DISCRIMINATING SYSTEMS

Gender, Race, and Power in AI

Sarah Myers West, AI Now Institute, New York University
Meredith Whittaker, AI Now Institute, New York University, Google Open Research
Kate Crawford, AI Now Institute, New York University, Microsoft Research

APRIL 2019

Cite as: West, S.M., Whittaker, M. and Crawford, K. (2019). Discriminating Systems: Gender, Race and Power in AI. AI Now Institute. Retrieved from https://ainowinstitute.org/discriminatingsystems.html.

AINOW

CONTENTS

RESEARCH FINDINGS

There is a diversity crisis in the AI sector across gender and race. Recent studies found only 18% of authors at leading AI conferences are women,[i] and more than 80% of AI professors are men.[ii] This disparity is extreme in the AI industry:[iii] women comprise only 15% of AI research staff at Facebook and 10% at Google. There is no public data on trans workers or other gender minorities. For black workers, the picture is even worse. For example, only 2.5% of Google's workforce is black, while Facebook and Microsoft are each at 4%. Given decades of concern and investment to redress this imbalance, the current state of the field is alarming.

The AI sector needs a profound shift in how it addresses the current diversity crisis. The AI industry needs to acknowledge the gravity of its diversity problem, and admit that existing methods have failed to contend with the uneven distribution of power, and the means by which AI can reinforce such inequality. Further, many researchers have shown that bias in AI systems reflects historical patterns of discrimination. These are two manifestations of the same problem, and they must be addressed together.

The overwhelming focus on 'women in tech' is too narrow and likely to privilege white women over others. We need to acknowledge how the intersections of race, gender, and other identities and attributes shape people's experiences with AI. The vast majority of AI studies assume gender is binary, and commonly assign people as 'male' or 'female' based on physical appearance and stereotypical assumptions, erasing all other forms of gender identity.

Fixing the 'pipeline' won't fix AI's diversity problems. Despite many decades of 'pipeline studies' that assess the flow of diverse job candidates from school to industry, there has been no substantial progress in diversity in the AI industry. The focus on the pipeline has not addressed deeper issues with workplace cultures, power asymmetries, harassment, exclusionary hiring practices, unfair compensation, and tokenization that are causing people to leave or avoid working in the AI sector altogether.

The use of AI systems for the classification, detection, and prediction of race and gender is in urgent need of re-evaluation. The histories of 'race science' are a grim reminder that race and gender classification based on appearance is scientifically flawed and easily abused. Systems that use physical appearance as a proxy for character or interior states are deeply suspect, including AI tools that claim to detect sexuality from headshots,[iv] predict 'criminality' based on facial features,[v] or assess worker competence via 'micro-expressions.'[vi] Such systems are replicating patterns of racial and gender bias in ways that can deepen and justify historical inequality. The commercial deployment of these tools is cause for deep concern.

i. Element AI. (2019). Global AI Talent Report 2019. Retrieved from https://jfgagne.ai/talent-2019/.
ii. AI Index 2018. (2018). Artificial Intelligence Index 2018. Retrieved from http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf.
iii. Simonite, T. (2018). AI is the future - but where are the women? *WIRED*. Retrieved from https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/.
iv. Wang, Y., & Kosinski, M. (2017). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology*.
v. Wu, X. and Zhang, X. (2016). Automated Inference on Criminality using Face Images. Retrieved from https://arxiv.org/pdf/1611.04135v2.pdf.
vi. Rhue, L. (2018). Racial Influence on Automated Perceptions of Emotions. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3281765.

RECOMMENDATIONS

Recommendations for Improving Workplace Diversity

1. Publish compensation levels, including bonuses and equity, across all roles and job categories, broken down by race and gender.
2. End pay and opportunity inequality, and set pay and benefit equity goals that include contract workers, temps, and vendors.
3. Publish harassment and discrimination transparency reports, including the number of claims over time, the types of claims submitted, and actions taken.
4. Change hiring practices to maximize diversity: include targeted recruitment beyond elite universities, ensure more equitable focus on under-represented groups, and create more pathways for contractors, temps, and vendors to become full-time employees.
5. Commit to transparency around hiring practices, especially regarding how candidates are leveled, compensated, and promoted.
6. Increase the number of people of color, women and other under-represented groups at senior leadership levels of AI companies across all departments.
7. Ensure executive incentive structures are tied to increases in hiring and retention of under-represented groups.
8. For academic workplaces, ensure greater diversity in all spaces where AI research is conducted, including AI-related departments and conference committees.

Recommendations for Addressing Bias and Discrimination in AI Systems

9. Remedying bias in AI systems is almost impossible when these systems are opaque. Transparency is essential, and begins with tracking and publicizing where AI systems are used, and for what purpose.
10. Rigorous testing should be required across the lifecycle of AI systems in sensitive domains. Pre-release trials, independent auditing, and ongoing monitoring are necessary to test for bias, discrimination, and other harms.
11. The field of research on bias and fairness needs to go beyond technical debiasing to include a wider social analysis of how AI is used in context. This necessitates including a wider range of disciplinary expertise.
12. The methods for addressing bias and discrimination in AI need to expand to include assessments of whether certain systems should be designed at all, based on a thorough risk assessment.

INTRODUCTION

There is a diversity crisis in the AI industry, and a moment of reckoning is underway. Over the past few months, employees have been protesting across the tech industry where AI products are created. In April 2019, Microsoft employees met with CEO Satya Nadella to discuss issues of harassment, discrimination, unfair compensation, and lack of promotion for women at the company.[1] There are claims that sexual harassment complaints have not been taken seriously enough by HR across the industry.[2] And at Google, there was an historic global walkout in November 2018 of 20,000 employees over a culture of inequity and sexual harassment inside the company, triggered by revelations that Google had paid $90m to a male executive accused of serious misconduct.[3]

This is just one face of the diversity disaster that now reaches across the entire AI sector. The statistics for both gender and racial diversity are alarmingly low. For example, women comprise 15% of AI research staff at Facebook and just 10% at Google.[4] It's not much better in academia, with recent studies showing only 18% of authors at leading AI conferences are women,[5] and more than 80% of AI professors are male.[6] For black workers, the picture is worse. For example, only 2.5% of Google's workforce is black,[7] while Facebook and Microsoft are each at 4%.[8,9] We have no data on trans workers or other gender minorities. Given decades of concern and investment to redress the imbalances, the current state of the field is alarming.

The diversity problem is not just about women. It's about gender, race, and most fundamentally, about power.[10] It affects how AI companies work, what products get built, who they are designed to serve, and who benefits from their development.

This report is the culmination of a year-long pilot study examining the scale of AI's current diversity crisis and possible paths forward. This report draws on a thorough review of existing literature and current research working on issues of gender, race, class, and artificial intelligence. The review was purposefully scoped to encompass a variety of disciplinary and methodological perspectives, incorporating literature from computer science, the social sciences, and humanities. It represents the first stage of a multi-year project examining the intersection of gender, race, and power in AI, and will be followed by further studies and research articles on related issues.

1 Tiku, N. (2019, Apr. 4). Microsoft Employees Protest Treatment of Women to CEO Nadella. *WIRED*. Retrieved from https://www.wired.com/story/microsoft-employees-protest-treatment-women-ceo-nadella/.

2 Gershgorn, D. (2019, Apr. 4). Amid employee uproar, Microsoft is investigating sexual harassment claims overlooked by HR. *Quartz*. Retrieved from https://qz.com/1587477/microsoft-investigating-sexual-harassment-claims-overlooked-by-hr/.

3 Statt, N. (2018, Nov. 2). Over 20,000 Google employees participated in yesterday's mass walkout. *The Verge*. Retrieved from https://www.theverge.com/2018/11/2/18057716/google-walkout-20-thousand-employees-ceo-sundar-pichai-meeting.

4 Simonite, T. (2018). AI is the future - but where are the women? *WIRED*. Retrieved from https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/.

5 Element AI. (2019). Global AI Talent Report 2019. Retrieved from https://jfgagne.ai/talent-2019/.

6 AI Index 2018. (2018). Artificial Intelligence Index 2018. Retrieved from http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf.

7 Google. (2018). Google Diversity Annual Report 2018. Retrieved from https://static.googleusercontent.com/media/diversity.google/en//static/pdf/Google_Diversity_annual_report_2018.pdf.

8 Williams, M. (2018, July 12). Facebook 2018 Diversity Report: Reflecting on Our Journey. Retrieved from https://newsroom.fb.com/news/2018/07/diversity-report/

9 Microsoft. (2019). Diversity & Inclusion. Retrieved from https://www.microsoft.com/en-us/diversity/default.aspx.

10 As authors of this report, we feel it's important to acknowledge that, as white women, we don't experience the intersections of oppression in the same way that people of color and gender minorities, among others, do. But the silence of those who experience privilege in this space is the problem: this is in part why progress on diversity issues moves so slowly. It is important that those of us who do work in this space address these issues openly, and act to center the communities most affected.

To date, the diversity problems of the AI industry and the issues of bias in the systems it builds have tended to be considered separately. But we suggest that these are two versions of the same problem: issues of discrimination in the workforce and in system building are deeply intertwined. Moreover, tackling the challenges of bias within technical systems requires addressing workforce diversity, and vice versa. Our research suggests new ways of understanding the relationships between these complex problems, which can open up new pathways to redressing the current imbalances and harms.

From a high-level view, AI systems function as systems of discrimination: they are classification technologies that differentiate, rank, and categorize. But discrimination is not evenly distributed. A steady stream of examples in recent years have demonstrated a persistent problem of gender and race-based discrimination (among other attributes and forms of identity). Image recognition technologies miscategorize black faces,[11] sentencing algorithms discriminate against black defendants,[12] chatbots easily adopt racist and misogynistic language when trained on online discourse,[13] and Uber's facial recognition doesn't work for trans drivers.[14] In most cases, such bias mirrors and replicates existing structures of inequality in society.

In the face of growing evidence, the AI research community, and the industry producing AI products, has begun addressing the problem of bias by building on a body of work on fairness, accountability, and transparency. This work has commonly focused on adjusting AI systems in ways that produce a result deemed "fair" by one of various mathematical definitions.[15] Alongside this, we see growing calls for ethics in AI, corporate ethics boards, and a push for more ethical AI development practices.[16]

But as the focus on AI bias and ethics grows, the scope of inquiry should expand to consider not only how AI tools can be biased technically, but how they are shaped by the environments in which they are built and the people that build them. By integrating these concerns, we can develop a more accurate understanding of how AI can be developed and employed in ways that are fair and just, and how we might be able to ensure both.

Currently, large scale AI systems are developed almost exclusively in a handful of technology companies and a small set of elite university laboratories, spaces that in the West tend to be extremely white, affluent, technically oriented, and male.[17] These are also spaces that have a history of problems of discrimination, exclusion, and sexual harassment. As Melinda Gates describes, "men who demean, degrade or disrespect women have been able to operate with such impunity—not just in Hollywood, but in tech, venture capital, and other spaces where

---

11 Alcine, J. (2015). Twitter. Retrieved from https://twitter.com/jackyalcine/status/615329515909156865.

12 Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016, May 3). Machine Bias. *ProPublica*, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

13 Vincent, J. (2016, Mar 24). Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. *The Verge*. Retrieved from https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

14 Melendez, S. (2018, Aug. 9). Uber driver troubles raise concerns about transgender face recognition. *Fast Company*, Retrieved from https://www.fastcompany.com/90216258/uber-face-recognition-tool-has-locked-out-some-transgender-drivers.

15 Narayanan, A. (2018). 21 fairness definitions and their politics. ACM Conference on Fairness, Accountability and Transparency. Retrieved from https://www.youtube.com/watch?v=jIXIuYdnyyk.

16 Vincent, J. (2019, Apr. 3). The Problem with AI Ethics. *The Verge*. Retrieved from https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech.

17 Crawford, K. (2016, June 25). Artificial Intelligence's White Guy Problem. *The New York Times*. Retrieved from https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html.

their influence and investment can make or break a career. The asymmetry of power is ripe for abuse."[18] Or as machine learning researcher Stephen Merity noted at the end of 2017, "Bias is not just in our datasets, it's in our conferences and community."[19]

Both within the spaces where AI is being created, and in the logic of how AI systems are designed, the costs of bias, harassment, and discrimination are borne by the same people: gender minorities, people of color, and other under-represented groups. Similarly, the benefits of such systems, from profit to efficiency, accrue primarily to those already in positions of power, who again tend to be white, educated, and male. This is much more than an issue of one or two bad actors: it points to a systematic relationship between patterns of exclusion within the field of AI and the industry driving its production on the one hand, and the biases that manifest in the logics and application of AI technologies on the other.

Addressing these complexities will take much more than the technically-driven problem solving that has thus far dominated the discussion of gender and race in AI. Our research points to the need for a more careful analysis of the ways in which AI constructs and amplifies systems of classification, which themselves often support and naturalize existing power structures,[20] along with an examination of how these systems are being integrated into our institutions, and how they may be experienced differently on the basis of one's identity. Such research requires looking at gender and race as categories "within which humans think about and organize their social activity, rather than a natural consequence of difference."[21] In short, in studies of discriminatory systems we need to ask: who is harmed? Who benefits? Who gets to decide?

It is critical that we not only seek to understand how AI disadvantages some, but that we also consider how it works to the advantage of others, reinforcing a narrow idea of the 'normal' person.[22] By tracing the way in which race, gender, and other identities are understood, represented, and reflected, both within AI systems, and in the contexts where they are applied, we can begin to see the bigger picture: one that acknowledges power relationships, and centers equity and justice.[23]

---

18 Kolhatkar, S. (2017). The Tech Industry's Gender Discrimination Problem. https://www.newyorker.com/magazine/2017/11/20/the-tech-industrys-gender-discrimination-problem.

19 Merity, S. (2017). Bias is not just in our datasets, it's in our conferences and community. *Smerity.com*. https://smerity.com/articles/2017/bias_not_just_in_datasets.html.

20 Bowker, G.C. and Star, S.L. (1999). *Sorting Things Out: Classification and its Consequences*. Cambridge: MIT Press.

21 Harding, S. (1986). *The Science Question in Feminism*. Ithaca: Cornell University Press, p. 17

22 While race and gender are key axes of identity, and are most commonly considered in discussions of AI bias, it is important to emphasize that they are far from the only identity categories that shape AI systems. For example, as the work of Virginia Eubanks makes clear, class-based discrimination is a particularly thorny challenge, highlighting the ways in which AI systems are entwined with surveillance of the poor. See: Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Punish and Police the Poor*. London: St. Martin's Press. In addition, in partnership with the NYU Center for Disability Studies and Microsoft, AI Now recently hosted a one-day workshop on Disability and Bias in AI. We will be releasing a report summarizing our discussion and examining the ways in which disability studies expand and complicate our notions of AI bias. An examination of disability in the context of AI bias is particularly productive in that it requires us to scrutinize what (and who) constitutes a "normal" body, how aberrance and normalcy are themselves defined (and by whom), how such normative classifications may be mapped onto bodies in different ways at different times throughout an individual's lifetime, and what the consequences of such classifications may be.

23 For thoughtful treatments of what a justice-oriented data science might look like, and how it differs from data ethics, see: Green, B. (2018). Data Science as Political Action: Grounding Data Science in a Politics of Justice, Retrieved from https://scholar.harvard.edu/files/bgreen/files/data_science_as_political_action.pdf, and Klein, L. and D'Ignazio, C. (2019). *Data Feminism*. Cambridge: MIT Press. Retrieved from https://bookbook.pubpub.org/pub/dgv16l22.

WHICH HUMANS ARE IN THE LOOP? HOW WORKFORCES AND AI SYSTEMS INTERACT

To understand the full scope of the diversity crisis, we must do more than ask whether humans are in the loop - a phrase commonly used in the AI community to refer to AI systems that operate under the guidance of human decision makers - but which humans are in the loop.

A growing body of research is highlighting the ways that AI systems can cause harm to under-represented groups and those with less power.[24] Anna Lauren Hoffmann describes this as data violence:[25] data science that enacts forms of administrative violence that disproportionately affect some of us more than others.[26]

A recent example illustrates this discriminatory system at work. In 2018, Reuters reported that Amazon had developed an experimental hiring tool to help rank job candidates. By learning from its past preferences, Amazon hoped that the resume scanning tool would be able to efficiently identify qualified applicants by comparing their applications to previous hires. The system quickly began to downgrade resumes from candidates who attended all-women's colleges, along with any resumes that included the word "women's". After uncovering this bias, Amazon engineers tried to fix the problem by directing the system to treat these terms in a "neutral" manner. The company eventually abandoned the tool when they were unable to ensure that the algorithm would not be biased against women.[27] Gender-based discrimination was built too deeply within the system – and in Amazon's past hiring practices - to be uprooted using a purely technical approach.

The Amazon resume scanning example is just one of many that show how the functional logics of a given technology echo the gender and racial dynamics of the industry that produced it.[28] Amazon's Rekognition facial analysis service previously demonstrated gender and racial biases worse than those of comparable tools, biases that took the form of literally failing to "see" dark-skinned women while being most proficient at detecting light-skinned men.[29] Amazon's initial response to such criticism has been to try and discredit the research behind it.[30] This reaction is evidence of the wider problem: the research was conducted by two well-regarded AI researchers who are women of color. By attempting to publicly discredit their expertise and research methods, Amazon is reinforcing the same kinds of prejudice and erasure that the research critiques.

24 Neff, G. (2018) Does AI Have Gender? *OII London Lecture*. Retrieved from https://www.oii.ox.ac.uk/events/oii-neff-lecture/.

25 Hoffmann, A.L. (2018, Apr. 30). Data Violence and How Bad Engineering Choices can Damage Society. *Medium*. Retrieved from https://medium.com/s/story/data-violence-and-how-bad-engineering-choices-can-damage-society-39e44150e1d4.

26 Keyes, O. (2019, Mar 24). Counting the Countless. Retrieved from https://ironholds.org/counting-writeup/.

27 Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, Retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/.

28 For a deeper dive into gender-based data bias, see: Perez, C.C. (2019). *Invisible Women: Data Bias in a World Designed for Men*. New York: Abrams Press.

29 Raji, I & Buolamwini, J. (2019). Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. *Conference on Artificial Intelligence, Ethics, and Society*. Retrieved from https://www.media.mit.edu/projects/actionable-auditing-coordinated-bias-disclosure-study/publications/.

30 Buolamwini, J. (2019, Jan. 25). Response: Racial and Gender bias in Amazon Rekognition—Commercial AI System for Analyzing Faces. *Medium*. Retrieved from https://medium.com/@Joy.Buolamwini/response-racial-and-gender-bias-in-amazon-rekognition-commercial-ai-system-for-analyzing-faces-a289222eeced.

These problems are not inevitable, nor are they natural: history shows us that they are a product of the distribution of power in society.[31] For example, the work of historian Mar Hicks meticulously documents how structural discrimination shifted the gender makeup of Britain's computing industry, marginalizing the work of female technical experts by molding them into a technical underclass. As Hicks describes, "throughout history, it has often not been the content of the work but the identity of the worker performing it that determined its status".[32]

Examples of bias and discrimination in the workforce can be found across all the leading tech companies that are driving the development of AI technologies:

- A class action suit led by Microsoft workers alleges that the company systemically failed to take hundreds of allegations of harassment and discrimination seriously.[33]
- A federal investigation is underway into gender discrimination at Uber.[34]
- Apple dismissed concerns about its lack of workplace diversity as a 'solvable issue'[35] while simultaneously calling proposals for diverse hiring practices 'too burdensome'.[36]
- An audit of Google's pay practices by the Department of Labor found six to seven standard deviations between pay for men and women in nearly every job category.[37]
- Black employees at Facebook recount being aggressively treated by campus security and dissuaded by managers from taking part in internal Black@ group activities.[38]
- A lawsuit filed against Tesla alleges gender discrimination, retaliation, and a hostile work environment. One worker recounts that there were more men named "Matt" in her group than women.[39]

These examples suggest that inequity and bias are not to be found in a single place, like a bug that can be located and fixed. These issues are systemic. There is a close relationship between these workplaces with discriminatory practices and discriminatory tools: a feedback loop that is shaping the AI industry and its tools. The products of the AI industry already influence the lives of millions. Addressing diversity issues is therefore not just in the interest of the tech industry, but of everyone whose lives are affected by AI tools and services.

---

31 See, for example: Greenbaum, J. (1990). *Windows on the Workplace: Computers, Jobs, and the Organization of Office Work in the Late Twentieth Century*. New York: Monthly Review Press. Oldenziel, R. (1999) *Making Technology Masculine: Men, Women, and Modern Machines in America, 1870-1945*. Amsterdam: Amsterdam University Press.; Ensmenger, N. (2015). Beards, Sandals, and Other Signs of Rugged Individualism: Masculine Culture within the Computing Professions. *Osiris*, 30(1): 38-65.

32 Hicks, M. (2017). *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*. Cambridge: MIT Press, 16.

33 Microsoft Gender Case (2019, Apr. 12). Retrieved from https://microsoftgendercase.com/.

34 Bensinger, G. (2018, July 16). Uber Faces Federal Investigation Over Alleged Gender Discrimination. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/uber-faces-federal-investigation-over-alleged-gender-discrimination-1531753191?mod=breakingnews.

35 Goldman, D. (2015, June 8). Tim Cook: You'll soon see more women representing Apple. *CNN*. Retrieved from https://money.cnn.com/2015/06/08/technology/tim-cook-women-apple/?iid=EL.

36 O'Brien, S.A. (2016, Jan. 15). Apple's board calls diversity proposal 'unduly burdensome and not necessary'. *CNN*. Retrieved from https://money.cnn.com/2016/01/15/technology/apple-diversity/index.html.

37 Kolhatkar, S. (2017, Nov. 13). The Tech Industry's Gender-Discrimination Problem. *The New Yorker*. Retrieved from https://www.newyorker.com/magazine/2017/11/20/the-tech-industrys-gender-discrimination-problem.

38 Luckie, M. (2018, Nov. 27). Facebook is failing its black employees and its black users. *Facebook*. https://www.facebook.com/notes/mark-s-luckie/facebook-is-failing-its-black-employees-and-its-black-users/1931075116975013/.

39 Kolhatkar, S. (2017, Nov. 13). The Tech Industry's Gender-Discrimination Problem. *The New Yorker*. Retrieved from https://www.newyorker.com/magazine/2017/11/20/the-tech-industrys-gender-discrimination-problem.

From this perspective, locating individual biases within a given technical system–and attempting to fix them by tweaking the system–becomes an exercise in futility. Only by examining discrimination through the lens of its social logics (who it benefits, who it harms, and how) can we see the workings of these systems in the context of existing power relationships.

In addition to asking when and how AI systems favor some identities over others we might also ask: what is the logic through which artificial intelligence "sees" and constructs gender and race to begin with? How does it engage in the production and enactment of new classifications and identities?[40] And how do AI systems replicate historical hierarchies by rendering people along a continuum of least to most "valuable"?

These questions point to the larger problem: it is not just that AI systems need to be fixed when they misrecognize faces or amplify stereotypes. It is that they can perpetuate existing forms of structural inequality even when working as intended.

To tackle these questions, our research traces the way gender and race surfaces in AI systems and workforces, and their interrelationship. First, we review what is known and not known about diversity in the field of AI, focusing particularly on how frames devoted to the STEM field 'pipeline' have dominated the discourse. Then, we provide a brief summary of existing literature on gender and racial bias in technologies and where this literature could be extended. Finally, we look at how calls for diversity in tech have been ignored or resisted, and how these discriminatory views have permeated many AI systems. We conclude by sharing new research findings that point to ways in which a deeper analysis of gender, race, and power in the field of AI can help to redress inequalities in the industry and in the tools it produces.

WHO MAKES AI?

The current data on the state of gender diversity in the AI field is dire, in both industry and academia. For example, in 2013, the share of women in computing dropped to 26%, below their level in 1960.[41] Almost half the women who go into technology eventually leave the field, more than double the percentage of men who depart.[42] As noted above, a report produced by the research firm Element AI found that only 18% of authors at the leading 21 conferences in the field are women,[43] while the 2018 Artificial Intelligence Index reports 80% of AI professors are men.[44] This imbalance is replicated at large tech firms like Facebook and Google, whose websites show

40 Kloppenburg, S. and van der Ploeg, I. (2018). *Securing Identities: Biometric Technologies and the Enactment of Human Bodily Differences*. Science as Culture.

41 Thompson, C. (2019, Feb. 13). The Secret History of Women in Coding. *New York Times Magazine*. Retrieved from https://www.nytimes.com/2019/02/13/magazine/women-coding-computer-programming.html?linkId=65692573.

42 Ashcraft, C., McLain, B. and Eger, E. (2016). Women in Tech: The Facts. National Center for Women in Information Technology. Retrieved from https://www.ncwit.org/sites/default/files/resources/womenintech_facts_fullreport_05132016.pdf.

43 Element AI. (2019). Global AI Talent Report 2019. Retrieved from https://jfgagne.ai/talent-2019/.

44 AI Index 2018. (2018). Artificial Intelligence Index 2018. Retrieved from http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf.

even greater imbalances, with women comprising only 15% and 10% of their AI research staff, respectively.[45,46] There is no reported data on trans workers or other gender minorities.

The state of racial diversity in AI is even worse. Only 2.5% of Google's full-time workers are black, and 3.6% latinx, with black workers having the highest attrition rate of all racial categories.[47] Facebook isn't much better: the company reported that with 4% black workers and 5% 'Hispanic' workers in 2018, the company's diversity is improving.[48] Microsoft reflects similar levels as Facebook, with 4% black workers, and 6% Latinx workers.[49] Machine vision researcher and co-founder of Black in AI, Timnit Gebru, said that when she first attended the preeminent machine learning conference NeurIPS in 2016, she was one of six black people – out of 8,500 attendees.[50] "We are in a diversity crisis for AI," Gebru explains. "In addition to having technical conversations, conversations about law, conversations about ethics, we need to have conversations about diversity in AI. This needs to be treated as something that's extremely urgent."[51]

Of course, artificial intelligence is a sub-field of computer science, and the broader discipline is experiencing an historic low point for diversity: as of 2015, women made up only 18% of computer science majors in the United States, a decline from a high of 37% in 1984.[52] No other professional field has experienced such a sharp decline in the number of women in its ranks.[53] At present, women currently make up 24.4% of the computer science workforce, and receive median salaries that are only 66% of the salaries of their male counterparts. These figures are similarly pronounced when race is taken into account; the proportion of bachelor's degree awards in engineering to black women declined 11% between 2000 and 2015.[54] The number of women and people of color decreased at the same time that the tech industry was establishing itself as a nexus of wealth and power. This is even more significant when we recognize that these shocking diversity figures are not reflective of STEM as a whole: in fields outside of computer science and AI, racial and gender diversity has shown a marked improvement.[55]

---

45 Simonite, T. (2018). AI is the future - but where are the women? *WIRED*. Retrieved from https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/.

46 The World Economic Forum's 2018 Global Gender Gap Report includes a section on diversity in AI that places its estimate much higher at 22%. However, the methodology for obtaining this figure raises some questions: it relies on LinkedIn users' inclusion of AI-related skills in their profiles as the primary data source. This requires several causal leaps: first, that a sample of LinkedIn users is representative of the global population of workers in the field of AI, and that these users accurately represented their skill set. Moreover, the study used a flawed mechanism to attribute gender on a binary basis to users on the basis of inference from their first name – a practice that is not only trans-exclusionary, but is particularly problematic in an analysis that includes names in non-English languages.

47 Google. (2018). Google Diversity Annual Report 2018. Retrieved from https://static.googleusercontent.com/media/diversity.google/en//static/pdf/Google_Diversity_annual_report_2018.pdf.

48 Williams, M. (2018, July 12). Facebook 2018 Diversity Report: Reflecting on Our Journey. Retrieved from https://newsroom.fb.com/news/2018/07/diversity-report/.

49 Microsoft. (2019). Diversity & Inclusion. Retrieved from https://www.microsoft.com/en-us/diversity/default.aspx.

50 Snow, J. (2018). "We're in a diversity crisis": cofounder of Black in AI on what's poisoning algorithms in our lives. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/s/610192/were-in-a-diversity-crisis-black-in-ais-founder-on-whats-poisoning-the-algorithms-in-our/.

51 Ibid.

52 National Academies of Sciences, Engineering, and Medicine. (2018). *Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine*. Washington, DC: The National Academies Press.

53 Misa, T. (2006). Gender Codes: Defining the Problem, in Misa, T. (Ed.) *Gender Codes: Why Women Are Leaving Computing*. Hoboken: IEEE Computer Society.

54 National Academies of Sciences, Engineering, and Medicine. (2018). *Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine*. Washington, DC: The National Academies Press.

55 Hayes, C.C. (2006). Computer Science: The Incredible Shrinking Woman, in Misa, T. (Ed.) *Gender Codes: Why Women Are Leaving Computing*. Hoboken: IEEE Computer Society.

Collectively, these statistics paint a grim picture. As Freda Kapor Klein has described it, "It's sobering to see the lack of progress...We have a problem, and we need to work together to solve it".[56]

Diversity Statistics in the AI Industry: Knowns and Unknowns

But the existing data on the state of diversity has real limitations. Over the past decade, the AI field has shifted from a primarily academic setting to a field increasingly situated in corporate tech environments. But it is simply harder to gain a clear view of diversity and decision making within the large technology firms that dominate the AI space due to the ways in which they tightly control and shape their hiring data. This is a significant barrier to research.

In making inferences based on company diversity reports, we are thus reliant on a specifically curated view of corporate diversity. For years, technology firms resisted releasing any diversity figures at all, engaging in legal battles with reporters and the Department of Labor to prevent access to their employment data.[57] In 2014, one year after a call by software engineer Tracy Chou to release the data received widespread public attention, Apple, Facebook, and Google released their first diversity reports, indicating that women and people of color were indeed systematically under-represented in all three companies, particularly in technical and leadership roles.[58,59]

Though these statistics provide some insight, they deserve close scrutiny. For one, statistics can only tell a part of the overall story: they do not account for the day to day experiences of workers within these companies, the structural factors that may be shaping whether or not these workers can succeed, and what work is and is not incentivized.[60] In other words, the picture may be even worse than the statistics reveal.

For example, the 2019 email thread by women at Microsoft exposed how dozens of women were repeatedly passed over for promotion, side-lined, or harassed. They reported being threatened unless they performed sexual acts, demeaned during meetings, and being dismissed by HR when making claims about unfair treatment.[61] Further, a 2018 class action suit brought by women in technical roles at Microsoft alleges the company handled complaints of harassment and discrimination in a 'lackluster' way, fostering a 'boys' club atmosphere' and forcing a female intern to work alongside a man who she alleged raped her, even after reporting the assault to the police, her supervisor, and HR. After investigating over 100 complaints of gender discrimination,

---

56 Guynn, J. (2018, Feb. 28). Tech industry's diversity efforts haven't lived up to promises. A new report explains why. *USA Today*. Retrieved from https://www.usatoday.com/story/tech/2018/02/28/diversity-freada-kapor-klein-kapor-center-report-leaky-pipeline/378295002/.

57 Pepitone, J. (2013). Black, female, and a Silicon Valley 'trade secret'. *CNN Business*. Retrieved from https://money.cnn.com/2013/03/17/technology/diversity-silicon-valley/index.html.

58 Chou, T. (2013, Oct. 11). Where are the numbers? *Medium*. Retrieved from https://medium.com/@triketora/where-are-the-numbers-cb997a57252.

59 Gutman, R. (2018, Feb. 3). The Origins of Diversity Data in Tech. *The Atlantic*. Retrieved from https://www.theatlantic.com/technology/archive/2018/02/the-origins-of-diversity-data-in-tech/552155/.

60 The Elephant in the Valley project has played a critical role in filling in these gaps, collecting anonymous stories from workers who have experienced discrimination and harassment in the workplace. For more see: https://www.elephantinthevalley.com/.

61 Gershgorn, D. (2019, Apr. 4). Amid employee uproar, Microsoft is investigating sexual harassment claims overlooked by HR. *Quartz*. Retrieved from https://qz.com/1587477/microsoft-investigating-sexual-harassment-claims-overlooked-by-hr/.

the company concluded only one was 'founded'.[62,63] The existence of these employees would certainly be accounted for in corporate diversity and inclusion statistics, but their experiences tell a radically different story about what it means to be included. These are precisely the accounts that most need to be listened to. As Kristian Lum put it in her own account of harassment and discrimination at the machine learning conference NeurIPs, "It is time for us to be publicly and openly appalled, not just attempting to tactfully deflect inappropriate advances and privately warning other women."[64]

While providing much needed insight, the diversity and inclusion data AI companies release to the public is a partial view, and often contains flaws. Former defense lawyer and gender diversity advocate Pat Gillette told reporters from the Center for Investigative Reporting that corporate diversity reports are easier to manipulate than the EEO-1 forms they are mandated to provide to the government, which break down companies' employees by race, gender, and job category. Companies rarely make these forms available to the public, but publishing their own reports gives them more leeway to massage the numbers based on how they define terms like race, gender, and role, Gillette said.[65] One researcher found that Google's diversity report was designed to artificially inflate the numbers of women and people of color employed by the company by only accounting for 80% of the company's full-time workforce.[66]

The data presented in these reports is also limited in scope, and historically excluded figures that would provide key insights into gender and race-based discrimination in tech companies. For example, analysis of data from the 2010-12 American Community Survey by the American Institute for Economic Research found that there are substantial pay disparities among high tech workers: on average, female software developers of color earn less than white, black, and Asian men, as well as white women. Latina software developers earned as much as 20% less annually than white male software developers.[67] Disparities in equity and ownership are even worse: an analysis of over 6,000 companies found that women hold only 9% of startup equity value, blocking them from streams of compensation that are often of greater worth than tech workers' annual salaries.[68]

The ways in which such data can be massaged to reflect company positions was on display in a 2019 claim by Google regarding its gender pay gap. In studying ways to remedy gender inequities, the company found that more men in junior engineering roles were underpaid than women.[69] This "counterintuitive" finding was widely reported, accompanied by questions about whether the gendered wage gap was, in fact, a problem.[70] On close examination, however, the claim being

62 Katherine Moussouris, Holly Muenchow and Dana Piermarini v. Microsoft Corporation. (2018). Retrieved from https://www.courtlistener.com/recap/gov.uscourts.wawd.220713/gov.uscourts.wawd.220713.381.0.pdf.

63 Solon, O. (2018, Mar. 13). Lawsuit claims sexual harassment rife in Microsoft's 'boys' club atmosphere'. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2018/mar/13/microsoft-sexual-harassment-lawsuit-lacklustre-response.

64 Lum, K. (2017, Dec. 14). Statistics, we have a problem. *Medium*, Retrieved from https://medium.com/@kristianlum/statistics-we-have-a-problem-304638dc5de5.

65 Evans, W. and Rangarajan, S. (2017, Oct. 19). Hidden figures: How Silicon Valley keeps diversity data secret. *Center for Investigative Reporting*. Retrieved from https://www.revealnews.org/article/hidden-figures-how-silicon-valley-keeps-diversity-data-secret/.

66 Lee, L. (2018, Mar. 29). Life as a Female Techie. *Unicorn Techie*. Retrieved from http://unicorntechie.com/.

67 American Institute for Economic Research. (2014, Oct. 9) H-1B Visas: No Impact on Wages. Retrieved from https://www.aier.org/research/h-1b-visas-no-impact-wages.

68 Kramer, E. (2018, Sept. 17). The Gap Table: Analyzing the gender gap in equity. *Carta*. Retrieved from https://carta.com/blog/gap-table/.

69 Barbato, L. (2019). Ensuring we pay fairly. *Google*. Retrieved from https://www.blog.google/inside-google/company-announcements/ensuring-we-pay-fairly-and-equitably/.

70 Wakabayashi, D. (2019, Mar. 4). Google Finds It's Underpaying Many Men as It Addresses Wage Equity. *New York Times*. Retrieved from https://www.nytimes.com/2019/03/04/technology/google-gender-pay-gap.html.

made is extremely narrow, focusing on one level of one job category, and not taking into account equity and bonus, which at senior levels of the company often comprise the majority of employee compensation.[71] It is also significant that Google made this information publicly available in early 2019. This comes at a time when Google is being investigated by the US Department of Labor, facing a lawsuit by women employees, and is still grappling with the ramifications of the protest where 20,000 of its workers walked out to protest discrimination, sexual harassment, and a hostile workplace culture.[72,73]

In the past, when the US Department of Labor sought to look into allegations that Google systematically underpaid women, the company reported it would be "financially burdensome and logistically challenging" to provide its salary records to the government. It released the company's internal pay equity analyses for the first time in 2016, asserting that there was no statistically significant difference between the compensation received by men and women at the company.[74] However, the report's methodology noted that 11% of the company's employees were left off of the analysis because the company limited its findings to job categories with 30 or more employees with at least five men and five women, thus excluding all employees with a rank above vice president. The company's highest-paying jobs, most of which are held by men, were not included in the report.[75]

Importantly, neither the report nor the 2019 findings accounted for the phenomenon of "underleveling," in which women and people of color are hired in junior roles even when they have the skills to perform in more senior (and better compensated) positions. An analysis of data provided by 177 leading tech companies to the Equal Employment Opportunity Commission revealed a systematic pattern of underleveling across the technology industry. Nearly a third of the firms included in the data had no executives who are women of color. Six had no female executives at all.[76,77]

Such phenomena aren't new. Margaret Rossiter documents similar problems for women entering scientific fields in the early 20th century: while women had gained some access to scientific training and jobs in these fields, they struggled to obtain equality once they were in the door, faced with "a pattern of segregated employment and under-recognition" that "limited [them] to positions just inside the entryway."[78]

---

71 Tiku, N. (2019, Mar. 4). Are men at Google paid less than women? Not really. *WIRED*. Retrieved from https://www.wired.com/story/men-google-paid-less-than-women-not-really/.

72 Tiku, N. (2017, Sept. 14). Bias suit could boost pay, open promotions for women at Google. *WIRED*. Retrieved from https://www.wired.com/story/bias-suit-could-boost-pay-open-promotions-for-women-at-google/.

73 Google Walkout For Real Change. (2018, Nov. 2). Google employees and contractors participate in global "walkout for real change". *Medium*. Retrieved from https://medium.com/@GoogleWalkout/google-employees-and-contractors-participate-in-global-walkout-for-real-change-389c65517843.

74 Naughton, E. (2017, Apr. 11). Our focus on pay equity. *Google*. Retrieved from https://www.blog.google/outreach-initiatives/diversity/our-focus-pay-equity/.

75 Colby, L. and Huet, E. (2018, Mar. 15). Google's Equal-Pay Claim Comes With an Asterisk. *Bloomberg*. Retrieved from https://www.bloomberg.com/news/articles/2018-03-15/google-s-equal-pay-claim-for-women-comes-with-an-asterisk.

76 Rangarajan, S. (2018, June 25). Here's the clearest picture of Silicon Valley's diversity yet: It's bad. But some companies are doing less bad. *The Center for Investigative Reporting*. Retrieved from https://www.revealnews.org/article/heres-the-clearest-picture-of-silicon-valleys-diversity-yet/.

77 See also: Gee, B., Peck, D., Wong, J. (2013). Hidden in Plain Sight: Asian American Leaders in Silicon Valley. *The Ascend Foundation*. Retrieved from https://c.ymcdn.com/sites/ascendleadership.site-ym.com/resource/resmgr/Research/HiddenInPlainSight_Paper_042.pdf.

78 Rossiter, M. (1986). Women Scientists in America: Struggles and Strategies to 1940, in Harding, S. (1986). *The Science Question in Feminism*. Ithaca: Cornell University Press, p. 60-62.

Many tech companies do publicly advocate for diversity, have corporate diversity officers, and fund initiatives to encourage more young girls to take up coding (although this excludes others who experience identity-based discrimination, such as the trans community). However, there are multiple examples where Silicon Valley's largest tech firms have privately hindered efforts by employees to advocate for diversity within the company. One recent report highlighted that employees at Google have expressed concerns that they will face consequences for voicing support for diversity initiatives[79] - concerns reinforced by studies that show that women are frequently penalized for advocating for diversity in the workplace.[80]

FROM WORKFORCES TO AI SYSTEMS: THE DISCRIMINATION FEEDBACK LOOP

Discrimination and inequity in the workplace have significant material consequences, particularly for the under-represented groups who are excluded from resources and opportunities. For this reason alone the diversity crisis in the AI sector needs to be urgently addressed. But in the case of AI, the stakes are higher: these patterns of discrimination and exclusion reverberate well beyond the workplace into the wider world. Industrial AI systems are increasingly playing a role in our social and political institutions, including in education, healthcare, hiring, and criminal justice. Therefore, we need to consider the relationship between the workplace diversity crisis and the problems with bias and discrimination in AI systems.

Fairness, accountability, and transparency research is playing an emerging role in documenting the scale and scope of gendered and racialized discrimination in AI systems. For example, a recent study found that mechanisms in Facebook's ad delivery systems led users to be shown ads for housing and employment in a discriminatory manner. With the same targeted audience, and without the advertisers intending or being aware, ads are delivered in a manner that aligns with gender and racial stereotypes: ads for jobs in the lumber industry were disproportionately shown to white male users, while ads for cashier positions at supermarkets were shown to female users and ads for taxi drivers to black users.[81] By experimenting with Google's search engine results, Safiya Noble demonstrated that Google search results retrieve highly sexualized imagery for searches on terms like "black girls" and "latina girls".[82] In a landmark study, Latanya Sweeney found that two search engines disproportionately serve ads for arrest records against searches for racially associated names.[83]

A 2019 study found significant racial bias in a widely used commercial algorithm used to determine whether patients will be enrolled in 'care management' programs that allocate considerable additional resources: white patients were far more likely to be enrolled in the

79 Conger, K. (2018, Feb. 21). Google Fired and Disciplined Employees for Speaking Out About Diversity. *Gizmodo*. Retrieved from https://gizmodo.com/google-fired-and-disciplined-employees-for-speaking-out-1822277125.

80 Hekman, D.R., Johnson, S.K., Foo, M. and Yang, W. (2016). Does Diversity-Valuing Behavior Result in Diminished Performance Ratings for Non-White and Female Leaders? *Academy of Management Journal*, 60(2).

81 Ali, M., Sapiezynski, P, Bogen, M, Korolova, A., Mislove, A., and Rieke, A. (2019). Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes. Retrieved from https://arxiv.org/pdf/1904.02095.pdf.

82 Noble, S.U. (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

83 Sweeney, L. (2013). Discrimination in Online Ad Delivery. arXiv. Retrieved May 20, 2018 from https://arxiv.org/abs/1301.6822.

program and to benefit from its resources than black patients in a comparable state of health.[84] The forensic examination of individual systems for bias and discrimination is an important area of research, and more studies like these are sorely needed.

There are also other aspects of discrimination by AI systems that need further attention. First, studies that focus on bias often adopt a definition of the term that can easily be operationalized technically, such as looking for a biased distribution of error rates on the basis of a single variable.[85] But we can benefit from broadening the way we think about the concept of bias: there are many forms of discrimination that might emerge from AI systems that would not fit within such definitions.[86] Harms are frequently defined in terms of economic disadvantage, with less attention paid to harms of how people are represented or interpreted by AI systems, and the downstream social and political consequences of such representation.[87] And, as Ben Hutchinson and Margaret Mitchell argue in a recent article, the political will to use scientific contributions in this policy arena may suffer if technical definitions of fairness stray too far from the public's perceptions.[88]

Taking a contextualized view may enable a more extensive account of bias to emerge. Future work could examine the politics of system design itself, and study AI systems in situated realities. Such work could ask why a system was designed in a particular way, how it was constructed, and whose interests shaped the metrics by which its success or failure is assessed. Rather than solely focusing on improving existing datasets or individual algorithms, future work could also more thoroughly account for how societal discrimination surfaces in data provenance, examining the history and process of dataset construction, and considering how cultural norms and stereotypes were numerated and represented at the time of data creation.[89]

For example, according to Han and Jain, while the popular Labeled Faces in the Wild (LFW) dataset contains over 15,000 images of faces, only 7% are images of black people.[90] When we examine the root of such unequal representation, we are led to the media landscape of the early 2000s from which these images were gleaned. The news media at the time predominantly featured white men in positions of celebrity and power. Drawing from this source, LFW's representation of "human faces" can be understood as a reflection of early 2000s social hierarchy, as reproduced through visual media. Similarly, Mishra and Srikumar argue that datasets in India

---

84 Obermeyer, Z. and Mullainathan, S. (2019). Dissecting Racial Bias in an Algorithm that Guides Health Decisions for 70 million people. FAT* '19: Conference on Fairness, Accountability, and Transparency, January 29–31, 2019, Atlanta, GA, USA Retrieved from https://dl.acm.org/citation.cfm?id=3287593.

85 Costanza-Chock, S. (2018, Jul. 27). Design Justice, A.I. and Escape from the Matrix of Domination. *Journal of Design and Science*. Retrieved from https://jods.mitpress.mit.edu/pub/costanza-chock.

86 Hutson, J. et al. (2018). Debiasing Desire: Addressing Bias & Discrimination on Intimate Platforms. Retrieved from https://arxiv.org/pdf/1809.01563.pdf.

87 Barocas, S., Crawford, K., Shapiro, A. and Wallach, H. 2017 'The Problem With Bias: Allocative Versus Representational Harms in Machine Learning', *SIGCIS Conference*, http://meetings.sigcis.org/uploads/6/3/6/8/6368912/program.pdf, Hutchinson, B. and Mitchell, M. (2019). 50 Years of Test (Un)fairness: Lessons for Machine Learning. FAT* '19: Conference on Fairness, Accountability, and Transparency, January 29–31, 2019, Atlanta, GA, USA. Retrieved from https://arxiv.org/abs/1811.10104.

88 Hutchinson, B. and Mitchell, M. (2019). 50 Years of Test (Un)fairness: Lessons for Machine Learning. FAT* '19: Conference on Fairness, Accountability, and Transparency, January 29–31, 2019, Atlanta, GA, USA. Retrieved from https://arxiv.org/abs/1811.10104.

89 Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumee III, H., and Crawford, K. (2018). Datasheets for Datasets. Retrieved from https://arxiv.org/abs/1803.09010. See also: Singh, J., Cobbe, J. and Norval, C. (2018). Decision Provenance: Harnessing Data Flow for Accountable Systems, *IEEE Access*. Retrieved from https://arxiv.org/abs/1804.05741 and Passi, S. and Barocas, S. (2019). Problem Formulation and Fairness. *Conference on Fairness, Accountability and Transparency 2019*. Retrieved from https://arxiv.org/abs/1901.02547.

90 Han, H. and Jain, A.K. (2014). Age, Gender and Race Estimation from Unconstrained Face Images. MSU Technical Report. Retrieved from http://biometrics.cse.msu.edu/Publications/Face/HanJain_UnconstrainedAgeGenderRaceEstimation_MSUTechReport2014.pdf.

are especially likely to suffer from a lack of representation and to replicate masculine hegemonic norms due to limited access to technology among women and the poor, leading them to be excluded from such datasets, and thereby unaccounted for in AI systems.[91] A recent paper from the AI Now Institute examines the data used in algorithmic predictive policing systems, and finds that in many cases such data is fraudulent, created through practices of racially biased law enforcement, thus embedding bias into the logics of such systems.[92] In all of these cases, understanding "bias" in data (and arguably fixing such bias) requires a thorough accounting of the social context through which the data was produced - in other words, how humans make data in context.

Second, many studies on bias and discrimination operate on a single axis rather than examining the intersections of multiple identity categories. This is likely to produce what Erica Joy Baker, a Senior Engineering Manager at Patreon, calls colorless diversity: without acknowledging the ways in which different forms of oppression intersect, diversity efforts that target women without acknowledging the role of race and other forms of identity (let alone the broader spectrum of gendered identity) will implicitly privilege white women.[93]

An example of a more intersectional study is the 2018 Gender Shades paper by Joy Buolamwini and Timnit Gebru, which looked at three commercial facial recognition systems that include the ability to classify faces by gender and found that they tend to exhibit higher error rates for darker-skinned women than for any other group, with the lowest error rates for light skinned men.[94] Such gender and racial bias has been attributed to the composition of the datasets used to train these systems, which, like Labeled Faces in the Wild, were overwhelmingly composed of lighter-skinned male-looking subjects. To measure this disparity, Buolamwini and Gebru developed a new dataset that is more balanced both in terms of gender and skin color.

While studies such as theirs clearly illustrate the ways in which AI systems can reflect existing patterns of gender and racial bias, it's important to emphasize that the problems exposed in this study are not the only issues that exist in these systems. Irrespective of their accuracy, the very existence of automated gender classification systems presents a number of problems: they functionally understand gender as an essential, biological, and binary identity that can be "detected" and affirmed through the lens of a commercialized technical system.[95] In this way, such systems (and the interests that create and profit from them) are positioned as the arbiters of identity, mapping static categories onto diverse bodies.

Third, many of the studies examining bias within AI systems adopt a binary view of gender. In a review of existing research in this space we found a handful of studies that addressed trans or non-binary gender identity, or even acknowledged genders beyond "male" and "female".

---

91 Mishra, V. and Srikumar, M. (2017). Predatory Data: Gender Bias in Artificial Intelligence, in Saran, S. (Ed.) Digital Debates: CyFy Journal 2017. New Delhi: Observer Research Foundation.

92 Richardson, R., Schultz, J. and Crawford, K. (forthcoming). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. NYU Law Review Online. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3333423.

93 Baker, E.J. (2015). #FFFFFF Diversity. *Medium*. Retrieved from https://medium.com/this-is-hard/ffffff-diversity-1bd2b3421e8a.

94 Buolamwini, J. and Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research 81:1-15.

95 See: Keyes, O. (2018). The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. Proceedings of the ACM on Human-Computer Interaction - CSCW. Retrieved from https://dl.acm.org/citation.cfm?doid=3290265.3274357 and Hamid, F., Scheuerman, M.K. and Branham, S.M. (2018). Gender Recognition or Gender Reductionism? The Social Implications of Automatic Gender Recognition Systems. *CHI 2018*. Retrieved from https://docs.wixstatic.com/ugd/eb2cd9_ff211774807946099e7e1dcd0023497d.pdf.

Acknowledging the fluid nature of gender in studies of bias is not only necessary for accuracy, it will prevent use cases that will result in a systematic erasure of trans people and their experiences.[96]

Fourth, these studies often focus solely on the technical system, locating all problems within the system itself and without attending to the context in which it is applied along with the incentives that propel its application. This points to a significant dilemma for researchers of AI bias: though improving the performance of AI systems might be a necessary step toward making them more inclusive, there are some contexts in which 'fixing' such inaccuracies may not fix the overall problems presented by such systems - and some problems that cannot be fixed by a technical solution at all.[97]

To this point, a focus on fixing technical systems in isolation, without examining the broader context of their use and the power dynamics that attend such use, is not only limited in its intervention: it can actively cause harm. For example, the facial and image recognition technologies that are the focus of many studies of AI bias are often applied in service of police surveillance, which disproportionately harms poor people and communities of color.[98] This prompts the question of whose needs are served by ensuring these technologies 'work' for everyone. As Zoé Samudzi notes, "it is not social progress to make black people equally visible to software that will inevitably be further weaponized against us. We are considered criminal and more surveillable by orders of magnitude; whatever claim to a right to privacy that we may have is diminished by a state that believes that we must always be watched and seen."[99]

Asking this question is particularly important given that practices involved in correcting such biases sometimes lead those developing such technologies (most often large corporations) to conduct invasive data collection on communities that are already marginalized with the goal of ensuring that they're represented. For example, facial recognition systems often have a challenging time recognizing the faces of people undergoing gender transition. This error has been a problem for trans Uber drivers, because the facial recognition system built in as a security feature by Uber has led their accounts to be suspended, preventing them from being able to work while they seek to get their accounts restored.[100]

These harms should be balanced against remedies that rely on unethical practices, or that propose mass data collection as the solution to bias. One approach that received particular pushback collected videos from transgender YouTubers without their consent in order to train facial recognition software to more accurately recognize people undergoing the process of transitioning.[101] In this case, allowing alternate means of account verification may be a better "fix" than continuing to rely on a system whose efficacy demands increased surveillance and worker control.

---

96 Ibid.

97 Powles, J. (2018, Dec. 7). The Seductive Diversion of 'Solving' Bias in Artificial Intelligence. *Medium*. Retrieved from https://medium.com/s/story/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53.

98 Garvie, C., Bedoya, A. and Frankle, J. (2016). The Perpetual Lineup: Unregulated Police Face Regulation in America. *Georgetown Law Center on Privacy and Technology*. Retrieved from https://www.perpetuallineup.org/.

99 Samudzi, Z. (2019, Feb. 8). Bots Are Terrible at Recognizing Black Faces. Let's Keep it That Way. *The Daily Beast*. Retrieved from https://www.thedailybeast.com/bots-are-terrible-at-recognizing-black-faces-lets-keep-it-that-way.

100 Urbi, J. (2018, Aug. 8). Some transgender drivers are being kicked off Uber's app. *CNBC*. Retrieved from https://www.cnbc.com/2018/08/08/transgender-uber-driver-suspended-tech-oversight-facial-recognition.html.

101 Vincent, J. (2017, Aug. 22). Transgender YouTubers had their videos grabbed to train facial recognition software. *The Verge*. Retrieved from https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset.

A growing number of scholars and advocates argue that some biased systems are undeserving of being 'fixed' and may need to be removed or reconsidered altogether.[102] Repeated efforts to develop machine learning methods to detect sexual orientation, for example, are illustrative of why 'diversifying' the dataset is not a de facto solution and may in fact exacerbate the problem by legitimizing harmful technologies.[103] Some systems should not be built at all.

Making AI systems more cognizant of race, gender, and other identity categories is often presented as a sign that those producing such systems embrace diversity and inclusion. However, such an expansion of classification also enables the monetization of identity as "market segments" for corporate profit. This can create new harms while driving ad tech. For example, Jacob Gaboury argues that by expanding the gender categories available in Facebook profiles from two to 58 options, Facebook is not just embracing a more inclusive view of gender, it is creating flattened and instrumentalized identity categories that can be used for the purpose of value extraction through targeted ads.[104]

A study by Rena Bivens found that the incorporation of these identity categories is surface level at best: at the database level, they are collapsed from 58 to only three: male, female, and other.[105] This means that users are given the capacity to signal their gender preferences to others on the site, but only Facebook has the power to define how these users are profiled by advertisers. The choice to expand categories to 58 discrete options rather than to offer text entry further suggests that Facebook was not looking to allow users to define their gender identity themselves, but to provide more options that ultimately mapped to Facebook's prescriptive (and profitable) taxonomy.

The study of racial and gender bias presents both methodological and political challenges: the AI technologies that touch our lives are very often produced at massive industrial scale, yet are highly personalized, making it extremely difficult to see how systems discriminate across thousands or millions of users. Many AI systems don't have "user facing" interfaces at all. Instead, they are integrated into the backend processes of institutional decision making, unseen and often unknown to those whose lives and opportunities they influence. Even when they are known, the companies that produce them almost always use trade secrecy justifications to render their inner workings opaque to outside inquiry.[106]

CORPORATE DIVERSITY: BEYOND THE PIPELINE PROBLEM

The lack of diversity in the AI sector and in tech more broadly has received widespread attention, and a number of popular approaches to solving the problem are now standard within industry. But these have yielded little change – diversity numbers have remained perilously low across the

102 See, for example: Aoun & Ahmed, 2018; Hoffmann, 2018; Pasquale, 2018, Powles, 2018.

103 Sharpe, A. and Raj, S. (2017). Using AI to determine queer sexuality is misconceived and dangerous. *The Conversation*. Retrieved from https://theconversation.com/using-ai-to-determine-queer-sexuality-is-misconceived-and-dangerous-83931.

104 Gaboury, J. (2013). A Queer History of Computing. *Rhizome*. Retrieved from http://rhizome.org/editorial/2013/feb/19/queer-computing-1/.

105 Bivens, R. (2017) The gender binary will not be deprogrammed: Ten years of coding gender on Facebook. New Media & Society, 19(6): 880-898.

106 Advancing this approach to research will likely face significant challenges due to access and legal barriers. Without greater transparency and more robust protections for academic researchers, it will be difficult for future empirical studies to make much headway - the ongoing Sandvig vs. Sessions case is critical to this effort.

AI sector. Given the lack of progress in the face of persistent efforts, we need to scrutinize the means by which the AI industry understands and contends with its lack of diversity. What might need to change to ensure real improvement?

In our research we examined the existing body of literature focused on questions relating to the representation of women in tech fields. Notably, the literature almost solely looked at gender, and represented gender as binary. It much less frequently examined race, or other identities, and even more rarely examined the intersection of such identities. This itself is worthy of scrutiny. The so-called "pipeline" studies - a term used in industry to reference the absence of diverse candidates in the hiring pool, and often to justify the inability of large firms to achieve diversity due to scarcity - commonly engage with questions such as:

- "Why are there so few women computer scientists?"
- "Why do women avoid computer science?"
- "Why are women leaving computing?"
- "Where are the women computer scientists?"
- "Where have all the girls gone?"
- "What draws women to and keeps women in computing?"
- "Why do some gender gaps remain while others do not?"
- "Will Computer Engineer Barbie® impact young women's career choices?"

Relying primarily on survey-based research conducted in educational settings, pipeline studies seek to understand the factors that lead to gender-based discrimination in computer science, more precisely by interrogating what drives women and people of color away from the field, and implicitly, what might make them stay. They have documented factors such as sociocultural dynamics and the role of stereotypes, structural barriers that inhibit access to STEM fields for women, particularly in K-12 education, and overt hostility toward women and people of color who demonstrate an interest in computing.[107,108,109]

Companies that are challenged on their lack of diversity frequently cite pipeline studies as proof of the persistent challenge in finding enough women and people of color to hire. But the evidence suggests otherwise. For example, in 2016 Facebook's Chief Diversity Officer wrote that "It has become clear that at the most fundamental level, appropriate representation in technology or any other industry will depend upon more people having the opportunity to gain necessary skills through the public education system".[110] But, as the Center for Investigative Reporting's study of tech company diversity data found, 91 large tech companies headquartered in Silicon Valley managed to hire higher percentages of black, Latino, and multiracial employees than Facebook that year.[111]

107 See, for example, Corbett & Hill, 2015; Hill, Corbett & Rose, 2010; Scott, Klein & Onovakpuri, 2017; Margolis & Fisher, 2001

108 American Association of University Women. (2008). Where the Girls Are: The Facts About Gender Equity in Education. Retrieved from https://www.aauw.org/research/where-the-girls-are/.

109 Margolis, J. (2010). *Stuck in the Shallow End: Education, Race, and Computing*. Cambridge: MIT Press.

110 Williams, M. (2016, July 14). Facebook Diversity Update: Positive Hiring Trends Show Progress. *Facebook*. Retrieved from https://newsroom.fb.com/news/2016/07/facebook-diversity-update-positive-hiring-trends-show-progress/.

111 Rangarajan, S. (2018, June 25). Here's the clearest picture of Silicon Valley's diversity yet: It's bad. But some companies are doing less bad. *The Center for Investigative Reporting*. Retrieved from https://www.revealnews.org/article/heres-the-clearest-picture-of-silicon-valleys-diversity-yet/.

This view is particularly prevalent among the male entrepreneurs who make up much of the top leadership in the AI industry: one survey of founders of venture-backed companies found that men were over twice as likely to blame the pipeline for the diversity problem in the tech industry than women.[112] But there has been no real change in diversity within tech companies, despite the volume of studies and their relatively consistent findings. Thus, the role pipeline research plays in justifying the diversity status quo within large companies and elite university programs deserves closer scrutiny.

Core Themes in Pipeline Research

What can we learn from these studies, what work are they doing, and for whom? A dominant theme in pipeline-focused research (which, as above, focuses primarily on gender as binary, to the exclusion of other identities) examines the role that cultural factors, and particularly stereotypes play in discouraging women from entering computer science.[113] In almost all cases, women are centered as the subject of concern; rarely are men given the same attention, even though male-dominated environments are a topic of frequent discussion in such studies. Centering the role of culture, this research suggests that a student's self-assessment of whether they are a good fit for the field is likely to influence whether they will leave computing[114,115] and is intertwined with stereotypes of computer scientists as singularly focused, asocial, competitive, and male.[116] Women tend to persist in computer science when they reject and find alternatives to the dominant culture of the field.[117]

Female students are much more likely than male students to observe these stereotypes, presumably because of this experience of dissonance. In one study, half of the computer science students interviewed believed there was a stereotypical geek, hacker, or nerd computer culture. Female students were much more likely to refute the geek image as applicable to themselves.[118]

Other studies suggest that gender is correlated with a person's motivations for pursuing a career in the field. Women, and particularly women from low socioeconomic status or minority backgrounds, are more likely to see computing as a versatile profession that provides an opportunity for secure employment, higher pay, and better social standing.[119,120] Moreover, their interests go beyond technical aspects of computing, focusing instead on the purpose and application of software. However, such interests are often de-emphasized in computer science curricula that prize technical skill and its applicability to industrial settings above all else.[121]

---

112 First Round. (2016). State of Startups. Retrieved from http://stateofstartups.firstround.com/2016/#highlights-diversity-problem.

113 American Association of University Women. (2010). Why So Few? Women in Science, Technology, Engineering, and Mathematics. Retrieved from https://www.aauw.org/resource/why-so-few-women-in-science-technology-engineering-and-mathematics-executive-summary/; American Association of University Women. (2015). Solving the Equation: The Variables for Women's Success in Engineering and Computing. Retrieved from https://www.aauw.org/research/solving-the-equation/.

114 Rodriguez, S.L., & Lehman, K.J. (2018). Developing the next generation of diverse computer scientists: The need for enhanced, intersectional computing identity theory. Computer Science Education (pp. 1–20). doi: https://doi.org/10.1080/08993408.2018.1.

115 Fouad, N.A. (2011). Stemming the Tide: Why Women Leave Engineering. *National Science Foundation*. Retrieved from https://www.energy.gov/sites/prod/files/NSF_Stemming%20the%20Tide%20Why%20Women%20Leave%20Engineering.pdf.

116 Lewis, C.M., Anderson, R.E. and Yasuhara, K. (2016). "I Don't Code All Day": Fitting in Computer Science When the Stereotypes Don't Fit. *ICER '16*.

117 Margolis, J. and Fisher, A. (2001). *Unlocking the Clubhouse: Women in Computing*. Cambridge: MIT Press.

118 Varma, R. (2007). *Women in Computing: The Role of Geek Culture*. Science as Culture, 16(4): 359-376.

119 Ibid.

120 Margolis, J. and Fisher, A. (2001). *Unlocking the Clubhouse: Women in Computing*. Cambridge: MIT Press.

121 Ibid.

Stereotypes may be introduced at many different junctures: the interpersonal relationships and competition among computer science students is an important factor, but so is the program curricula. The kinds of examples used in problem sets and the physical educational environment can shape whether a student develops a sense of belonging and interest in computer science. Often, researchers observed that these influences served both to convey geek stereotypes and to code them as masculine.[122] However, these stereotypes are not universally held: one study of Malaysian computer science students observed that women constitute half of all computer science students in higher education in the country, and espouse a unique, though nevertheless gendered, perspective on the field. In contrast to the West, computer science is seen among Malaysian students as a suitable career path for women because it involves office work which keeps them indoors.[123]

Some studies also examine deeper structural factors influencing gender-based discrimination, such as influences shaping the technical competency of male and female computer science students. Male students are more likely to enter computer science programs with existing programming skills, creating a sense among female computer science students that they are constantly behind. The source of this discrepancy starts at an early age: researchers have found that access to computers both in the home and in school settings is gendered - female students have a harder time gaining access to computers than their male counterparts, so they are at a disadvantage when trying to acquire these skills. Students interested in learning about computing but who lack parental instruction, resources at home, and a peer computing community are likely to lose the most when school resources are inadequate.[124] These challenges have understandable effects on female students' confidence in their skills in computer science. Many female students underestimate their own capabilities, and both male and female students tend to incorrectly believe that male computer science majors have higher GPAs.[125]

Such studies suggest that this lack of confidence in their technical competency, regardless of how the student is actually doing, has significant downstream effects on the likelihood of a female student continuing in the field. One study found that women are under-represented in STEM fields where innate intellectual talent is believed to be necessary for success, and that computer science ranks relatively high among these fields within STEM.[126] Furthermore, a study of computer science majors at Carnegie Mellon University (CMU), a leading university in the field of artificial intelligence, found that many female students enter with high levels of confidence in their abilities, but this is eroded over the course of their freshman and sophomore years. Small injuries were likely to hurt female and minority students more and make them much more likely to drop out of the major, even when their GPA did not bear out the lack of confidence.[127]

Sometimes these injuries are not so small: a study by the National Academies of Science, Engineering and Medicine found that between 20 and 50% of female students in STEM fields, and over 50% of faculty reported experiencing harassment. LGBTQ women and women of color

122 Ibid.

123 Lagesen, V.A. (2008). A Cyberfeminist Utopia?: Perceptions of Gender and Computer Science among Malaysian Women Computer Science Students and Faculty. *Science, Technology, & Human Values*. 33(1): 5-27.

124 Margolis, J. and Fisher, A. (2001). *Unlocking the Clubhouse: Women in Computing*. Cambridge: MIT Press.

125 Beyer, S., Rynes, K. and Haller, S. (2004). Deterrents to Women Taking Computer Science Courses. *IEEE Technology and Society Magazine*.

126 Meyer, M., Cimpian, A. and Leslie, S. (2015). Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in Psychology*, 6(235): 1-12.

127 Margolis, J. and Fisher, A. (2001). *Unlocking the Clubhouse: Women in Computing*. Cambridge: MIT Press.

were more likely than straight or white women to have been harassed.[128] A study that focused specifically on populations of computer science students in minority-serving institutions found that hostility directed at students by male peers and faculty was a prominent factor in computer science students' decision to change majors for all groups surveyed except white males.[129] Institutions can account for these forms of discrimination and harassment: for example, CMU dropped its admissions requirements for prior programming experience in favor of leadership experience, and found this alone increased its enrollment of female students dramatically. Importantly, doing so did not result in compromising the quality of the students as measured by key indicators like GPA or scores on standardized tests.

Limitations of Pipeline Research

Despite the contribution of pipeline studies to a better understanding of the factors influencing participation in technical fields, they can have significant limitations that often go unacknowledged. Some of these issues are methodological. Often they rely on self-reporting by students, and thus readers need to consider the conditions under which the research was conducted and the way in which a narrow population of students is generalized to represent an entire identity, in addition to questions around how such an identity was chosen as a focus of study and how it was defined. What was the relationship between the researcher and participant? Was the researcher a professor at the institution the student attended, and could this relationship shape the student's response to survey questions? What incentives were provided to participate, and what was the environment in which the survey or interview took place? Many of these studies also rely on relatively small samples that may not be representative of the broader experiences of people in the field, or the studies are conducted at a single university that may have its own particularities that are not reflective of conditions elsewhere.

Moreover, the persistent focus on gender as a binary often results in treating it as a biologically essential category that maps to certain attributes. Few of these studies adequately investigate the experiences of students through an intersectional lens. Within this narrow frame, such research almost always focuses on women, and often implies that the problem is one that resides within women's individual psychology, whether it be a lack of confidence or a lack of prior experience, as opposed to an issue with the institutions and their cultures. They also frequently represent the experiences of mostly white students at research universities. This inadequately accounts for the experiences of students of color, students of minority-serving institutions, and community college students, reflecting an already existing dominance in the industry of a few elite universities.[130]

Roli Varma's work illustrates why it's so important to expand whose perspectives are accounted for in diversity research. By focusing on the experiences of members of five major ethnic groups from seven minority-serving institutions, her work illustrates a complex matrix of influences shaping how students of different genders and ethnicities come to view themselves

---

128 National Academies of Sciences, Engineering, and Medicine. (2018). *Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine*. Washington, DC: The National Academies Press.

129 Varma, R. (2007). *Women in Computing: The Role of Geek Culture*. Science as Culture, 16(4): 359-376.

130 See, for example, research by Richard Kerby that indicates 40% of venture investors went to Stanford or Harvard. Kerby, R. (2018, Jul. 30). Where did you go to school? *Medium*, retrieved from https://blog.usejournal.com/where-did-you-go-to-school-bde54d846188?stream=top.

as computer scientists.[131,132] More work that addresses the nuances of how identity shapes students' relationship to computer science is sorely needed, as is more work that interrogates the construction of privileged identities (male, white, etc.). We also need to re-examine the dominant culture of tech as the subject of study, connecting these questions to the current incentives and power relationships that undergird the industry and the field as a whole. Such studies could look not only at who is underserved within the current tech ecology, but who benefits from its present construction and how these dynamics might be untangled to more clearly understand the state of diversity.

In focusing overwhelmingly on university settings, these studies also rely on samples of convenience, though there are other critical juncture points in the pipeline that would be of value to examine in more detail. For example, Wynn & Correll studied the recruitment sessions from tech companies at a prominent university on the West Coast, finding several indicators of gender bias in the recruiting process (they did not look at race or other identities): presenters at the sessions were overwhelmingly male, and in the cases where a female engineer was present, she would rarely talk, or only talk about company culture. Technical material was presented as a male domain, as were the perks emphasized, like foosball tables and beer fridges. By and large, these gendered patterns were reflected in students' engagement in the sessions; presentations that were gender-neutral often led to increased participation by female students.[133] More studies examining recruitment, promotion, and workplace environment would provide a fuller view into the influences shaping the experiences of women and gender non-conforming people in technology.

It is worth considering the scope of these studies' recommendations and context in relation to the central role these studies play within the diversity discourse overall. By and large, the recommendations they issue are limited, targeted at the administrators of university computer science programs seeking to broaden the diversity of their student body. Though important, this is a narrow frame through which to view potential solutions to barriers to inclusion; it does not address the companies that hire computer science students, the peers responsible for promulgating stereotyped views or engaging in hostile behavior, or the broader social conditions that may influence students' success in computer science programs - let alone what awaits them in the corporate environment. Nonetheless, such studies are persistently funded, often through corporate-sponsored initiatives, and are frequently cited by those within corporate environments to justify their own lack of diversity, as they situate the locus of change outside of the corporation itself. As such, pipeline studies are disproportionately emphasized as a part of the broader research agenda on diversity and technology.

131 Ibid.

132 Varma, R. (2010). Why so few women enroll in computing? Gender and ethnic differences in students' perception. *Computer Science Education*, 20(4): 301-316.

133 Wynn, A. and Correll, S. (2018). Puncturing the pipeline: Do technology companies alienate women in recruiting sessions? *Social Studies of Science*, 48(1) 149-164.

Pipeline Dreams: After Years of Research, The Picture Worsens

In 2019, the diversity crisis is now well documented. Despite decades of research, there has been little meaningful headway in remedying these problems within industry, or within academia. In fact, diversity numbers within both industry and academia have either declined over the last decade or stagnated. Evidence suggests that a focus on researching the pipeline problem has not translated into meaningful action by tech companies. A recent survey of 32 leading tech companies found that though many express a desire to improve diversity, only 5% of 2017 philanthropic giving was focused on correcting the gender imbalance in the industry, and less than 0.1% was directed at removing the barriers that keep women of color from careers in tech. This meant that out of $500 million in total philanthropic giving by these companies that year, only $335,000 – across 32 tech companies – went to programs focused on outreach to women and girls of color.[134,135]

So what is motivating the production of so many similar pipeline studies? And, as importantly, does the overwhelming focus on the pipeline, and narrowly on women in computer science, come at the cost of more impactful research and initiatives?

The pipeline frame tends to place the onus to solve issues of discrimination in Silicon Valley on those who are discriminated against, rather than the perpetrators.[136] As Sandra Harding puts it, discussing such research in the context of the sciences more generally: “traditional gender-role research has formulated the problem as lack of success by girls and women, rather than the obstacles that masculine-dominated social institutions raise to women’s success.”[137] As such, it also enables those within this space to engage with the problem of diversity without addressing the deeper, more complex issues at hand. These issues include asking not only who is harmed, but who benefits from the dominant structures governing the current technology ecosystem.

The emphasis on diversity and inclusion can also serve to distract from dealing with actually existing racism and misogyny.[138] Such pipeline discourses are prominent within the emergence of what Sarah Banet-Weiser describes as ‘popular feminism’ - in which the focus is solely on women as a stand-alone identity. As she describes it, “the inclusion of women becomes the solution for all gender problems, not just those of exclusion or absence. It is, of course, important to have bodies at the table, but their mere presence doesn’t necessarily challenge the structure that supports, and builds, the table in the first place”.[139]

---

134 Wittemeyer, R., Nowski, T., Ellingrud, K. and Conway, M. (2018). Rebooting Representation: Using CSR and Philanthropy to Close the Gender Gap in Tech. Retrieved from https://www.rebootrepresentation.org/wp-content/uploads/Rebooting-Representation-Report.pdf.

135 As an indicator of just how small this number is, it represents 0.4% of the $90 million payout Google gave to Andy Rubin, the creator of Android, upon leaving the company amid accusations of sexual misconduct. See: Wakabayashi, D. and Benner, K. (2018, Oct. 25). How Google Protected Andy Rubin, the ‘Father of Android’. *The New York Times*. Retrieved from https://www.nytimes.com/2018/10/25/technology/google-sexual-harassment-andy-rubin.html.

136 Noble, S.U. (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

137 Harding, S. (1986). *The Science Question in Feminism*. Ithaca: Cornell Press, p. 77

138 Jessie Daniels lays out a compelling argument against the growing interest in adopting ‘race-neutral’ or ‘colorblind’ strategies to addressing algorithmic bias, illustrating the importance of racial literacy among industry leaders in order to adequately participate in discussions of racial inequality. See: Daniels, J. (2019, Apr. 3). “Color-blindness” is a bad approach to solving bias in algorithms. *Quartz*. Retrieved from https://qz.com/1585645/color-blindness-is-a-bad-approach-to-solving-bias-in-algorithms/.

139 Banet-Weiser, S. (2018). *Empowered: Popular Feminism and Popular Misogyny*. Durham: Duke University Press, p. 12.

This problem was called out recently in a memo by former Facebook employee Mark S. Luckie outlining systemic failures by the company that impact both black employees and black users. Luckie observed that in some Facebook buildings, there are more "Black Lives Matter" posters than actual black people - a vivid illustration of Sara Ahmed's claim that diversity initiatives often act as a way of rebranding organizations that allow inequalities to be concealed behind a veneer of marketing, and thus reproduced.[140,141]

Addressing tech companies' inclusion problems is a necessary goal and a step toward meaningful change. But diversity initiatives must be accompanied by efforts to address workplace cultures and the logics of how tech systems are designed: cultures of exclusion that have been frequently documented, but that remain woefully unaddressed. As one Google worker recounted to a journalist following the Google Walkout, "I feel like I'm leading young girls and boys to the slaughter. I mean, why would you want to go into tech if it's like this?"[142]

WORKER-LED INITIATIVES

While pipeline approaches have not been an effective mechanism for motivating change in the tech industry, worker-led initiatives are beginning to play a critical role in advocating for corporate diversity.

Founded informally in 2014, the Tech Workers Coalition emerged in 2017 as a hub of tech worker organizing, which often included advocating for more inclusive and equitable tech and tech cultures.[143] Among its efforts was a proposal brought to the board of Alphabet Inc – and rejected by it – that would specify protections for anyone involved in an internal HR investigation, require the company to do more to promote civil discourse on internal message boards, and tie pay to the company's ability to meet its diversity goals.[144] Another proposal that would accelerate Apple's recruitment policy to increase diversity on its board and senior management was also rejected by the company's board as "unduly burdensome and not necessary".[145]

Google has been a prominent home for such worker-led efforts. In October 2018 the New York Times published an account of Google's handling of cases of sexual assault and harassment, documenting its long track record of ignoring such claims, and even rewarding executives

140 Ahmed, S. (2012). *On Being Included: Racism and Diversity in Institutional Life*. Durham: Duke University Press.

141 Years earlier, Luckie wrote another post about his experiences at Twitter, there outlining the profound influence of culture on the exclusion of people of color in tech companies: culture both influences who gets hired (a product of an emphasis on 'culture fit') and shapes the experiences of employees of color, who are often the only people of color in the room during meetings. See: Luckie, M.S. (2015, Sept. 15) What it's actually like to be a Black employee at a tech company. *Medium*. Retrieved from https://medium.com/@marksluckie/what-it-s-actually-like-to-be-a-black-employee-at-a-tech-company-e32bb222818b.

142 Weaver, M., Hern, A., Bekiempis, V., Hepler, L, and Fermoso, J. (2018, Nov. 1). Google walkout: global protests after sexual misconduct allegations *The Guardian*. Retrieved from https://www.theguardian.com/technology/2018/nov/01/google-walkout-global-protests-employees-sexual-harassment-scandals.

143 Weigel, M. (2017, Oct. 31). Coders of the world, unite: can Silicon Valley workers curb the power of Big Tech?. *The Guardian*. Retrieved from https://www.theguardian.com/news/2017/oct/31/coders-of-the-world-unite-can-silicon-valley-workers-curb-the-power-of-big-tech.

144 Reuters. (2018, June 6). Alphabet shareholders reject diversity proposal backed by employees. *Reuters*. Retrieved from https://www.reuters.com/article/us-alphabet-inc-agm/alphabet-shareholders-reject-diversity-proposal-backed-by-employees-idUSKCN1J22BS.

145 O'Brien, S.A. (2016). Apple's board calls diversity proposal 'unduly burdensome and not necessary'. *CNN*. Retrieved from https://money.cnn.com/2016/01/15/technology/apple-diversity/index.html.

accused of this behavior with millions of dollars in exit pay.[146] The article served as a catalyst for workers at the company, and precipitated the Google Walkout for Real Change, in which 20,000 Google employees in 50+ cities around the world walked out of work in protest of the company's actions.

Initially conceived as a women's march, the organizers of the walkout quickly acknowledged that at the core of sexual harassment and discrimination are entrenched abuses of power.[147,148] The movement took on an intersectional and worker-driven agenda that acknowledged that race, class, and sexuality are intertwined with the forms of gender-based discrimination foregrounded in the news reporting on the walkout. Participants made explicit choices to center the needs of the company's temps, vendors and contractors – employees who lack the job security and benefits of more privileged tech workers.

The protesters outlined a list of five demands, including an end to forced arbitration of harassment and discrimination cases, a commitment to pay and opportunity equity, a publicly disclosed sexual harassment report, a clear, uniform, and globally inclusive process for reporting sexual misconduct, and the promotion of the Chief Diversity Officer to answer directly to the CEO and appointment of an Employee Representative to the Board. Of these, Google has met part of the first demand, announcing an end to forced arbitration for full-time workers, and has made a commitment to improve pay and benefits for some TVC workers starting in 2022.[149,150,151]

The walkout is notable for achieving a policy outcome and galvanizing political activism among tech employees, as well as for the intersections the participants found with other worker-driven movements in the tech industry: many called out their opposition to Project Maven, a Department of Defense contract to develop AI systems to be used in drone warfare, and Dragonfly, a Google-developed Chinese search engine that would censor search results in compliance with government requests.

Google is not alone as the home for such efforts, however. In 2019, Microsoft employees participated in a protest at a company meeting asking CEO Satya Nadella to address claims of discrimination and sexual harassment against women in the company.[152] These efforts resulted in Microsoft leadership agreeing publish statistics about concerns raised and violations, as well as data around career progression. This step on the part of Microsoft echoes the demand by

---

146 Wakabayashi, D. and Benner, K. (2018, Oct. 25). How Google Protected Andy Rubin, the 'Father of Android'. *New York Times*. Retrieved from https://www.nytimes.com/2018/10/25/technology/google-sexual-harassment-andy-rubin.html.

147 O'Donovan, C. and Mac, R. (2018, Oct. 30). Google Engineers Are Organizing A Walkout To Protest The Company's Protection Of An Alleged Sexual Harasser. *Buzzfeed*. Retrieved from https://www.buzzfeednews.com/article/carolineodonovan/googles-female-engineers-walkout-sexual-harassment?bftwnews&utm_term=4ldqpgc#4ldqpgc.

148 Meredith Whittaker, one of the co-founders of AI Now, was one of the organizers of the Google Walkout.

149 End Forced Arbitration. (2019, Mar. 21). Google makes arbitration policy change … employees push for progress. *Medium*. Retrieved from https://medium.com/@endforcedarbitration/google-makes-arbitration-policy-change-but-employees-push-for-progress-797eb9faa534.

150 Pichai, S. (2018, Nov. 8). A note to our employees. *Google*. Retrieved from https://www.blog.google/inside-google/company-announcements/note-our-employees/.

151 Birnbaum, E. (2019, April. 4). Google requiring temporary workers, contractors get health care coverage, parental leave. Retrieved from https://thehill.com/policy/technology/436939-google-will-require-healthcare-parental-leave-for-extended-workforce.

152 Tiku, N. (2019, Apr. 4). Microsoft Employees Protest Treatment of Women to CEO Nadella. *WIRED*. Retrieved from https://www.wired.com/story/microsoft-employees-protest-treatment-women-ceo-nadella/.

Google workers that their company create a “Sexual Harassment Transparency Report”. And both Microsoft and Amazon employees issued letters to corporate management demanding that the companies discontinue contracts that support US Immigration and Customs Enforcement after it was revealed the agency was separating migrant parents and children at the border with Mexico.[153,154] These are a handful of the many examples of tech worker-led organizing that have emerged over the past year and a half. Through them we see clearly the ways in which organizing around issues of power and inequity within the tech industry directly relates to issues of power and inequity expressed in and through the technologies the industry is engaged in building.

At the same time, it is worth considering what it means that workers who experience a great deal of privilege are driving protests at AI companies. They have a responsibility to center the voices of those who are most at risk of harm from AI systems and are largely excluded from conversations about AI ethics, including the temps, vendors and contractors who make up an ever growing part of the workforce at these companies.[155]

A larger question remains: will diversifying the ranks of tech company workers necessarily address the deeper structural challenges AI systems pose to communities? This question emphasizes why a deeper analysis of power is so critical to examining the relationship between technology development and the lived experiences of individuals of different racialized, gendered, and classed identities.[156]

THE PUSHBACK AGAINST DIVERSITY

It is a critical time to be addressing the diversity crisis in AI, because we now see diversity itself being weaponized. Over the past year and a half, evidence of systemic discrimination and harassment at tech companies and conference spaces has entered the public debate, much of it exposed by worker-led initiatives and whistleblowers. This growing awareness, accompanied by demands for inclusion and equity, has led to some change, but there has also been resistance, especially among those implicitly privileged by the status quo.

Those questioning and even rejecting the idea that racism, misogyny, and harassment are problems within the AI field and the tech industry have appropriated the language of diversity to argue that efforts to improve inclusion are in fact ‘exclusionary’, and that addressing the deeper structural challenges posed by racism, sexism, and inequity is misguided. For example, some

153 Frenkel, S. (2018, June 19). Microsoft Employees Protest Work With ICE, as Tech Industry Mobilizes Over Immigration. *The New York Times*. Retrieved from https://www.nytimes.com/2018/06/19/technology/tech-companies-immigration-border.html.

154 Shaban, H. (2018, June 22) Amazon employees demand company cut ties with ICE. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/the-switch/wp/2018/06/22/amazon-employees-demand-company-cut-ties-with-ice/?utm_term=.65c50c634d97.

155 Google Walkout For Real Change. (2018, Dec. 5). Invisible no longer: Google’s shadow workforce speaks up. *Medium*. Retrieved from https://medium.com/@GoogleWalkout/invisible-no-longer-googles-shadow-workforce-speaks-up-9ea04b7bcc41.

156 For examples, see: Keyes, O. (2018). The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. Retrieved from https://ironholds.org/resources/papers/agr_paper.pdf, Costanza-Chock, S. (2018, Jul. 27). Design Justice, A.I. and Escape from the Matrix of Domination. *Journal of Design and Science*. Retrieved from https://jods.mitpress.mit.edu/pub/costanza-chock.

AI researchers greeted the announcement of the Black in AI workshop at NeurIPS, a leading machine learning conference, by questioning whether the event was necessary and arguing that it would be discriminatory.[157,158]

Such pushback often centers calls for "cognitive diversity" or "viewpoint diversity," the idea that individual differences in the ways people think and understand the world are distinctions that should be counted alongside, or instead of, other identity categories such as race and gender. As Bärí A. Williams puts it, "a dozen white men, so long as they were not raised in the same household and don't think identical thoughts, could be considered diverse."[159]

These arguments work by centering "identity" while flattening or ignoring power relationships. For example, in 2017 Facebook VP of Engineering Regina Dugan said that "the ultimate goal is cognitive diversity, and cognitive diversity is correlated with identity diversity. That means it's not just about [getting] women in tech. It's about broad voices, broad representation. But we can't step away from the idea that in the workplace, diversity also looks like identity diversity. You have to get to the place where you aren't made comfortable by the fact that everyone is the same, but rather feel inspired by how different we are."[160]

Instead of looking at historical patterns of marginalization, calls for cognitive diversity argue that all differences are equal. Such arguments are circulating among some of the leaders of the AI industry, as was recently exemplified in the controversy over Google's appointment of Heritage Foundation CEO Kay Coles James to its Advanced Technology External Advisory Council. Google's reasoning for the appointment of James was ostensibly to ensure 'diversity of thought' by including a conservative viewpoint on the Council. James is also a black woman, thus adding racial and gender diversity to the panel. But the pushback following James' inclusion focused on her policy positions, citing specifically her vocal anti-LGBTQ and anti-immigrant views and highlighted why cognitive diversity is a particularly limited lens.[161] In a letter opposing the appointment, a group of Google workers calling themselves Googlers Against Transphobia and Hate responded to the idea that 'diversity of thought' justified James' addition to the council: "This is a weaponization of the language of diversity. By appointing James to the ATEAC, Google elevates and endorses her views, implying that hers is a valid perspective worthy of inclusion in its decision making. This is unacceptable."[162,163]

---

157 Kahn, J. and Bass, D. (2017, Oct. 20). Black AI Workshop Becomes Latest Flashpoint in Tech's Culture War. *Bloomberg*. https://www.bloomberg.com/news/articles/2017-10-20/black-ai-workshop-becomes-latest-flashpoint-in-tech-s-culture-war.

158 Though we don't discuss it in detail here, NeurIPS has become a particular site of controversy around issues of diversity and inclusion in the field of AI. In addition to the events described above, there have been allegations by attendees of harassing behavior and patterns of discrimination at the conference. 2018 also saw an ongoing debate over the acronym used to describe the conference, which for years drove inappropriate jokes that many community members felt led to a hostile environment. Almost 2,000 people signed a petition advocating a change to the conference name, eventually leading to the adoption of the new acronym.

159 Williams, B.A. (2017, Oct. 16). Tech's Troubling New Trend: Diversity Is in Your Head. *The New York Times*. Retrieved from https://www.nytimes.com/2017/10/16/opinion/diversity-tech-women-silicon-valley.html.

160 Fast Company. (2017, Jan. 9) Facebook Engineering VP Explains Why "Cognitive Diversity Is the Most Powerful Tool". Retrieved from https://www.fastcompany.com/3066345/facebooks-vp-of-engineering-cognitive-diversity-is-the-most-powerful-tool.

161 Shead, S. (2019, Apr. 5). Google A.I. Panel Member Says Google 'Pulled The Plug Rather Than Defend Themselves'. *Forbes*. Retrieved from https://www.forbes.com/sites/samshead/2019/04/05/google-a-i-panel-member-says-google-pulled-the-plug-rather-than-defend-themselves/#27d051439554.

162 Googlers Against Transphobia and Hate. (2019, Apr. 1). Retrieved from https://medium.com/@against.transphobia/googlers-against-transphobia-and-hate-b1b0a5dbf76.

163 AI Now co-founder Meredith Whittaker was one of the organizers of the Googlers Against Transphobia and Hate letter

The idea of cognitive diversity is mobilized by some to support the conclusion that the AI field and the tech industry are already diverse, even going so far as to support claims that not including identities like "white" and "male" constitutes discrimination. A July 2017 memo written by James Damore, a Software Engineer at Google, is illustrative of such pushback.[164] Titled "Google's Ideological Echo Chamber" and published on an internal mailing list, the memo critiqued the company's diversity policies, arguing that biological differences between men and women (rather than bias and discrimination) help explain gender disparities at the company. Damore's objective in writing the memo was to make a case that policies designed to achieve equal representation are "unfair, divisive, and bad for business."[165] Supporters of Damore's point of view at times even drew on the rhetoric of the pipeline to make the case that diversity initiatives are in fact discriminatory: they argue, incorrectly, that if there aren't qualified candidates in the pipeline, then hiring those who are unqualified on the basis of identity discriminates against those who are qualified.[166] Damore went on to file an unsuccessful complaint with the National Labor Relations Board asserting wrongful termination,[167] and sued Google in a class action lawsuit for alleged discrimination against white conservative men.[168,169]

In an update to the memo, Damore himself asserted that he values "diversity and inclusion, [is] not denying that sexism exists, and [doesn't] endorse using stereotypes".[170] But his primary concern was cognitive diversity. Echoing Weiss, one respondent to a thread on the anonymous industry forum Blind said, "to suggest that adding women to the mix increases diversity is to suggest that women are, on the whole, different from men. Isn't that just a sexist stereotype? It seems a hypocritical contradiction to me. Since all individuals are in reality different (whether a man or a woman), surely adding another man to the workforce contributes to diversity just as much as adding another woman?"[171]

This quote illustrates the effect of flattening diversity and inclusion discourses: 'diversity' becomes an empty signifier stripped of the histories and lived experiences of systemic discrimination and repurposed around ideology, rather than bodies. A deeper analysis of the link between power inequities and the historical practices that value some identities more highly than others is needed, particularly as they emerge within technical communities. Indeed, within hours of the memo's publication, harassment targeting minority advocates who pushed back against

---

164 Conger, K. (2017, Aug. 5). Exclusive: Here's The Full 10-Page Anti-Diversity Screed Circulating Internally at Google. *Gizmodo*. Retrieved from https://gizmodo.com/exclusive-heres-the-full-10-page-anti-diversity-screed-1797564320.

165 Emerson, S., Matsakis, L. and Koebler, J. (2017, Aug. 5). Internal Reactions to Google Employee's Manifesto Show Anti-Diversity Views Have Support. *Motherboard*. Retrieved from https://motherboard.vice.com/en_us/article/ywpamw/internal-reaction-to-google-employees-manifesto-show-anti-diversity-views-have-support.

166 In fact, evidence shows that there are ample qualified candidates - for a more extensive refutation of this claim, see Daniels, J. (2019, Apr. 3). "Color-blindness" is a bad approach to solving bias in algorithms. *Quartz*. Retrieved from https://qz.com/1585645/color-blindness-is-a-bad-approach-to-solving-bias-in-algorithms/ and Grant, N. (2018, June 13). The Myth of the 'Pipeline Problem'. *Bloomberg*. Retrieved from https://www.bloomberg.com/news/articles/2018-06-13/the-myth-of-the-pipeline-problem-jid07tth.

167 Matsakis, L. (2018, Feb. 16). Labor Board Rules Google's Firing of James Damore Was Legal. *WIRED*. Retrieved from https://www.wired.com/story/labor-board-rules-google-firing-james-damore-was-legal/.

168 James Damore vs. Google: Class Action Lawsuit. Retrieved from https://www.scribd.com/document/368688363/James-Damore-vs-Google-Class-Action-Lawsuit#from_embed.

169 This suit has now moved out of court and in to arbitration; see: https://cdn.vox-cdn.com/uploads/chorus_asset/file/13287023/18CV321529.pdf.

170 Damore, J. (2017, July).Google's Ideological Echo Chamber: How bias clouds our thinking about diversity and inclusion. Retrieved from https://assets.documentcloud.org/documents/3914586/Googles-Ideological-Echo-Chamber.pdf.

171 Shablamo. (2017, Aug. 4). Any googlers wanna talk about this manifesto? *Blind*. Retrieved from https://www.teamblind.com/article/any-googlers-wanna-talk-about-this-manifesto-gEa43hGg.

the claims in the memo began, with a particular focus on queer and trans workers.[172,173] Google's Vice President of Diversity even locked down her Twitter account shortly after Damore's firing, responding to a barrage of threats describing her as a "police Nazi".[174]

Damore's memo also stated that "the distribution of preferences and abilities of men and women differ in part due to biological causes and that these differences may explain why we don't see equal representation of women in tech and leadership."[175] This assertion hinges on a flawed assumption that identities like gender and race are essential and fixed biological attributes, and that inequalities are at least in part the product of such irreducible differences.

Biological determinism, like that of Damore, comes from a long history. Stephen Jay Gould describes it as the idea that "the social and economic differences between human groups – primarily races, classes, and sexes – arise from inherited, inborn distinctions and that society, in this sense, is an accurate reflection of biology."[176] The periodic resurgence of biological determinism is, according to Gould, more a product of political circumstances than a set of historical arguments marked by new or inventive logics – it resurfaces during "episodes of political retrenchment, particularly with campaigns for reduced government spending on social programs, or at times of fear among ruling elites, when disadvantaged groups sow serious unrest or even threaten to usurp power."[177]

So it is notable that similar determinist logics are currently emerging within AI systems themselves. There have been many recent examples: from a 2016 paper (widely maligned) that claimed a machine learning model could predict whether an individual was a criminal from their ID photo,[178] to Faception, a commercial AI vendor that markets their facial analysis systems as capable of determining, via an image, whether someone is an "extrovert, a person with High IQ, Professional Poker Player or a threats [sic],"[179] to mainstream face recognition systems that claim to recognize ethnicity, gender, and emotion.[180] Such systems locate identity, character - and, often, social worth - in physical, biological, and externally "knowable" attributes, effectively making a value judgement about a person based on their mannerisms and the appearance of their body. Based on these determinations, such AI systems are increasingly tasked with sorting those who are worthy from those who are not – be it for school admission, release from prison, or job interviews. Rarely are such decisions contestable or even visible to the people most at risk of harm.

---

172 Tiku, N. (2018, Jan. 26). The Dirty War Over Diversity Inside Google. *WIRED*. Retrieved from https://www.wired.com/story/the-dirty-war-over-diversity-inside-google/.

173 Swisher, K. (2017, Aug. 10). Google CEO Sundar Pichai cancels all-hands meeting about gender controversy due to online harassment. *Recode*. Retrieved from https://www.recode.net/2017/8/10/16128380/google-cancels-all-hands-meeting-controversy-memo.

174 Ghosh, S. (2017, Aug. 8). Google's diversity VP has locked down her tweets after receiving racist and sexist insults. Business Insider. Retrieved from https://www.businessinsider.com/google-diversity-vp-danielle-brown-protected-tweets-harassment-2017-8.

175 Damore, J. (2017, July).Google's Ideological Echo Chamber: How bias clouds our thinking about diversity and inclusion. Retrieved from https://assets.documentcloud.org/documents/3914586/Googles-Ideological-Echo-Chamber.pdf.

176 Gould, S.J. (1981). *The Mismeasure of Man*. New York: W.W. Norton & Co., p.54.

177 Gould, S.J. (1981). *The Mismeasure of Man*. New York: W.W. Norton and Co., p. 28.

178 Wu, X. and Zhang, X. (2016). Automated Inference on Criminality using Face Images. Retrieved from https://arxiv.org/pdf/1611.04135v2.pdf.

179 See: https://www.faception.com/.

180 See: https://azure.microsoft.com/en-us/services/cognitive-services/face/.

Biological determinism has now re-emerged both as a pushback against calls for equity and inclusion, and as a foundational idea within AI system design. Given the work that biological determinism has done historically – justifying the 'superiority' of white people over black people, men over women, and extant social hierarchies as a product of biological destiny, it is urgent that these ideas are rejected as spurious. Most importantly, we need to consider who benefits and who bears the cost of the widespread automation of such logics.

CONCLUSION

The diversity crisis in AI is well-documented and wide-reaching. It can be seen in unequal workplaces throughout industry and in academia, in the disparities in hiring and promotion, in the AI technologies that reflect and amplify biased stereotypes, and in the resurfacing of biological determinism in automated systems.

Throughout this report, we've outlined the scope and scale of the problem, tracing how the diversity crisis in the industry and the problems of bias in AI systems are interrelated aspects of the same issue. In the past, these topics were commonly examined in isolation, but increasing evidence shows that they are closely intertwined. By studying these connections further, we can open new pathways to redressing imbalances and harms.

Our analysis surfaced two prominent responses to the diversity crisis: on the one hand, a worker-driven movement focused on addressing inequities is showing promise in driving change. On the other hand, we observe a small but vocal counter-movement that actively resists diversity in the industry and uses arguments from biological determinism to assert that women are inherently less suited to computer science and AI.

This is a critical moment for the AI industry to decide what it will do. As AI systems are embedded in more social domains, they are playing a powerful role in the most intimate aspects of our lives: our health, our safety, our education, and our opportunities. It's essential that we are able to see and assess the ways that these systems treat some people differently than others, because they already influence the lives of millions.

In the initial findings and recommendations from our multi-year research project, we seek to trace out a positive path forward. Our objective should not be to simply diversify the privileged class of technical workers engaged in developing AI systems in the hope that this will result in greater equity. Nor should it be to develop bespoke technical approaches to systemic problems of bias and error, hoping that others won't come along. Instead, by broadening our frame of reference and integrating both social and technical approaches, we can begin to chart a better path forward.

ACKNOWLEDGMENTS

We wish to acknowledge Genevieve Fried, Mike Ananny, Roel Dobbe, Casey Gollan, Joan Greenbaum, Elizabeth Kaziunas, Os Keyes, Varoon Mathur, Rashida Richardson, and Lucy Suchman for their generous feedback on the draft. Pivotal Ventures, founded by Melinda Gates, provided support for this project.

AI NOW