Reframing Opportunity and Fairness in the AI Diversity Pipeline

Armanda Lewis
New York University
al861@nyu.edu

CEUR-WS.org/Vol-2812/RDAI-2021_paper_7.pdf

Abstract

This whitepaper reframes the AI diversity pipeline to address persistent structural imbalances and achieve sustainable practices that support social justice, scientific, and technological advancements.

Recent statistics on the representation of historically underrepresented minority (URM) populations within the field of Artificial Intelligence (AI) reveal extreme disparities (Shoham et al. 2018). Though one observes disparities across all STEM fields, scholars highlight the AI case as particularly alarming given the rate and breadth at which AI technologies are adopted within society and the potential for AI technologies to replicate historical biases (Brundage et al. 2018; West, Whittaker, and Crawford 2019). Often, these reports focus on workforce implications, evoking the primacy of the school-to-professional pipeline to address disparities over time (Allen-Ramdial and Campbell 2014).

This paper offers a critical discussion of the AI diversity pipeline, highlighting existing approaches to fostering inclusivity, and articulating points of tension that may hinder sustainability and wider participation. Reframing how the AI pipeline functions is essential to achieving substantive, durable practices that support social justice, scientific, and technological advancements. We look to current higher educational research on learning, fairness, and inclusivity within STEM fields to discuss how broadscale representation can be balanced with the technical and bottomline priorities of AI development and deployment. At the heart is the need to identify the causes and intersecting effects of structural racism on AI as a field and as a community. One aim is to reconcile existing representational imbalances of URM populations with some of the critical discussions on the potential detrimental effects of AI tools and processes that may exacerbate racial and ethnic inequalities, unleash models with opaque processes, and create new forms of disparate impact.

An overview of existing AI diversity pipeline initiatives reveal interventions that

- are linear. Such interventions offer a small set of entry points, typically at the undergraduate or high school levels. They privilege a model based on professional development, positing that earlier and broader exposure to AI will increase underrepresented populations within the workforce. While logical, such processes tend to be leaky, excluding participants due to the limited number of pathways or disengaging members if they pursue non-hierarchical or delayed trajectories. (Scott et al. 2018);
- do not recognize holistically the perspectives of the URM member traversing AI. Existing initiatives offer technical training and professional overviews, but often fail to recognize the importance of inclusive cohort building and social identity development. These interventions may also not connect to issues of inclusion in post-educational environments;
- may target students from Minority Serving Institutions (MSIs), but do not account for complex challenges and disproportionate inequities observed at these institutions. Relatedly, recruitment efforts of URM faculty and staff may not acknowledge the plethora of commitments that mark their academic and professional lives; and
- may inadvertently replicate existing inequalities by resisting explicit discussion of race, gender, and intersectional identities.

Recently, McGee (McGee 2020) examines structural racial inequities within STEM higher education, positing that most diversity programs do not make sustained impact due to a lack of dialogue on structural racism and its effects. McGee argues for an in depth look at how historical exclusionary practices within STEM linger to the present time, and what links exist to scientific knowledge and education. The advanced pace at which AI is developing marks an urgent need to build better pathways that recognize the persistent and oft latent impacts of historical practices. Within ethical AI and related research, scholars have begun to reference the need for inclusion within AI, but focus on general topics like justice and fairness, transparency, professional responsibility, and promotion of human values (Fjeld et al. 2020; Jobin, Ienca, and Vayena 2019). Racial and identity-based equality is referenced, though in the context of marginalized identities as a singular group.

The AI diversity pipeline should take into account more contextual realities of URM learners and scholars, and incorporate existing best practices about STEM education pipelines that value various levels of mentorship, curricular and co-curricular supports, and professional training. The

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

pipeline must be specific to AI as an interdisciplinary, emergent, and multivalent field, and therefore look to a diverse set of participants who impact and are impacted by AI. A new AI diversity pipeline, therefore, should

- take place within an ethical framework dedicated to social justice within AI education, technical development, and professional practices. There should be a concerted effort to tie diversity initiatives to the ethical systemic change in AI curriculum development and evaluation;
- allow for multiple entry points at varying levels of technical preparation;
- offer more meaningful joint participatory initiatives between MSI institutions and majority-serving institutions and corporations;
- take into consideration structural imbalances that happen at the institutional level, particularly for partnerships between different institutional types;
- leverage current learning science research that is perspectival and contextual;
- dig deeper into the educational pipeline by building strong connections at the pre-secondary school level. Looking beyond undergraduate and even secondary education recognizes the value of early interventions as younger learners are developing their sense of selves and basic conceptual knowledge;
- include stakeholders at individual, department, institutional, and corporate levels, adding diverse disciplinary and role-based representation. Incorporating current stakeholder theory (Jones, Wicks, and Freeman 2017) is important to ground ethical value creation, though there need to be AI-specific models; and
- connect explicitly to research developments, which should in turn address diversity implications.

Current higher education scholarship calls for an acknowledgement of the role that systemic racism plays in disparities during and beyond the educational experience. Within the AI space, this acknowledgement would naturally lead to critiquing the meritocracy perspective that privileges technical expertise. This would also reposition AI research to introduce ethical implications more systematically in the quest to advance system or model accuracy, speed, and capacity.

Creating a successful AI diversity pipeline extends beyond recruiting URM members to populate our professional ranks and pulls together diverse stakeholders with a vested interest in making AI more inclusive. A sustainable model makes a place for social identity, and offers holistic experiences that resist the model of participants learning to fit within the existing mold of the AI professional or researcher or student.

References

Allen-Ramdial, S.-A. A.; and Campbell, A. G. 2014. Reimagining the Pipeline: Advancing STEM Diversity, Persistence, and Success. *BioScience* 64(7): 612–618.

Brundage, M.; Avin, S.; Clark, J.; Toner, H.; Eckersley, P.; Garfinkel, B.; Dafoe, A.; Scharre, P.; Zeitzoff, T.; Filar, B.; Anderson, H.; Roff, H.; Allen, G. C.; Steinhardt, J.; Flynn, C.; hÉigeartaigh, S. ; Beard, S.; Belfield, H.; Farquhar, S.; Lyle, C.; Crootof, R.; Evans, O.; Page, M.; Bryson, J.; Yampolskiy, R.; and Amodei, D. 2018. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. *arXiv:1802.07228 [cs]* URL http://arxiv.org/abs/1802.07228. ArXiv: 1802.07228.

Fjeld, J.; Achten, N.; Hilligoss, H.; Nagy, A.; and Srikumar, M. 2020. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. SSRN Scholarly Paper ID 3518482, Social Science Research Network, Rochester, NY. URL https://papers.ssrn.com/abstract=3518482.

Jobin, A.; Ienca, M.; and Vayena, E. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1(9): 389–399. ISSN 2522-5839. doi:10.1038/s42256-019-0088-2. URL https://www.nature.com/articles/s42256-019-0088-2.

Jones, T. M.; Wicks, A. C.; and Freeman, R. E. 2017. Stakeholder Theory: The State of the Art. In *The Blackwell Guide to Business Ethics*, 17–37. John Wiley & Sons, Ltd. ISBN 9781405164771. doi:10.1002/9781405164771.ch1. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781405164771.ch1.

McGee, E. O. 2020. Interrogating Structural Racism in STEM Higher Education. *Educational Researcher* 49(9): 633–644.

Scott, A.; Klein, F. K.; McAlear, F.; Martin, A.; and Koshy, S. 2018. The Leaky Tech Pipeline: A Comprehensive Framework for Understanding and Addressing the Lack of Diversity across the Tech Ecosystem. Technical report, Kapor Center for social Impact. URL https://leakytechpipeline.com/wp-content/themes/kapor/pdf/KC18001_report_v6.pdf.

Shoham, Y.; Perrault, R.; Brynjolfsson, E.; Clark, J.; Manyika, J.; Niebles, J. C.; Lyons, T.; Etchemendy, J.; Grosz, B.; and Bauer, Z. 2018. The AI Index 2018 Annual Report. Technical report, AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA.

West, S. M.; Whittaker, M.; and Crawford, K. 2019. Discriminating Systems: Gender, Race, and Power in AI. White Paper, AI Now Institute.