Building Fair and Trustworthy Biomedical AI: A Tool for Identifying Key Decision Points

Nicole Foti

Stanford Center for Biomedical Ethics, Stanford University
Stanford, CA, USA
Email: nfoti@stanford.edu

Janet K. Shim

Department of Social and Behavioral Sciences, University of California San Francisco
San Francisco, CA, USA

Caitlin McMahon

Division of Ethics, Columbia University
New York, NY, USA

Sandra Soo-Jin Lee

Division of Ethics, Columbia University
New York, NY, USA

Recent advancements in artificial intelligence (AI) have transformed biomedicine, offering tools for improved diagnostics, drug discovery, and patient care. Yet these innovations raise pressing ethical concerns, including bias, inequitable outcomes, and privacy risks, which highlight the need for deliberate attention to fairness, trust, and trustworthiness in AI development. In this paper, we argue that ethical responsibility should be embedded at both institutional and individual levels, and that multi-stakeholder engagement, especially with underrepresented groups, is essential to ensure AI tools meet diverse needs. Building on a framework originally developed for precision medicine research, we present an adapted decision-mapping tool—the Trustworthy AI Decision Map—that can anchor and structure dialogue about the ethical implications of specific AI tools. The map identifies key decision points across the AI life cycle that impact fairness and trustworthiness and facilitates dialogue among stakeholders. In making these decisions visible, the map seeks to enable teams to anticipate downstream consequences, integrate multiple perspectives, and support institutional accountability. We illustrate its potential through a case involving the deployment of AI in rural healthcare settings. Moving forward, we suggest that empirical testing with stakeholders is needed to validate and refine the map's utility in biomedical AI contexts to promote fair and trustworthy AI practices.

Keywords: Biomedical AI, Fairness, Trust, Trustworthiness, Stakeholder engagement

1. Introduction

Artificial intelligence is transforming biomedicine. From improving diagnostic accuracy to accelerating drug discovery, AI offers powerful tools for advancing research and clinical care. Yet this potential is entangled with a range of ethical concerns, including biased training data that can exacerbate health disparities, threats to privacy, and the risk of unintended harms when algorithms are used in ways not originally intended. These concerns raise critical questions about responsibility and process: At what points in the development cycle are ethical issues identified, and through what mechanisms? Who bears responsibility for addressing them? And how can interventions be structured to anticipate downstream consequences rather than reacting only once problems emerge?

For many working in biomedical AI, responsibility for ethical outcomes is either unclear or assumed to lie elsewhere. Some developers believe their products are unlikely to cause harm, while others may recognize the risks but feel ill-equipped to address them. A recent study on software startup environments found that ethical concerns are often deferred or displaced—framed as someone else's problem, or something to be handled "later," if at all.[1] Even when researchers or developers want to build more fair and trustworthy systems, they may lack the structures, support, or tools to identify where their decisions have ethical consequences.

This ambiguity reflects a deeper issue, in which the burden of ethical AI is often placed on individuals when it should be understood as a shared, institutional responsibility. Our work responds to this challenge. Building on a framework originally developed for use in precision medicine research (PMR), we offer a tool for identifying key decision points across the life cycle of biomedical AI/ML development that shape fairness and trustworthiness. By making these decision points visible, the tool can support both individual reflection and institutional accountability by facilitating dialogue about when and where to intervene.

1.1. *Fairness, trust, and trustworthiness in AI*

In recent years, there has been a surge of interest in "fair" and "trustworthy" AI.[2,3] Yet these terms are often invoked without careful definition, or treated in ways that gloss over their complexity. Even less attention is paid to the fact that fairness, trust, and trustworthiness can have very different meanings depending on social, cultural, and institutional context. What one group views as fair or trustworthy may be contested by another. For instance, a widely cited study of fairness in AI shows how applying different fairness metrics to the same case can yield opposing conclusions, with some assessing the case as fair and others as unfair.[4]

For the purpose of this paper, we follow a sociological perspective that treats trust as a *relational* process.[5] In this view, trust exists when one party believes another has incentives to act in their best interest.[6] This perspective distinguishes trust from trustworthiness. Trust refers to the belief or expectation held by one party that the other will act in their best interest, while

trustworthiness encompasses the actual practices and qualities that justify that belief. Kraft and colleagues emphasize this distinction in precision medicine, arguing that *building trust* and *being trustworthy* are two distinct actions.[7] Institutions may cultivate trust through strategic outreach or branding, but this does not guarantee trustworthiness. To be trustworthy requires substantive commitments: maintaining transparency, ensuring accountability, minimizing conflicts of interest, addressing systemic inequities, and engaging meaningfully with participants' lived experiences and histories.

Fairness is related but distinct, focusing attention on how the benefits and harms of AI are distributed and, importantly, on the broader systems of inequality in which technologies take shape. Fair AI is defined in some frameworks as tools and models that are impartial, do not introduce bias, and are applied equitably to produce equitable outcomes across populations.[8] Other frameworks go beyond these somewhat more technical dimensions to also include principles of transparency (comprised of qualities of interpretability, explainability, accountability) and inclusion (including completeness of information to detect bias and responsible collection of data, patient and family engagement) into proposed definitions of fairness in AI.[9] In this latter, more fulsome definition of fairness, there is substantial alignment with notions of trustworthiness.

Together, fairness and trustworthiness are integral in the evaluation of ethical AI. But because these concepts are situated and contested, principle-driven approaches can be limited. To ensure fair and trustworthy AI, it is essential to bring in perspectives from patients, members of the public, and especially individuals from marginalized communities.[10,11] Thus, we need tools that can help elicit, compare, and negotiate diverse stakeholder understandings of what fair and trustworthy AI requires in particular contexts, such as in clinical care.

1.1.1. *Multi-stakeholder perspectives*

Engagement with AI developers, patients, researchers, healthcare professionals, policymakers, and marginalized groups is crucial for understanding the varied and context-specific meanings of trust, fairness, and trustworthiness in the development and deployment of AI in health and biomedicine. As noted, these concepts are not universal; they are shaped by local conditions, histories, and relationships, and their interpretations can differ significantly across multiple stakeholder groups. Such differences directly influence whether, and under what circumstances, communities are willing to adopt AI technologies and how these technologies might affect them. Assumptions about shared values, potential benefits, or possible harms may overlook important concerns and priorities. Instead, these must be surfaced through meaningful engagement.

Effective engagement extends beyond one-way collection or dissemination of information to forms of collaboration in which communities and stakeholders have genuine influence and power over decisions. Community and stakeholder engagement can take many forms, from focus groups

and participatory workshops to long-term advisory boards, community-based participatory research, or co-design processes that embed stakeholders throughout every stage of a project.[12] Importantly, engagement should not be viewed as a box-checking exercise—it requires time, resources, and careful planning.[13] When done haphazardly, it can erode trust, especially if participants perceive the process as performative or if they do not see their contributions as leading to meaningful outcomes; however, strong engagement practices can often strengthen trust and trustworthiness.[14,15] Thus, our mapping tool offers one pathway and a set of procedures toward this goal. Unlike principles-based frameworks that rely on predefined ethical standards, our map is designed to facilitate dialogue among diverse stakeholders about their own definitions of fairness, trust, and trustworthiness.

1.2. *Current tools for ethical AI in biomedicine*

Several tools have been proposed to address ethical considerations in biomedical AI. Ng and colleagues adopt a lens related to ours, focusing on the AI life cycle to systematically identify ethical concerns across each stage.[16] Importantly, they emphasize the interdependence of these stages, arguing that ethical implications should be assessed holistically rather than in isolation. A team at Stanford Health Care developed an assessment to identify Fair, Useful, and Reliable AI Models (FURMs), a framework used "to routinely estimate the achievable benefits of AI model-guided workflows before deployment."[17] In addition to technical, infrastructural, and impact considerations, each FURM assessment includes a component to assess ethical considerations, including responsibility, equity, traceability, reliability, governance, nonmaleficence, and autonomy. Trotsyuk and colleagues offer yet another approach, focusing on preventing and mitigating misuse of AI in biomedicine.[18] Their framework identifies current guidance and regulations, recommends existing "off-the-shelf" mitigation strategies, and considers design-specific interventions to address potential harms.

Given that most of these tools focus on identifying specific ethical issues to guide the development of fair and trustworthy AI, in this paper, we explore the possibility of adapting a tool we created for a different domain—supporting cross-stakeholder dialogue and collaboration in precision medicine research (PMR)—to the context of AI development and deployment in biomedicine. Our Diversity Decision Map (DDM) was designed to provide a platform to elicit and compare multiple stakeholders' understandings of diversity and equity, and to examine how these values are integrated into research decisions. We believe this approach has potential utility for structuring multi-stakeholder discussions around the concepts of fairness, trust, and trustworthiness in biomedical AI, and for integrating these values into governance practices.

Below, we first describe the development of the DDM—how it was created, how it can be applied in PMR, and an illustrative example emphasizing community engagement. Building on this foundation, we introduce a new, adapted map for AI development that highlights key decision points affecting fairness, trust, and trustworthiness. Although this AI-focused map has not been reviewed

by stakeholders for feedback like the DDM, we show how it could be used in practice, using a real-world example of AI that again underscores the importance of diverse stakeholder engagement throughout the development life cycle—from problem identification and team formation through model maintenance and monitoring.

2. The Diversity Decision Map

From 2018 to 2023, our research team conducted the Ethics of Inclusion (EOI) Study,[19–22] where our goal was to determine how definitions of diversity impact precision medicine research practices such as retention, engagement, and return of results. We assessed how approaches towards diversity and inclusion are managed in tandem with other research goals, and the implications of specific trade-offs that result. As part of our effort to create data-informed guidance for PMR, we developed an empirically derived draft dialogical tool, the Diversity Decision Map (DDM), to support multi- and cross-stakeholder conversations and more collaborative, deliberate decision-making around how to achieve goals of diversity, inclusion, and equity in PMR.[23]

2.1. *Methods and development of DDM*

The methods for the development of the DDM are described fully elsewhere.[23] Briefly, we analyzed our EOI Study data to identify and create a visual display of key decision-making nodes that influenced and/or impacted goals of diversity and equity. We took inspiration from the burgeoning literature on designing biomedical research for diversity and equity[24–31] but sought to offer something distinct from the existing literature, motivated by the decision-making support needs expressed by our participants. They indicated needing ways to help anticipate the future consequences of research choices and tradeoffs for diversity, inclusion, and equity; a platform and opportunities for greater multi-stakeholder dialogue; and tools to retrospectively reflect on different study choices—paths taken and not taken—that ended up impacting the inclusivity of their research practices and diversity of participant samples in unanticipated ways. To assess the potential utility of the DDM, we conducted three engagement sessions with regulatory and ethics staff at an academic medical center, precision medicine investigators who are part of a genetic research consortium, and community partners with experience advising a clinical translational science institute. Results indicated that the DDM would catalyze the kinds of discussions and raise the kinds of questions we hoped it would, and supported the proof of principle for the DDM.

2.2. *Orientation to the DDM*

In Figure 1, we show our current draft of the DDM. It represents our EOI Study findings that PMR studies often engaged in different research stages at the same time, as investigators iterated on study procedures after encountering issues or new conditions. Significantly, we found that multiple

seemingly minor research decisions made upstream incrementally shaped and constrained consequences for participant diversity, the kinds of data collected, relations with participants or community advisors, and the potential to aggregate data with those of other studies. Similarly, choices made at one time about one study activity often looped back to shape other activities and options available in the future. Thus, we constructed the DDM to indicate how research activities (in boxes) across the research life course (related but different from the research life cycle[23]) connect to and influence one another, through double-headed arrows that represent the potential mutual influence of different study stages and research activities. Not all possible arrows are depicted, and the study stages described in the boxes in the DDM are not intended to be exhaustive, but rather highlight common domains of research identified in our EOI data that were impacted by and influenced goals of diversity.
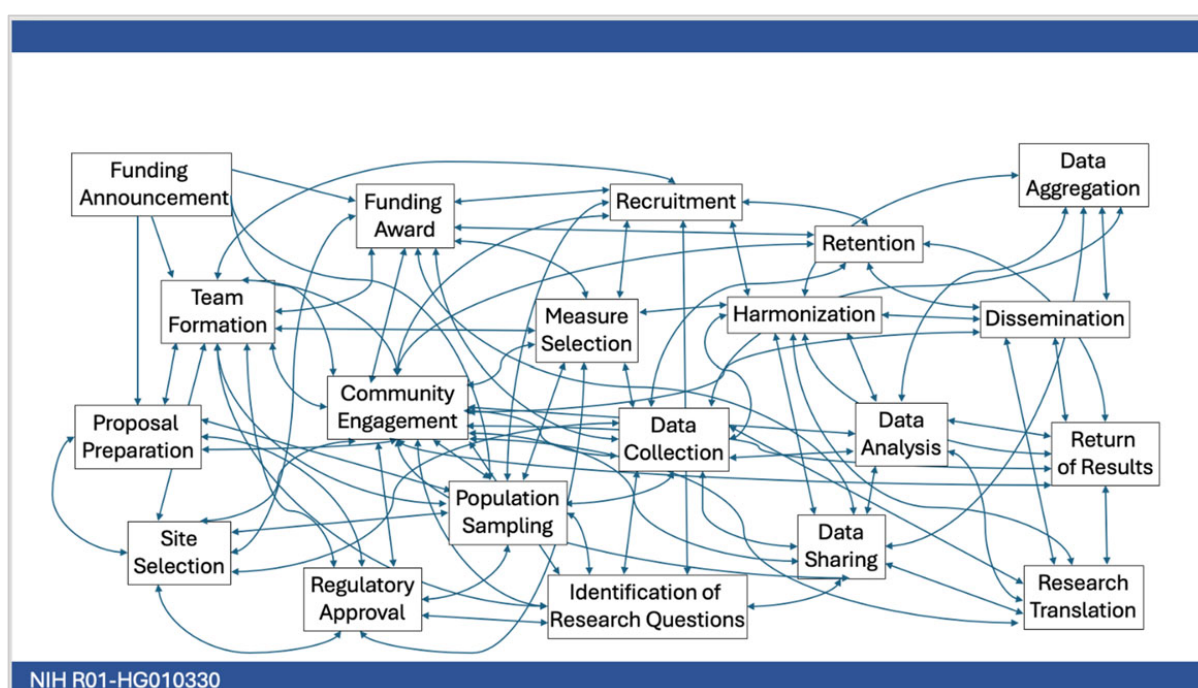

Fig. 1. Diversity decisions across the research life course

The DDM is not meant as a prescriptive, normative guide for what specific research decisions can and should be made to optimize equity. Instead, we developed the DDM to serve as a useful heuristic and tool to enable stakeholders to identify tradeoffs and facilitate stakeholder dialogue about those tradeoffs, at multiple stages of a study and even when the research is well underway (see our DDM User Guide[32]). To do so, we found that the DDM allowed stakeholders to locate where they are in the research process (much like a map reader who must pinpoint where they are before navigating their way). Stakeholders then could weigh competing interests, consider methodological and resource tradeoffs, anticipate the future consequences of different choices for

inclusion and equity, and seek to make informed decisions. The DDM thus helped guide the systematic consideration of the potential effects of study decisions about one set of research activities on others, by attending to how boxes are and might be connected to and influence others.

As a brief example of how PMR teams might use the DDM, we illustrate this by focusing on community engagement, located in the center-left of the DDM. Examining potential arrows between Community Engagement and other boxes leads to posing questions about whether, when, and how community engagement is or is not incorporated into the governance of the study throughout the research life course. The degree of commitment to Community Engagement affected choices about Team Formation, as represented in the arrow between those two boxes (see Figure 2): our EOI findings suggest that strong commitments to an engaged approach led to community partners needing to be part of study leadership, whereas understandings of community input as advisory meant that community engagement could happen more separately.[13] The relative integration of community input often affected who enrolled in PMR studies, as seen by arrows connected to Recruitment and Retention (both in top middle) and Population Sampling (bottom middle). Community Engagement could also be linked to other research stages such as Identification of Research Questions (bottom center): empirically, we found that decisions about who to engage, when, and how shaped whether community advisors provided input on what sorts of research questions were pursued, and therefore what data were collected and how.
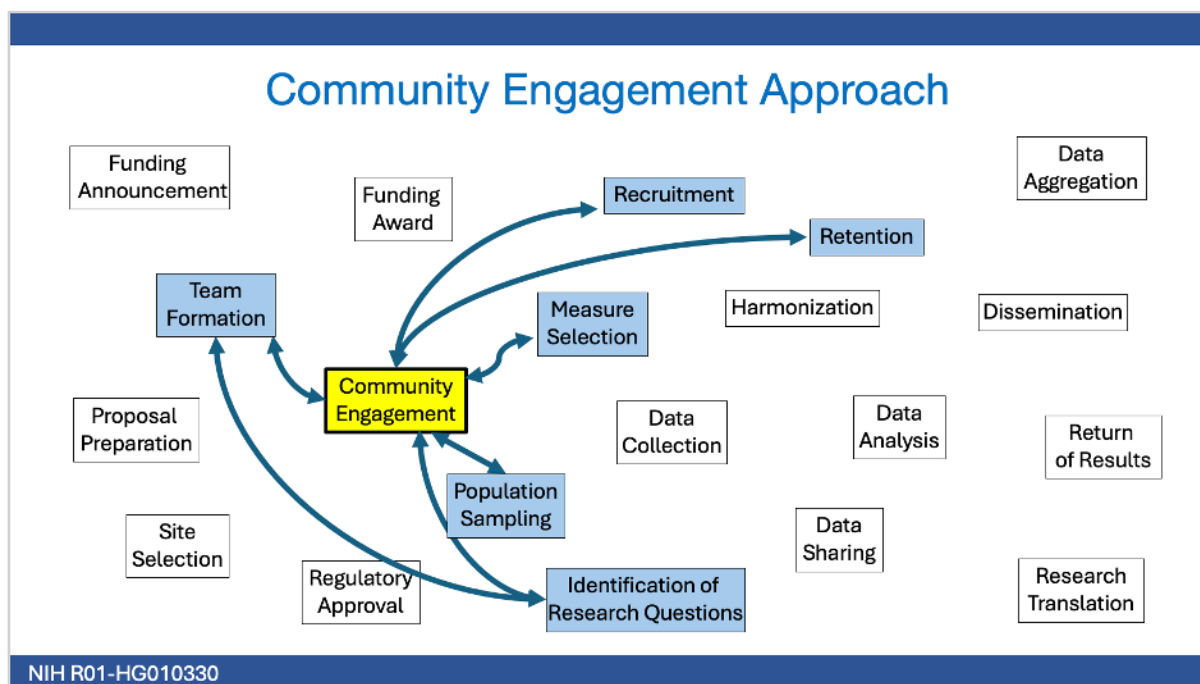

Fig. 2. Example of the DDM to analyze whether, when, and how community engagement is integrated into PMR studies

3. Mapping Key Decision Points in AI Development for Fair and Trustworthy AI

We believe that many of the same contexts, constraints, and issues that we observed in our EOI Study, with respect to values and practices related to diversity and equity in PMR, also play out in the development and deployment of AI tools in biomedicine. This domain is similarly marked by the competing interests and priorities of innovation and those of communities. Yet despite the growing influence of AI in health and medicine, there remain few formal mechanisms to ensure that considerations of fairness and trustworthiness, and their meanings by multiple stakeholders, are consistently integrated into decision-making.

3.1. *Applying the decision map to AI*

To support reflection and deliberation in this space, we adapted the original DDM into a new tool we call the *Trustworthy AI Decision Map,* or AI-DM for short (Figure 3). While several other frameworks exist for guiding ethical AI development—most notably those organized around high-level phases of the AI life cycle and principle-based ethical analysis—our approach differs in important ways. Rather than treating the stages of AI development as discrete or linear, the AI-DM emphasizes the dense network of *micro-decisions* that occur within and across those stages. Drawing from our prior findings, we argue that decisions impacting fairness and trustworthiness are not resolved at singular points or stages of the life cycle, but are continually shaped through seemingly routine or technical choices: how a problem or need is identified and defined, how a dataset is created, how missing data is handled, or how a model is implemented in a specific context. The AI-DM is designed to make these granular choices more visible and open to scrutiny, enabling teams to better anticipate downstream consequences and identify opportunities for intervention.
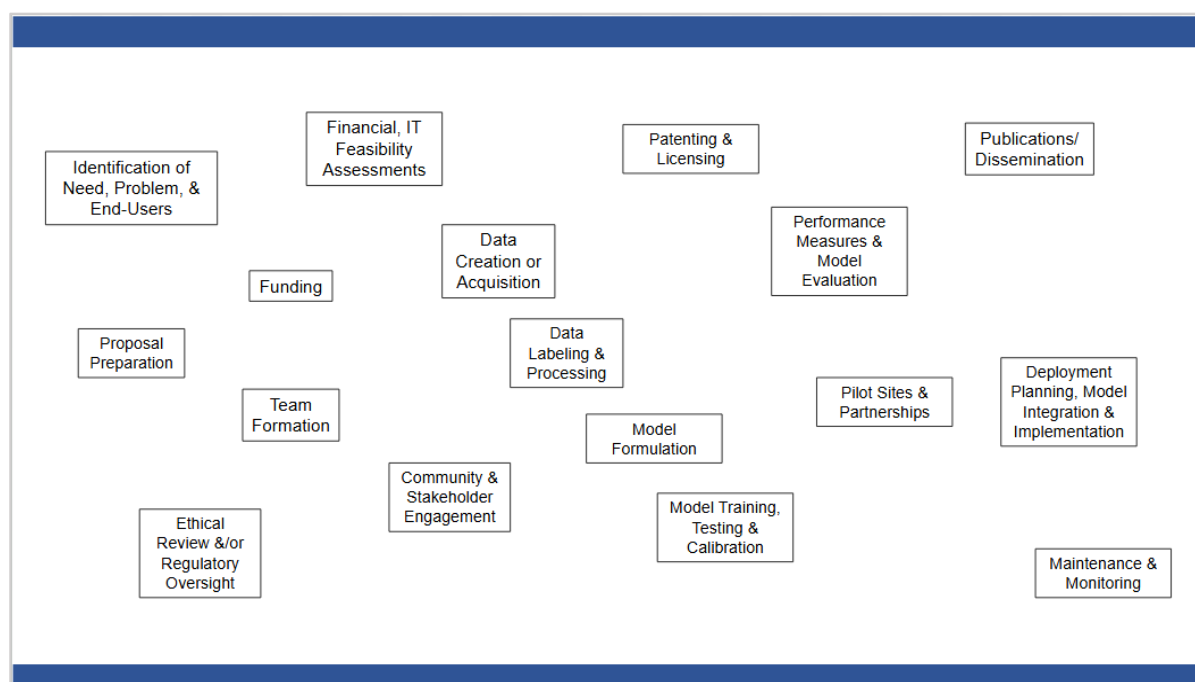
Fig. 3. Map of decision points impacting fairness and trustworthiness in biomedical AI development

The AI-DM is intended to help structure reflections on and dialogue about where and how issues around fairness and trustworthiness arise and where outcomes that impact these qualities are being shaped. While retrospection is always important, the key intervention we hope the AI-DM can have is to elicit conversations in proactive and anticipatory ways among multiple stakeholders. To do so, the AI-DM may be used by academic medical centers, health systems, research and other institutions to convene and conduct discussions focused on the fair and responsible development and deployment of an AI tool.

Data scientists, AI developers, researchers and/or clinicians, regulatory staff, and patient and community representatives, among others, can participate in facilitator-led discussions that are anchored by and oriented to the AI-DM. Advance prep work and review of accompanying resources would familiarize participants with the AI-DM, its purpose, and the charge or agenda for the discussion. A facilitated session would then further introduce and explain the AI-DM and demonstrate its use. Participants would then discuss (in breakout groups depending on size) such prompts as:

● What stage is the AI tool in right now, and how does this impact perceptions of its fairness and trustworthiness?

Sub-questions: Have different stakeholder groups' definitions of fairness and trustworthiness been elicited, and how might these definitions differ? Are tradeoffs or

conflicts between fairness and trustworthiness explicitly acknowledged (e.g., an approach that appears "fair" across groups may still lack accountability or responsiveness, and thus be seen as untrustworthy)? What decisions, in what other boxes on the AI-DM, have already been made? Who has been involved in these decisions, whose input has been meaningfully considered, and to what extent has there been transparency around the reasoning for these decisions? What impacts do these previous decisions have on the current stage for fairness (e.g., new biases, representativeness) and trustworthiness (e.g., clarity of documentation, perceived responsiveness to concerns)? What options remain open at the current stage, and what have been closed off due to previous decisions? How do these open and closed options affect how the AI tool is seen as fair and trustworthy, by whom, and for whom? What additional ethical implications, such as equity and inclusion, are potentially at play at this decision point?

- After choosing (or being assigned) a decision-making point to discuss, consider at least two different options or pathways.

  Sub-questions: By locating and considering different boxes on the AI-DM, what are the potential downstream consequences related to fairness and trustworthiness that the different pathways might impact? For example, which stakeholder groups are most affected by this decision point and the different pathways? Are there processes in place to ensure these groups are aware of these potential effects, and to what extent will they be engaged in shaping future decisions about what paths will be pursued? To what extent do stakeholders see clear reasoning and accountability in how choices were made? How might each pathway differently affect trust-building with various communities? What fairness concerns might emerge from each option (e.g., are benefits and harms equitably distributed)? What perspectives on fairness and trustworthiness from affected stakeholders might be missing from this analysis? By tracing pathways between the current stage and other boxes, systematically walk through how a specific decision at the current stage could impact options at future stages. Generate explicit discussion about potential unintended ethical consequences by tracing pathways from the current stage to boxes that appear to be indirectly connected or located farther downstream.

- Synthesize and think iteratively and systematically:

  Sub-questions: Look back upstream: Could earlier decisions be modified or redirected in ways that might change the downstream pathways available, and their effects for fairness and trustworthiness? What different pathways forward have been identified, and what are their potential tradeoffs and consequences? What has not been considered on the AI-DM? Are there boxes or pathways that have not been explored or discussed that could impact fairness and trustworthiness?

The facilitated session would then end with reflections and strategizing next steps based on the discussion.

3.2. *Example of applying the AI-DM: Thinking through the consequences of stakeholder engagement on ethical AI*

To illustrate the potential utility of the AI-DM, we consider a real-world example (anonymized here for confidentiality) in which a developer created an AI-enabled chatbot assistant for doctors, designed specifically to support rural clinicians. The chatbot was designed to be trained on institution-specific content to assist with clinical questions about diagnoses and treatments, clinical guidelines, drug information, and more. Keeping clinicians up to date on rapidly evolving medical knowledge is a persistent challenge, especially in rural contexts, and such a tool could help rural doctors provide higher-quality care for their patients. Yet, realizing this potential requires careful consideration of how rural medical practices can differ from urban settings. Underresourced clinics and hospitals often face serious constraints in infrastructure and staffing, which means that questions of fairness and trustworthiness may emerge differently than in better-resourced settings. We see this not as a reason to exclude rural clinics from AI deployment; if AI implementation occurs only in well-resourced institutions, it risks compounding inequities in access to the benefits of AI. Instead, we suggest that our tool may provide a mechanism to surface and consider how AI development can be responsive to local conditions.

The AI-DM (Figure 4) can help make potential ethical and technical issues that impact fairness and trustworthiness visible by drawing out important connections between different decision points in the AI development life cycle. Community and stakeholder engagement is especially critical as it connects with upstream stages such as identifying needs and end-users, and with downstream areas like deployment, evaluation, and maintenance. By involving rural clinicians, patients, and administrators early, developers can surface potential mismatches between what the tool is designed to deliver and what is actually feasible in underresourced contexts. For example, presumptions about access to certain technologies (e.g., stable Wi-Fi), or requirements that place added burdens on patients—such as traveling long distances for follow-up tests—can be identified and addressed before the tool is deployed. Similarly, engagement can clarify whether rural clinics have the infrastructure and staff to test, integrate, and sustain a tool of this kind.

The *Community and Stakeholder Engagement* box is positioned as a point of connection, not an isolated stage, and links to multiple other decision points in the life cycle. At the stage of *Identifying Need, Problem, and End-users*, rural clinicians and patients, as well as other case managers and providers of social services, for example, can clarify what kinds of assistance would be most useful or practical, and flag and modify advice from the chatbot that might be unrealistic in underresourced contexts. With respect to *Data Acquisition* and *Model Training*, stakeholder

engagement with experienced rural physicians and providers, perhaps even epidemiologists and other experts in rural health, could help ensure that conditions more prevalent in rural populations are represented, and/or assessing whether the tool might inadvertently be optimized for urban patient populations, given that clinical research is predominantly conducted by academic medical centers located in urban areas. In *Deployment Planning and Model Integration*, stakeholders with diverse perspectives and deep experiences in rural healthcare can surface infrastructural barriers that can introduce inequities such as bandwidth limitations or IT support needs, or inappropriate assumptions about clinic workflow and operations, that may otherwise derail uptake. In *Performance Measures and Evaluation*, they can shape definitions of success that extend beyond accuracy (e.g., correct diagnosis) to those identified by affected stakeholders, such as whether the tool eases clinical burdens or improves healthcare encounters from the perspective of patients. Even downstream areas such as *Maintenance and Monitoring* could benefit from multi-stakeholder dialogue, such as who is responsible for adapting the chatbot to local hospital protocols, keeping it up to date, and ensuring that the tool remains sustainable if internal systems shift, and where will *Funding* and other resources come from to support these key tasks?
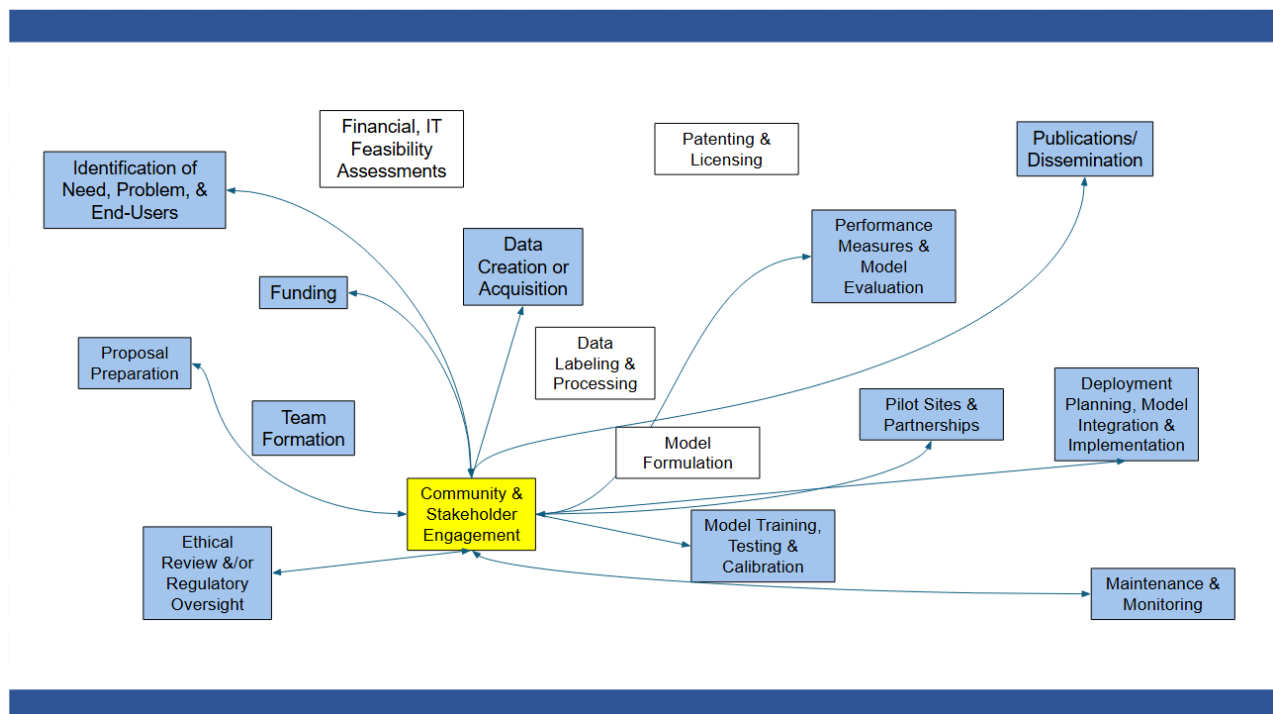

Fig. 4. Example of the Trustworthy AI Decision Map focusing on community and stakeholder engagement

In this way, the AI-DM helps identify where ethical and practical challenges may arise that impact fairness and trustworthiness, while also structuring cross-stakeholder engagement so that these issues can be surfaced early and revisited as they evolve. It prompts developers, health system

administrators, and other leaders in decision-making positions to work with diverse and multiple groups of stakeholders to identify, consider, and keep sight of downstream consequences and future implications of present choices, fostering more intentional and forward-looking decision-making across the AI development life cycle.

4. Conclusion

This paper has shown that fairness, trust, and trustworthiness in biomedical AI are situated concepts that require dialogue rather than fixed definitions. By adapting our Diversity Decision Map, we propose a tool to elicit diverse perspectives, support multi-stakeholder engagement, and guide decision-making throughout the AI life cycle that promotes fairness and trustworthiness. We argue that responsibility rests at the institutional level, and leaders and researchers need tools and support to carry this work forward. Future testing of the AI-DM with stakeholders will be key to evaluating its value for fostering accountability and shared responsibility in biomedical AI in order to promote fair and trustworthy AI.

5. Acknowledgments

References

1. Vakkuri V, Kemell KK, Jantunen M, Abrahamsson P. "This is Just a Prototype": How Ethics Are Ignored in Software Startup-Like Environments. *Agile Processes in Software Engineering and Extreme Programming*. 2020;383:195-210. doi:10.1007/978-3-030-49392-9_13

2. Richardson B, Gilbert JE. A Framework for Fairness: A Systematic Review of Existing Fair AI Solutions. *arXiv*. Preprint posted online December 10, 2021. doi:10.48550/arXiv.2112.05700

3. Zhang J, Zhang Z ming. Ethics and governance of trustworthy medical artificial intelligence. *BMC Medical Informatics and Decision Making*. 2023;23(1):7. doi:10.1186/s12911-023-02103-9

4. Verma S, Rubin J. Fairness definitions explained. In: *Proceedings of the International Workshop on Software Fairness*. ACM; 2018:1-7. doi:10.1145/3194770.3194776

5. Cook KS, Santana JJ. Trust: Perspectives in Sociology. In: Simon J, ed. *The Routledge Handbook of Trust and Philosophy*. Routledge; 2020.

6. Cook KS, Hardin R, Levi M. *Cooperation Without Trust?* Russell Sage Foundation; 2005. Accessed August 30, 2025. https://www.jstor.org/stable/10.7758/9781610441353

7. Kraft SA, Cho MK, Gillespie K, et al. Trustworthiness in Untrustworthy Times: Response to Open Peer Commentaries on Beyond Consent. *The American Journal of Bioethics*. 2018;18(5):W6-W8. doi:10.1080/15265161.2018.1461953

8. U.S. Department of Human Services. *Trustworthy AI (TAI) Playbook*.; 2021. https://digitalgovernmenthub.org/wp-content/uploads/2023/08/hhs-trustworthy-ai-playbook.pdf

9. Sikstrom L, Maslej MM, Hui K, Findlay Z, Buchman DZ, Hill SL. Conceptualising fairness: three pillars for medical algorithms and health equity. *BMJ Health Care Inform*. 2022;29(1):e100459. doi:10.1136/bmjhci-2021-100459

10. Parthasarathy S, Katzman J. Bringing Communities In, Achieving AI for All. *Issues in Science and Technology*. 2024;XL(4):41-44.

11. Banerjee S, Alsop P, Jones L, Cardinal RN. Patient and public involvement to build trust in artificial intelligence: A framework, tools, and case studies. *Patterns*. 2022;3(6):100506. doi:10.1016/j.patter.2022.100506

12. Domecq JP, Prutsky G, Elraiyah T, et al. Patient engagement in research: a systematic review. *BMC Health Serv Res*. 2014;14(1):89. doi:10.1186/1472-6963-14-89

13. Shim JK, Foti N, Vasquez E, et al. Community Engagement in Precision Medicine Research: Organizational Practices and Their Impacts for Equity. *AJOB Empirical Bioethics*. 2023;14(4):185-196. doi:10.1080/23294515.2023.2201478

14. Holzer JK, Ellis L, Merritt MW. Why We Need Community Engagement in Medical Research. *Journal of Investigative Medicine*. 2014;62(6):851-855. doi:10.1097/JIM.0000000000000097

15. Pratt B, Cheah PY, Marsh V. Solidarity and community engagement in global health research. *The American Journal of Bioethics*. 2020;20(5):43-56. doi:10.1080/15265161.2020.1745930

16. Ng MY, Kapur S, Blizinsky KD, Hernandez-Boussard T. The AI life cycle: a holistic approach to creating ethical AI for health decisions. *Nat Med*. 2022;28(11):2247-2249. doi:10.1038/s41591-022-01993-y

17. Callahan A, McElfresh D, Banda JM, et al. Standing on FURM Ground: A Framework for Evaluating Fair, Useful, and Reliable AI Models in Health Care Systems. *NEJM Catalyst*. 2024;5(10). doi:10.1056/CAT.24.0131

18. Trotsyuk AA, Waeiss Q, Bhatia RT, et al. Toward a framework for risk mitigation of potential misuse of artificial intelligence in biomedical research. *Nat Mach Intell*. 2024;6(12):1435-1442. doi:10.1038/s42256-024-00926-3

19. Lee SSJ, Fullerton SM, McMahon CE, et al. Targeting Representation: Interpreting Calls for Diversity in Precision Medicine Research. *Yale J Biol Med*. 2022;95(3):317-326.

20. Jeske M, Vasquez E, Fullerton SM, et al. Beyond inclusion: Enacting team equity in precision medicine research. *PLOS ONE*. 2022;17(2):e0263750. doi:10.1371/journal.pone.0263750

21. Shim JK, Bentz M, Vasquez E, et al. Strategies of inclusion: The tradeoffs of pursuing "baked in" diversity through place-based recruitment. *Social Science & Medicine*. 2022;306:115132. doi:10.1016/j.socscimed.2022.115132

22. McMahon CE, Foti N, Jeske M, et al. Interrogating the Value of Return of Results for Diverse Populations: Perspectives from Precision Medicine Researchers. *AJOB Empirical Bioethics*. 2024;15(2):108-119. doi:10.1080/23294515.2023.2279965

23. Shim JK, McMahon CE, Saco L, Bentz M, Foti N, Lee SSJ. A qualitative study of diversity in precision medicine research: The development and stakeholder assessment of a Diversity Decision Map. *Journal of Clinical and Translational Science*. 2025;9(1):e78. doi:10.1017/cts.2025.45

24. Claw KG, Anderson MZ, Begay RL, Tsosie KS, Fox K, Garrison NA. A framework for enhancing ethical genomic research with Indigenous communities. *Nat Commun*. 2018;9(2957):1-7. doi:10.1038/s41467-018-05188-3

25. Rebbeck TR, Bridges JFP, Mack JW, et al. A Framework for Promoting Diversity, Equity, and Inclusion in Genetics and Genomics Research. *JAMA Health Forum*. 2022;3(4):e220603. doi:10.1001/jamahealthforum.2022.0603

26. Watson KS, Cohn EG, Fair A, et al. Adapting a conceptual framework to engage diverse stakeholders in genomic/precision medicine research. *Health Expectations*. 2022;25(4):1478-1485. doi:10.1111/hex.13486

27. Hudson M, Beaton A, Milne M, et al. *Te Mata Ira: Guidelines for Genomic Research with Māori*. Te Mata Hautū Taketake - Māori & Indigenous Governance Centre, University of Waikato; 2016.

28. Tindana P, Yakubu A, Staunton C, et al. Engaging research ethics committees to develop an ethics and governance framework for best practices in genomic research and biobanking in Africa: the H3Africa model. *BMC Medical Ethics*. 2019;20(1):69. doi:10.1186/s12910-019-0398-2

29. Blanchard J, Hiratsuka V, Beans JA, et al. Power sharing, capacity building, and evolving roles in elsi: the center for the ethics of indigenous genomic research. *Collaborations: A Journal of Community-Based Research and Practice*. 2020;3(1):18. doi:10.33596/coll.71

30. Joosten YA, Israel TL, Williams NA, et al. Community Engagement Studios: A Structured Approach to Obtaining Meaningful Input From Stakeholders to Inform Research. *Academic Medicine*. 2015;90(12):1646-1650. doi:10.1097/ACM.0000000000000794

31. Tuttle KR, Knight R, Appelbaum PS, et al. Integrating Patient Priorities with Science by Community Engagement in the Kidney Precision Medicine Project. *CJASN*. 2021;16(4):660-668. doi:10.2215/CJN.10270620

32. The Ethics of Inclusion: Diversity in Precision Medicine Research. Department of Medical Humanities and Ethics. July 25, 2023. Accessed August 30, 2025. https://www.mhe.cuimc.columbia.edu/ethics/research/research-projects/ethics-inclusion-diversity-precision-medicine-research