH, Char DS. President Biden's Executive Order on Artificial Intelligence—implications for health care organizations. JAMA. 2024 Jan 2;331(1):17–8. doi.org/10.1001/jama.2023.25051.

[21] Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model cards for model reporting. In: Association for Computing Machinery, editor. FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency; 2019 Jan 29; New York, USA. New York: Assoc Computing Machinery; 2019. 220–229. doi: 10.1145/3287560.3287596.

[22] Charpignon ML, Byers J, Cabral S, Celi LA, Fernandes C, Gallifant J, et al. Critical bias in critical care devices. Critical Care Clinics. 2023 Mar 7;39(4):795–813. doi: 10.1016/j.ccc.2023.02.005.

[23] Nazer LH, Zatarah R, Waldrip S, Ke JXC, Moukheiber M, Khanna AK, et al. Bias in Artificial Intelligence algorithms and recommendations for mitigation. PLOS Digital Health. 2023 Jun 22;2(6):e0000278. doi: 10.1371/journal.pdig.0000278.

[24] Gallifant J, Zhang J, Whebell S, Quion J, Escobar B, Gichoya J, et al. A new tool for evaluating health equity in academic journals; the diversity factor. PLOS Global Public Health. 2023 Aug 14;3(8):e0002252. doi: 10.1371/journal.pgph.0002252.

[25] Harding S. Whose science? Whose knowledge?: thinking from women's lives. Ithaca (USA): Cornell University Press; 2016. 336p. https://muse.jhu.edu/pub/255/monograph/book/48914

[26] Tuhiwai Smith L. Decolonizing methodologies : research and indigenous peoples. London (UK): Zed Books; 2021. 291p. doi: 10.5040/9781350225282.

[27] Haraway D. Situated knowledges: the science question in feminism and the privilege of partial perspective. Feminist Studies. 1988;14(3):575–99. doi: 10.2307/3178066.

[28] Adas M. Machines as the measure of men: science, technology, and ideologies of western dominance. Ithaca (USA): Cornell University Press; 2015. https://muse.jhu.edu/book/55299

[29] Chilisa, B. Indigenous research methodologies. Thousand Oaks (USA): SAGE Publications; 2020. 343p.

[30] Berger R. Now I see it, now I don't: researcher's position and reflexivity in qualitative research. Qualitative Research. 2013 Jan 3;15(2):219–34. doi: 10.1177/1468794112468475.

[31] Liboiron M. Pollution Is colonialism [Internet]. Durham (USA): Duke University Press; 2021 [cited 2024 Mar 24]. 216p. Available from: https://www.dukeupress.edu/pollution-is-colonialism

[32] Liamputtong P. Researching the vulnerable: a guide to sensitive research methods. London (UK): Sage Publications; 2007. doi: 10.4135/9781849209861.

[33] Wilson S. Research is ceremony: indigenous research methods. Halifax (Canada): Fernwood Publishing; 2008. 144p.

[34] Gray J, Witt A. A feminist data ethics of care for machine learning: The what, why, who and how. FM. 2021 Dec 6; 26(12):1. doi: 10.5210/fm.v26i12.11833.

[35] Franks M. Feminisms and cross-ideological feminist social research: standpoint, situatedness and positionality – developing cross-ideological feminist research [Internet]. Journal of International Women's Studies. 2002 [cited 2024 Mar 24];3(2):38–50. Available from: https://vc.bridgew.edu/cgi/viewcontent.cgi?article=1601&context=jiws

[36] Cave S, Dihal K. The Whiteness of AI. Philos Technol. 2020 Dec 1;33(4):685–703. doi: 10.1007/s13347-020-00415-6.

[37] Freire A, Porcaro L, Gómez E. Measuring diversity of Artificial Intelligence conferences. arXiv. 2021 Mar 22. doi: 10.48550/arxiv.2001.07038.

[38] Zhang J, Whebell S, Gallifant J, Budhdeo S, Mattie H, Lertvittayakumjorn P, et al. An interactive dashboard to track themes, development maturity, and global equity in clinical artificial intelligence research. The Lancet Digital Health. 2022 Apr;4(4):e212–3. doi: 10.1016/S2589-7500(22)00032-2.

[39] Newman M. Co-authorship networks and patterns of scientific collaboration. PNAS. 2004 Apr 6;101(1): 5200–5205. doi: 10.1073/pnas.0307545100.

[40] Stahl GK, Maznevski ML, Voigt A, Jonsen K. Unraveling the effects of cultural diversity in teams: a meta-analysis of research on multicultural work groups. J Int Bus Stud. 2009 Nov 26;41(4):690–709. doi: 10.1057/jibs.2009.85.

[41] Davison SC, Ekelund BZ. Effective team processes for global teams. In Lane HW, Maznevski ML, Mendenhall ME, McNett T, editors. The Blackwell handbook of global management. Oxford (UK): Blackwell Publishing ; 2017. p.227–49.

[42] Harvard Law School Center for Health Law Policy Innovation responds to RIN 0945-AA17, nondiscrimination in health and health education programs or activities [Internet]. The Center for Health Law and Policy Innovation at Harvard Law School. 2022 Oct 3 [cited 24 Mar 2024]. Available from: https://chlpi.org/wp-content/uploads/2022/10/Section-1557-Clinical-Algorithms-Comment.pdf

[43] Hoffman S. The Emerging Hazard of AI-Related Health Care Discrimination. The Hastings Center Report. 2020 Dec 14;51(1):8–9. doi: 10.1002/hast.1203.

[44] Chen IY, Szolovits P, Ghassemi M. Can AI help reduce disparities in general medical and mental health care? AMA Journal of Ethics. 2019 Feb 1;21(2):167–79. doi: 10.1001/amajethics.2019.167.

[45] Braveman, P. Health disparities and health equity: concepts and measurement. Annual Review of Public Health. 2006 Apr;27(1): 167–94. doi: 10.1146/annurev.publhealth.27.021405.102103.

[46] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019 Oct 25;366(6464):447–53. doi: 10.1126/science.aax2342.

[47] Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight — reconsidering the use of race correction in clinical algorithms. New England Journal of Medicine. 2020 Aug 27;383(9):874–82. doi: 10.1056/NEJMms2004740.

[48] Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: a call for open science. Patterns. 2021 Oct 8;2(10). doi. 10.1016/j.patter.2021.100347.

[49] Adamson AS, Smith A. Machine Learning and health care disparities in dermatology. JAMA Dermatology. 2018 Nov 1;154(11):1247–8. doi: org/10.1001/jamadermatol.2018.2348.

[50] Sarkar R, Martin C, Mattie H, Gichoya JW, Stone DJ, Celi LA. Performance of intensive care unit severity scoring systems across different ethnicities in the USA: a retrospective observational study. The Lancet Digital Health. 2021 Apr 1;3(4):e241–9. doi: 10.1016/S2589-7500(21)00022-4.

[51] Neighbors HW, Jackson JS, Campbell L, Williams D. The influence of racial factors on psychiatric diagnosis: a review and suggestions for research. Community Ment Health J. 1989 Dec;25(4):301–11. doi: 10.1007/BF00755677.

[52] Meints SM, Cortes A, Morais CA, Edwards RR. Racial and ethnic differences in the experience and treatment of noncancer pain. Pain Management. 2019 May 29;9(3):317–34. doi: 10.2217/pmt-2018-0030.

[53] Gallifant J, Celi LA, Pierce RL. Digital determinants of health: opportunities and risks amidst health inequities. Nat Rev Nephrol. 2023 Aug 25;19(12):749–50. doi: 10.1038/s41581-023-00763-4.

[54] Gichoya JW, Banerjee I, Bhimireddy AR, Burns JL, Celi LA, Chen LC, et al. AI recognition of patient race in medical imaging: a modelling study. The Lancet Digital Health. 2022 May 11;4(6):e406–14. doi: 10.1016/S2589-7500(22)00063-2.

[55] Matthew DB. Just medicine: a cure for racial inequality in American health care [Internet]. New York: NYU Press; 2015 [cited 2024 Mar 25]: Available from: https://muse.jhu.edu/pub/193/monograph/book/76222

[56] Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Predicting cardiovascular risk factors from retinal fundus photographs using deep learning. Nat Biomed Eng. 2018 Feb 19;1(2):158–164. doi: 10.1038/s41551-018-0195-0.

[57] Gallifant J, Nakayama LF, Gichoya JW, Pierce R, Celi LA. Equity should be fundamental to the emergence of innovation. PLOS Digital Health. 2023 Apr 10;2(4):e0000224. doi: 10.1371/journal.pdig.0000224.

[58] Gallifant J, Kistler EA, Nakayama LF, Zera C, Kripalani S, Ntatin A, et al. Disparity dashboards: an evaluation of the literature and framework for health equity improvement. The Lancet Digital Health. 2023 Nov;5(11):e831–9. doi: 10.1016/S2589-7500(23)00150-4.

[59] Celi LA, Cellini J, Charpignon ML, Dee EC, Dernoncourt F, Eber R, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. PLOS Digital Health. 2022 Mar 31;1(3):e0000022. doi: 10.1371/journal.pdig.0000022.

[60] Matos J, Nakayama L, Charpignon ML, Gallifant J, Kashkooli M, Carli F, et al. The medical knowledge oligarchies. medRxiv. 2023 Jun 5. doi: 10.1101/2023.06.02.23290881.

[61] Shambe I, Thomas K, Bradley J, Marchant T, Weiss HA, Webb EL. Bibliometric analysis of authorship patterns in publications from a research group at the London School of Hygiene & Tropical Medicine, 2016-2020. BMJ Glob Health. 2023 Feb;8(2):e011053. doi: 10.1136/bmjgh-2022-011053.

[62] Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med. 2023 Jul 6;6(1):1–6. doi: 10.1038/s41746-023-00873-0.

[63] Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Iii HD, et al. Datasheets for datasets. Commun ACM. 2021 Dec;64(12):86–92. doi: 10.1145/3458723.

[64] Boer A, De Beer L, Van Praat F. Algorithm assurance: auditing applications of artificial intelligence. In: Berghout E, Fijneman R, Hendriks L, De Boer M, Butijn BJ, editors. Advanced digital auditing [Internet]. Springer International Publishing; 2023 [cited 2024 Mar 25]:149–83. Available from: https://link.springer.com/10.1007/978-3-031-11089-4_7

[65] Dossett LA, Kaji AH, Cochran A. SRQR and COREQ reporting guidelines for qualitative studies. JAMA Surgery. 2021 Sep 1;156(9):875–6. doi: 10.1001/jamasurg.2021.0525.

[66] Friesen P, Douglas-Jones R, Marks M, Pierce R, Fletcher K, Mishra A, et al. Governing AI-driven health research: are IRBs up to the task? Ethics & Human Research. 2021 Mar;43(2):35–42. doi: 10.1002/eahr.500085.

[67] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation) [Internet].Brussels: Official Journal of the European Union; 2016 Apr 27 [cited 2024 Mar 24]. Available from: http://data.europa.eu/eli/reg/2016/679/oj

[68] Holmes AG. Researcher positionality - a consideration of its influence and place in qualitative research - a new researcher guide. Shanlax Int. Journal of Education. 2020 Sep 1;8(4):1. doi: 10.34293/education.v8i4.3232.