A Call for Universities to Develop Requirements for Community Engagement in AI Research

Emily Black*
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15232, USA
emilybla@andrew.cmu.edu

Michael A. Madaio
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15232, USA
mmadaio@andrew.cmu.edu

Joshua Williams*
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15232, USA
jnwillia@andrew.cmu.edu

Priya L. Donti
Computer Science Department
Department of Engineering & Public Policy
Carnegie Mellon University
Pittsburgh, PA 15232, USA
pdonti@andrew.cmu.edu.

* Equal contribution.

Community Engagement
In this piece, we use the term "community engagement" to refer to processes that involve stakeholders outside of academia, industry, or the government in AI research, with the purpose of guiding its design, ideation, and implementation. This engagement should enable "those who must live with the consequences of a decision to make it together" [7]. As Asad et al. observed in their work on citizen engagement in civic technology, levels of engagement may range from merely "tokenistic" to meaningful citizen input and control [3], inspired by Arnstein's Ladder [2] of civic participation. We thus use "engagement" to refer to levels of interaction above simply "informing or placating" [2], for mechanisms that provide citizens higher levels of control.

Every submission will be assigned their own unique DOI string to be included here.

Introduction
AI systems are being used in many facets of public life, in areas as diverse as policing, healthcare, and housing [10, 16]. However, many of these systems are developed largely in isolation of the communities they are meant to serve. In the best case, this may lead to applications that are improperly specified or scoped, and are thereby ineffective; in the worst case, it can lead (and has led [14]) to harmful, biased outcomes for marginalized populations.

In response, a growing set of voices has called for meaningful community engagement in the design of public-facing AI research (i.e., AI research likely to impact the public) (e.g., see [19]). However, despite emerging HCI methods for engaging stakeholders throughout the AI design process [1, 11], members of impacted communities are too often asked for feedback only after deployment [5]. We believe this disconnect occurs in large part because academic AI researchers lack organizational incentives to actually use existing community engagement methods, as has been seen with industry AI practitioners adopting AI fairness methods [12].

In light of this, we call for universities to develop and implement requirements for community engagement in AI research. These requirements should ensure that AI researchers designing public-facing systems make the needs and inter-

Lack of Community Engagement in Public-facing AI Projects

Too often, academic AI projects do not engage affected populations during the design process, often leading to harmful outcomes and community dissatisfaction. Prominent examples include predictive policing in Chicago and LA, a healthcare allocation tool in Arkansas, and pre-trial risk assessment tools in Ohio [17, 10, 9]. To the best of our knowledge, the creators of these tools did not engage impacted communities during the design process [17, 10, 9].

When deployed, these tools caused significant harm and were met with backlash from local communities. The Arkansas healthcare tool cut off resources for in-home care from hundreds of citizens, often putting them in danger [10]. In Chicago and LA, grassroots groups Erase the Database[a] and StopLAPD Spying[b] were created in response to the use of predictive policing technologies.

[a] http://erasethedatabase.com/about/

[b] https://stoplapdspying.org

ests of impacted communities a fundamental part of their work—and, crucially, engage community members throughout the design and deployment of this work. We propose that universities create these requirements so that: (1) university-based AI researchers will be incentivized to incorporate meaningful community engagement *throughout* the research life cycle, (2) the resulting research is more effective at serving the needs and interests of *impacted communities*, not simply the stakeholders with greater influence, and (3) the AI field values the process and challenge of community engagement as an important contribution in its own right.

Community Engagement in the AI Pipeline

The history of research in human-centered technology design (e.g., HCI) suggests that it is critical for stakeholders to be involved throughout the process of technology creation. For instance, many HCI methods are available for community-based participatory design [8] and for community engagement in the design of civic technology [3]. For high-stakes AI systems, involvement early in the design process is particularly critical, as stakeholders may be harmed before systems are fully deployed. For example, randomized control trials of predictive policing algorithms may adversely affect marginalized populations during even the trial phase.[1]

While AI researchers conducting community-oriented research are often free to choose whether and how to adopt participatory methods (e.g., [1, 11]), current cultural incentives disincentivize meaningful, sustained community engagement in practice. The time-intensive nature of community engagement is often at odds with the extreme emphasis on rapid publication in top-tier AI conferences, e.g., ICML and NeurIPS.

[1] American Civil Liberties Union. Statement of Concern About Predictive Policing by ACLU and 16 Civil Rights Privacy, Racial Justice, and Technology Organizations. https://www.aclu.org/other/statement-concern-about-predictive-policing-aclu-and-16-civil-rights-privacy-racial-justice. Accessed: 2019-02-11.

In addition, community engagement is often not viewed as a necessary component of AI systems, or as a technical contribution in its own right. Exacerbated by institutional factors such as the large number of papers required to acquire an academic job or receive tenure, this may inhibit the practice of community engagement in academic AI research. Similar patterns hold true in industry; as prior work on industry AI teams has found, practitioners may have the best intentions to develop ethical systems, but may not be incentivized by their organizational culture to actually use existing fairness methods [12]. These practical realities necessitate explicit guidelines or requirements to ensure that researchers meaningfully engage communities in the design of their research.

Existing University-based Research Guidelines

Existing university mechanisms for protecting the public welfare are not satisfactory for representing community interests. The most common such mechanism, institutional review boards (IRBs), can approve, reject, or require modifications of projects that involve human subjects [13]. Yet, in doing so, these boards do not often garner substantial involvement from the community, as only *one* IRB member is required to represent “non-scientific” concerns [15]. The IRB’s goal to reduce harm to individual subjects, even if it includes this single “non-scientific” voice, is not equivalent to meaningfully involving community stakeholders in the research process (to, e.g., shape research questions and goals).

Notably, there is a growing movement within the medical field to develop guidelines, institutes, and incentive structures at universities that incorporate community engagement into the medical research process; see [4] for a review. Crucially, universities have also taken steps to strongly incentivize these guidelines—for example, by emphasizing com-

IRB and AI: The Need for AI-Specific Requirements
Despite analogous requirements for human subjects research (i.e., the IRB) and community-engaged medical research, there are unique challenges associated with AI research, suggesting a need for *AI-specific* community engagement requirements. A prominent example of the failure of existing requirements to address AI is the "gay-face" project by Wang et al. [18]. This project used facial recognition to predict sexual orientation and, as noted by the authors, was approved by the IRB. This example illustrates that the framing of "human-subjects research" is not effective in protecting the public from potential harms of AI. Research that may affect thousands of people directly (e.g., predictive policing, healthcare) or indirectly (e.g., discriminatory advertising[a]) can be deployed without being considered human-subjects research.

[a]Katie Benner, Glenn Thrush, and Mike Isaac. Facebook engages in housing discrimination with its ad practices, US says. *The New York Times.* 28th Mar (2019).

munity engaged research in tenure qualifications.[2] The steps that the medical community is taking to center their research around communities may prove to be instructive in creating similar requirements and incentive structures for AI.

Proposal for Community-engaged AI Research

Introducing community engagement to AI research brings with it a set of unique challenges apart from analogous standards in other fields. Unlike fields such as medicine, the current AI research life cycle often does not include interactions with stakeholders; working only with data instead of people means that AI researchers have the ability to develop systems that change hundreds of lives [10] or impact culture [6] without ever involving anyone from impacted communities in their research. This suggests some questions for discussion:

1) What sorts of academic AI research should be considered "public-facing" (and thus require community engagement)?
2) What requirements might lead AI researchers to meaningfully engage communities in their research?
3) How might universities develop incentives to induce AI researchers to adopt community-engaged research methods, or, if they are requirements, how might universities mandate that researchers adopt them with sufficient fidelity?
4) What are open methodological or theoretical questions for the field of human-centered AI that may be precursors to workable community-engaged AI research guidelines, and how might HCI, AI, and social science research each contribute to developing these methods?

AI systems can have large, and potentially harmful, effects on the public. In order to make this technology work *for* the public, we must meaningfully incorporate public concerns and desires into the design process of AI systems. We believe that engaging members of impacted communities throughout the AI research process will result in AI systems that are fairer and more effective at serving the needs of impacted communities (or, in cases where it is the best course of action, to prevent harmful systems from ever being designed or deployed at all). However, if AI researchers are not required or incentivized to engage with community members (and provided with the resources needed to do so), current developments in human-centered AI methods may not be employed in practice. As such, we propose the development and implementation of community engagement requirements for university-based AI research. We cannot rely on the goodwill of individual researchers to meaningfully engage members of impacted communities; instead, we believe that institutional requirements may be most likely to shift the culture around community engagement in AI research.

[2]University of Washington, Population Health Initiative. Recognizing community-engaged research in promotion and tenure guidelines. https://www.washington.edu/populationhealth/2018/08/29/recognizing-community-engaged-research-in-promotion-and-tenure-guidelines/

REFERENCES

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, Jaime Teevan, Kikin-Gil Ruth, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proc. 2019 CHI Conf. on Human Factors in Computing Systems*. 1–13.

[2] Sherry R Arnstein. 1969. A ladder of citizen participation. *Journal of the American Institute of planners* 35, 4 (1969), 216–224.

[3] Mariam Asad, Christopher A Le Dantec, Becky Nielsen, and Kate Diedrick. 2017. Creating a sociotechnical API: Designing city-scale community engagement. In *Proc. 2017 CHI conference on human factors in computing systems*. 2295–2306.

[4] Joyce E Balls-Berry and Edna Acosta-Perez. 2017. The Use of Community Engaged Research Principles to Improve Health: Community Academic Partnerships for Research. *P R Health Sci. Jrnl.* 36, 2 (2017), 84.

[5] Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. 2019. Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services. In *Proc. 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[6] Noel Carroll. 2014. In search we trust: exploring how search engines are shaping society. *Intl. Jrnl. of Knowledge Society Research* 5, 1 (2014), 12–27.

[7] Archon Fung. 2009. *Empowered participation: Reinventing urban democracy*. Princeton Univ. Press.

[8] Christina Harrington, Sheena Erete, and Anne Marie Piper. 2019. Deconstructing Community-Based Collaborative Design: Towards More Equitable Participatory Design Engagements. *Proc. ACM on Human-Computer Interaction* 3, CSCW (2019), 1–25.

[9] Edward J Latessa, Richard Lemke, Matthew Makarios, and Paula Smith. 2010. The creation and validation of the Ohio Risk Assessment System (ORAS). *Fed. Probation* 74 (2010), 16.

[10] Colin Lecher. 2018. What happens when an algorithm cuts your health care. *The Verge* (2018).

[11] Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, Alexandros Psomas, and others. 2019. WeBuildAI: Participatory framework for algorithmic governance. *Proc. ACM on Human-Computer Interaction* 3, CSCW (2019), 1–35.

[12] Michael A Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. In *Proc. CHI Conference on Human Factors in Computing Systems*.

[13] Margaret R Moon. 2009. The History and Role of Institutional Review Boards: A Useful Tension. *AMA Journal of Ethics* 11, 4 (2009), 311–316.

[14] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 6464 (2019), 447–453.

[15] US Department of Health, Human Services, and others. 2018. Code of federal regulations (45 CFR 46). *Sub-part D: Additional Protections for Children Involved as Subjects in Research* (2018).

[16] Department of Housing and Urban Development. 2019. HUD's Implementation of the Fair Housing Act's Disparate Impact Standard. *Federal Register* (2019).

[17] David Robinson and Logan Koepke. 2016. Stuck in a Pattern: Early evidence on 'predictive policing'and civil rights. *Upturn report* (2016), 1–29.

[18] Yilun Wang and Michal Kosinski. 2018. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of personality and social psychology* 114, 2 (2018), 246.

[19] Meredith Whittaker, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Mysers West, Rashida Richardson, Jason Schultz, and Oscar Schwartz. 2018. *AI now report 2018*. AI Now Institute at New York University.