# **Abstract**

This paper advocates for inclusive AI development, emphasizing its necessity for global equity, ethical soundness, and social relevance. We detail MIT Critical Data's approach to equitable AI development, focusing on healthcare. Our methods prioritize diverse collaboration and community engagement. Through global datathons, open-source datasets, and accessible education, we empower the global majority to actively participate in shaping AI that benefits all. Significant results, including numerous publications and established community hubs, demonstrate the impact of our approach. We argue that inclusivity in AI is not only achievable but crucial for its future success and fairness, particularly in serving the global majority.

# **The Importance of AI from the Global Majority**

While artificial intelligence (AI) influences society globally, its development and deployment are concentrated in technologically and economically dominant regions, leaving the majority of the global population underrepresented (World Health Organization, 2024). This disparity results in AI systems that do not reflect the diversity of the global majority. Consequently, these systems may perpetuate and exacerbate biases and inequities, further marginalizing already vulnerable populations (Shaffer, Alenichev, & Faure, 2023). The potential for AI to drive positive change is immense, but only if it is developed responsibly through a process that is participatory, inclusive, reflexive and reflective.

Inclusivity in AI development is not only a matter of equity, but is essential for any system that is ethically sound, socially relevant, and economically beneficial to all (Hendl & Shukla, 2024; Jansky, Hendl, & Nocanda, 2024). By involving diverse voices in the AI creation process, we ensure that these technologies are reflective of and responsive to the varied experiences and needs of different populations. At MIT Critical Data, we have taken these challenges head-on by employing a grassroots, local-first approach that prioritizes diversity and inclusivity in AI development. Through our initiatives, we aim to build a more equitable AI landscape that benefits everyone, not just a privileged few. This paper highlights our methods and the tangible outcomes of our work, demonstrating how inclusivity in AI is not only possible but essential for the technology’s future success and fairness.

# **MIT Critical Data’s Approach to Achieving Equitable Development, Transparency, and Accountability for AI**

At MIT Critical Data, we recognize that engaging diverse communities is essential to combating bias in healthcare AI. Our approach is derived from five distinct core values: rigorous and innovative research, multi-level and accessible teaching, building and networking communities of primary stakeholders, reimagining legacy systems of power, and advocacy for epistemic humility and health equity. We strive to unite the full range of professional, empirical, and cognitive backgrounds to foster collaborative imagination.

**Pioneering Research Methods in Healthcare AI**

We conduct our research under the premise that AI has equal capacity to revolutionize healthcare as it does to harm it. Relying solely on model prediction accuracy as the final arbiter for its implementation is short-sighted, not generalizable,

and risks significant harm to populations traditionally excluded from research and model training (Futoma, Simons, Panch, Doshi-Velez, & Celi, 2020). Rather than merely developing highly accurate models using robust methodologies, we prioritize addressing foundational challenges in machine learning for healthcare and incorporating any model development into the broader context of the data. Recognizing the many biases inherent to healthcare AI across all five stages of the pipeline (Gichoya et al., 2023), we have developed a well-validated checklist called TRIPOD-LLM (Gallifant et al., 2024). This tool helps quantify the severity of bias in published studies using LLM models, and also serves as a framework for responsibly designing prospective healthcare LLM studies. Key considerations before AI model design, validation, and

evaluation include identifying and involving community members who would be most impacted by the research, collaborating with co-authors from diverse backgrounds, openly discussing conflicts of interest of private industry, deeply understanding the data’s story and fidelity, mitigating “hidden signals” in the data (Gichoya et al., 2023), and committing to the replicability of digital research through open science (Seastedt et al., 2022; Watson et al., 2023).

AI research transcends any single cognitive or organizational domain and should not be developed, appraised, or regulated in a vacuum. There are no individual experts in AI; only collective wisdom exists. Our research ranges widely, including large language models, AI model error interrogation, causal reinforcement learning, scientometric analysis, network science, epistemic research, time-series deep learning of electronic medical record data, ethics, vector embeddings, scientometric studies and implementation science. As such, to ensure a holistic approach, we collaborate globally with a diverse array of experts including social scientists, computer engineers, network scientists, ethicists, philosophers, physicians, veterinarians, pharmacists, data scientists, and statisticians. We believe that healthcare AI research should be cultivated within the global majority through a crowd-sourced approach that bridges communities and disciplines, and advances the decentralization and democratization of health equity research. This mission is furthered by teaching knowledge and skills, empowering others to pass this understanding forward.

**Multi-level and Accessible Teaching**

Our teaching approach includes a wide range of training, education, expertise, age groups, and demographics. We partner with local and distant high schools and community colleges to advance healthcare AI education, sharing model development coding notebooks, providing access to open data sets, and offering tools to assess expected bias and harm. Our lab hosts a rotating cohort of visiting students from all over the world. We also teach at the Harvard School of Public Health, MIT, offer a freely available edX course, involve medical residents at Beth Israel Deaconess Medical Center, and engage in many more educational venues. We then translate this approach to durable, community-focused educational initiatives, particularly through global datathons, as discussed below.

*“Celebrating my new certification for successfully completing HST.953x: Collaborative Data Science for Healthcare on edX…. It is my first certification in Data Science. I started it a few days after meeting Leo Anthony Celi during a symposium organized by Mbarara University Data Science Research Hub back in February 2024. His passion for data*

*science in healthcare has opened up a new world to me of constant learning and adventure. I am eternally grateful to [the team] for putting up this course.”*

* *Testimonial from a student from Mbarara University of Science & Technology*

**Investment and Networking of Relevant Communities**

AI in healthcare cannot succeed without recentering the global majority to the forefront of AI development and regulation. Our approach centers on elevating primary actors involved in AI model development by incorporating their perspective into the TRIPOD-LLM bias assessment tool and validating a team scorecard applicable to any healthcare AI project. We work to establish community hubs–organically scaled networks that bring together people from neighboring countries and regions. These hubs serve as grassroots initiatives, fostering a community of committed individuals committed to advancing equitable AI. By building these networks, we ensure that capacity is already present within the communities,

and our role is to connect and empower them. At the local level we nurture the next generation of AI leaders, equipping them with the critical perspectives needed to challenge prevailing biases in healthcare datasets. Our ethos is that critical thinking cannot thrive in a room where everyone thinks the same way. We believe that diversity in thought and experience is key to developing AI that is truly inclusive and effective.

We also enhance the technical capacity of stakeholder communities through a global network of datathons and policy camps (Aboab et al., 2016). These events provide spaces where interdisciplinary teams can critically engage with open health datasets. Our datathons are immersive, multi-day events held in countries across the globe. Participants collaborate to uncover and address biases that could influence AI models, ensuring that these technologies prioritize health equity. These events are not just confined to capital cities; we also host them in smaller towns and regions that are often overlooked in global initiatives. This approach allows us to engage talented individuals from various backgrounds, ensuring that the AI solutions reflect the communities they are designed to serve. Furthermore, our events are conducted in local languages, enabling participants to communicate and collaborate effectively, regardless of their linguistic backgrounds.

## **Reimagine Legacy Systems of Power and Expertise**

We stand for the decentralization of medical knowledge and the democratization of clinical data sharing. To achieve this we’ve advocated for alternative metrics beyond the traditional impact factor to evaluate the impact of scientific journals, promoted open access, and supported open science to maximize scientific replicability. Our focus on data which is Findable Accessible Interoperable and Reusable (FAIR) reflects our dedication to transparency (Jacobsen et al., 2020). We host PhysioNet, a continually-building collection of 314 large physiological and clinical datasets (at time of writing), over 50 related open-source software packages, and over 30 tutorials and reference guides. Among these datasets is the well-known Medical Information Mart for Intensive Care (MIMIC) now in its fourth iteration (MIMIC-IV), which includes data on 12,881 patients and 13,941 ICU stays from 2010-2018. Branches of this data set include raw CXR images, ECG waveforms, echocardiograms, emergency department encounters, and free-text clinical notes for large language models. All code is freely available and access is regulated through a data use agreement. Preliminary data shows that MIMIC

datasets are cited significantly more often than several proprietary publicly available datasets, with citation numbers ranging from 48.8-2,523.7 times higher, an advantage that grows further when adjusting for funding received.

## **Advocacy for Epistemic Humility and Digital Health Equity**

Regulating AI in healthcare requires recognizing that there is no one-size-fits-all solution; however, several common principles should be generally understood. First and foremost, the authority to regulate these systems must be primarily informed by those most affected. For instance, if a healthcare AI model is to be trained and applied to people with HIV in South Africa, then people with HIV in South Africa must have a seat at the table for every stage of its development. Secondly, transparency throughout the AI model’s lifecycle–from the conceptualization to implementation–must be proactively and universally available. Third, rather than evaluate the AI bias of a model *post-hoc*, there may be value in mandating a prospective, systematic evaluation of AI bias for every proposal. However, this alone does not guarantee a fair outcome. Data harmonization in response to such evaluations introduces another layer of bias with which to contend.

We are currently developing model interrogation tools to identify groups that might be harmed by false negative and positive predictions during the model validation stage. Nevertheless, creating a repository for those who were harmed in real-world implementation is challenging, as it loses the counterfactual, which significantly impairs interpretation. Numerous studies have identified “accurate” models in training and testing stages using conventional performance metrics, yet these models have ultimately caused harm or contributed to patient mortality when applied in real-world settings (McDermott, Hansen, Zhang, Angelotti, & Gallifant, 2024). While we have suggested some alternatives (Gallifant et al., 2023), these are still under development.

Designing solutions to AI harm is limited only by our collective imagination, and potential is lost when we surround ourselves with people who think exactly as we do. To counter this, we create symposia for epistemic humility and policy camps where individuals from any discipline can come together to discuss the broader ethical, regulatory, and societal implications of AI in healthcare. We have learned that before regulation of AI health equity among the global majority can be more fully addressed, there are clear structural and systemic challenges to engage. We must challenge the current power structure of academia which rewards publications, titles, and pedigree over the open dissemination of scientific knowledge. We need to continue developing AI error interrogation tools and alternative performance metrics that capture the humanity inherent to the data. It is essential to incentivize peer-reviewed journals to reject manuscripts which only report accuracy of yet another new AI model. Educated community actors in AI must replace policy-makers and public funding administrators who are content with a poor understanding of what AI is and how to use it safely. We must advocate for making science accessible from behind paywalls and ensuring it is understandable to those without the privilege of academic immersion. Health equity should be tied to healthcare reimbursements. Medical device laws must be reformed to accommodate AI application, prevent the use of AI by insurance payers to propagate discrimination, and we must intimately understand the story of our data in every manifestation. Collaboration in all forms, across disciplines, cognitive domains, cultures, religions, quantities of education, race/ethnicity, industries, and nations is essential to fully open the gates keeping AI from the global majority.

**3. Results and Impact of MIT Critical Data’s Approach**

Our results have been significant, both in terms of academic output and real-world impact. Since 2014 we have hosted 46 datathons in 21 unique countries, including Singapore, Taiwan, the Philippines, Mexico, and more. Over 2,000 publications have been produced and a formal network effect assessment is also underway. These papers not only advance the field of healthcare AI but also ensure that contributions come from a broad spectrum of voices, particularly those from the underrepresented global majority. There are over 9,000 citations from over 40,000 people using the over 300 open-source datasets hosted on the PhysioNet Platform. These citations reflect the widespread adoption and influence of the datasets we maintain, which are used by researchers globally to develop AI solutions. Importantly, many of these citations come from researchers affiliated with low- and middle-income countries (LMICs) and minority-serving institutions (MSIs) in the United States, highlighting our success in promoting greater authorship representation from these regions.

The establishment of critical hubs has played a pivotal role in our initiative’s success. By creating organically scaled networks that connect people across neighboring countries, we have fostered a sustainable and resilient community of AI practitioners. These hubs are not reliant on external funding guarantees but are instead driven by the shared commitment of their members to advance equitable AI. Further concrete outcomes of our approach are reflected in the high-impact publications that have emerged from our initiatives (Collins et al., 2024; Ellen et al., 2024; Gottesman et al., 2019; Gottlieb, Ziegler, Morley, Rush, & Celi, 2022; Komorowski, Celi, Badawi, Gordon, & Faisal, 2018; Wong et al., 2021; Wu et al., 2022). These publications are not just a measure of academic success; they represent real-world advances in how AI can be used to improve healthcare for diverse populations. By involving diverse stakeholders in the co-creation process, we have developed AI solutions that are not only technically robust but also aligned with the needs and realities of the communities they are designed to serve.

**4. Lessons Learned and Future Directions**

Through our work at MIT Critical Data, we have demonstrated that inclusivity in AI development is not only achievable but also essential for creating equitable healthcare solutions. Some challenges have included both ensuring sustained engagement from participants in underrepresented regions and bridging the gap between diverse linguistic and cultural contexts. We continually adapt our methods to ensure that our initiatives remain accessible and relevant. The need for sustained engagement underscores the importance of building long-term relationships with local communities, rather than relying on one-time events. Similarly, the diversity of linguistic and cultural contexts enriches the AI solutions developed through our initiatives, as they are informed by a broader range of perspectives and experiences.

As we look to the future, we urge the global AI community to recognize the value of engaging with diverse populations and to make a concerted effort to include voices from the global majority in AI decision-making processes. The future of AI in healthcare depends on our collective ability to build systems that are not only technologically advanced but also equitable and just. Together, we can create an AI landscape where every voice is heard, and every community benefits.