



PARTNERSHIP ON AI

RESEARCH
PAPER

After the Offer: The Role of Attrition in AI's 'Diversity Problem'

Dr. Jeffrey Brown

APRIL 13, 2022



Contents

Executive Summary	3
Key Takeaways	4
Recommendations	6
Introduction	9
Why Study Attrition of Minoritized Workers in AI?	9
Background	11
Problems Due to Lack of Diversity of AI Teams	11
More Diverse Teams Yield Better Outcomes	12
Current Level of Diversity in Tech	12
Diversity in AI	15
What Has Been Done	15
Looking at the Pipeline	16
What Leads to Attrition?	17
Attrition in Tech	18
Current Study and Methodology	19
Recruitment	19
Participants	19
Measure	20
Procedure	20
Analysis	21
Results	22
Domain: Attrition	22
Domain: Culture	25
Domain: Efforts to Improve Inclusivity	28
Key Takeaways From Across All Domains	31
Summary and the Path Forward	32
Acknowledgements	33
Appendices	34
Appendix 1: Recruitment Document	35
Appendix 2: Privacy Document	37
Appendix 3: Research Protocol	38
Appendix 4: Important Terms	41
Notes	45

Executive Summary

Amid heightened attention to society-wide racial and social injustice, organizations in the AI space have been urged to investigate the harmful effects that AI has had on marginalized populations. It's an issue that engineers, researchers, project managers, and various leaders in both tech companies and civil society organizations have devoted significant time and resources to in recent years. In examining the effects of AI, organizations must consider who exactly has been designing these technologies.

Diversity reports have revealed that the people working at the organizations that develop and deploy AI lack diversity across several dimensions. While organizations have blamed pipeline problems in the past, research has increasingly shown that once workers belonging to minoritized identities get hired in these spaces, systemic difficulties affect their experiences in ways that their peers from dominant groups do not have to worry about.

Attrition in the tech industry is a problem that disproportionately affects minoritized workers.* In AI, where technologies already have a disproportionately negative impact on these communities, this is especially troublesome.

We are left wondering: What leads to these folks leaving their teams, organizations, or even the AI field more broadly? What about the AI field in particular influences these people to stay or leave? And what can organizations do to stem this attrition to make their environments more inclusive?

The current study uses interviews with folks belonging to minoritized identities across the AI field, managers, and DEI (diversity, equity, and inclusion) leaders in tech to get rich information about what aspects of cultures within an organization promote inclusion or contribute to attrition. Themes that emerged during these interviews formed 3 key takeaways:

1. Diversity makes for better team climates
2. Systemic supports are difficult but necessary to undo the current harms to minoritized workers
3. Individual efforts to change organizational culture fall disproportionately on minoritized folks who are usually not professionally rewarded for their efforts

* This report uses *minoritized workers* as an umbrella term to refer to people whose identities (in categories such as race, ethnicity, gender, or ability) have been historically marginalized by those in dominant social groups. The minoritized workers in this study include people who identified as minoritized within the identity categories of race and ethnicity, gender identity, sexual orientation, ability, and immigration status. Because this study was international in scope, it is important to note that these categories are relative to their social context.

In line with these takeaways, the study makes 4 recommendations about what can be done to make the AI field more inclusive for workers:

1. Organizations must systemically support Employee Resource Groups (ERGs)
2. Organizations must intentionally diversify at leadership and management levels
3. DEI trainings must be specific in order to be effective and be more connected to the content of AI work
4. Organizations must interrogate their values as practiced and fundamentally alter them to include the perspectives of people who are not White, cis, or male

These takeaways and recommendations are explored in more depth below.

Key Takeaways

1 Diversity makes for better team climate

Across interviews, participants consistently expressed that managers who belonged to minoritized identities or who took the time to learn about working with diverse identities were more supportive of their needs and career goals. Such efforts reportedly resulted in teams that were also more diverse, inclusive, interdisciplinary, and engendering of a positive team culture/climate. In these environments, workers belonging to minoritized identities thrived. A diversity in backgrounds and perspectives was particularly important for AI teams that needed to solve interdisciplinary problems.

Conversely, the negative impact of work environments that were sexist or where participants experienced acts of prejudice such as microaggressions was also a recurring theme.

While collaborative or positive work environments were also a common theme, such environments did not in themselves negate predominant cultures which deprioritized “DEI-focused” work, work that was highly interdisciplinary, or work that did not serve the dominant group. Negative organizational cultures seemed to exacerbate experiences of prejudice or discrimination on AI teams.

2 Systemic supports are more difficult but necessary to undo the current harms to minoritized workers

Participants belonging to minoritized identities said that they either left or intended to leave organizations that did not support their continued career growth or possessed values that did not align with their own. Consistent with this, participants described examples of their organizations not valuing the content of their work.

Participants also tied their desires to leave with instances of prejudice or discrimination, which may also be related to “toxic” work environments. Some participants reported instances of being tokenized or being subject to negative stereotypes about their identity groups, somewhat reflective of wider contexts in tech beyond AI.

Systemic supports include incentive structures that allow minoritized workers to succeed at every level, from the teams that they work with actively validating their experiences to their managers finding the best ways for them to deliver work products in accordance with both individual and institutional needs. Guidelines for promotion that recognize the barriers these workers face in environments mostly occupied by dominant group norms are another important support.

3 Individual efforts to change organizational culture fall disproportionately on minoritized folks who are usually not professionally rewarded for their efforts

Individuals discussed ways in which they tried to make their workplaces or teams more inclusive or otherwise sought to incorporate diverse perspectives into their work around AI. Participants sometimes had to contend with bias against DEI efforts, reporting that other workers in their organizations would dismiss their efforts as lacking rigor or focus on the product.

There were some institutional efforts to foster a more inclusive culture, most commonly DEI trainings. DEI trainings that were very specific to some groups (e.g., gender diverse folks, Black people) were reported as being the most effective. However, even when they were specific, DEI trainings seemed to be disconnected from some aspects of the workplace climate or the content of what teams were working on.

Participants who mentioned Employee Resource Groups (ERGs) uniformly praised them, discussing the huge positive impact they had on a personal level, forming the bases of their social support networks in their organizations and having a strong impact on their ability to integrate aspects of their identities or other “DEI topics” they were passionate about into their work.



Teams that are not diverse do not attract and keep minoritized individuals.

Recommendations

1 Organizations must systemically support Employee Resource Groups (ERGs)

Employees specifically named ERGs as one of their main sources of support even in work environments that were otherwise toxic. Additionally, ERGs provided built-in mentorship for those who did not have ready access to mentors or whose supervisors had not done the work to understand the kinds of support needed for those of minoritized identities to thrive in predominantly White and male environments.

WHAT MAKES THIS RECOMMENDATION WORK?

Within these ERGs, there existed other grass-roots initiatives that supported workers, such as informal talking circles and networks of employees that essentially provided peer mentoring that participants found crucial to navigating White- and male-dominated spaces. The mentorship provided by ERGs was also essential when HR failed to provide systemic support for staff and instead prioritized protecting the organization.

WHAT MUST BE IN PLACE?

While participants uniformly praised ERGs, they required large amounts of time from staff members that detracted from their work. Such groups also ran the risk of getting taken over by leadership and having their original mission derailed. Institutions should seek a balance between supporting these groups and giving them the freedom to organize in pursuit of their own best interests.

WHAT WON'T THIS SOLVE?

ERGs will not necessarily make an organization's AI or tech more inclusive. Rather, systematically supporting ERGs will provide more support and community for minoritized workers, which is meant to promote a more inclusive workplace in general.

2 Organizations must intentionally diversify at leadership and management levels

WHAT MAKES THIS RECOMMENDATION WORK?

Participants repeatedly pointed to managers and upper-level leaders who belonged to minoritized identities (especially racial ones) as important influences, changing policy that permeated through various levels of their organizations. A diverse workforce may also bring with it multiple perspectives, including those belonging to people from different disciplines who may be interested in working in the AI field due to the opportunity for interdisciplinary collaboration, research, and product development. Bringing in folks from various academic, professional, and technical backgrounds to solve problems is especially crucial for AI teams.

WHAT MUST BE IN PLACE?

There must be understanding about the reasons behind the lack of diversity and the “bigger picture” of how powerful groups more easily perpetuate power structures already in place. Participants spoke of managers who did not belong to minoritized identities themselves but who took the time to learn in depth about differences in power and privilege in the tech ecosystem, appreciating the diverse perspectives that workers brought. These managers, while not perfect, tended to take advocating for their reports very seriously, particularly female reports who often went overlooked.

WHAT WON'T THIS SOLVE?

Intentionally diversifying leadership and managers will not automatically create a pipeline for diversity at the leadership level, nor will it automatically override institutional culture or policies that ignore DEI best practices.

3 DEI trainings must be specific in order to be effective and be more connected to the content of AI work

WHAT MAKES THIS RECOMMENDATION WORK?

Almost all participants reported that their organizations mandated some form of DEI training for all staff. These ranged widely, from very general ones to very specific trainings that discussed cultural competency about more specific groups of people (e.g., participants reported that there were trainings on anti-Black racism). Participants discussed that the more specific trainings tended to be more impactful.

WHAT MUST BE IN PLACE?

Organizations must invest in employees who see the importance of inclusive values in AI research and product design. Participants pointed to the importance of managers who had an ability to foster inclusive team values, which was not something that HR could mandate.

WHAT WON'T THIS SOLVE?

As several participants observed, DEI trainings will not uproot or counteract institutional stigmas against DEI. It would take sustained effort and deliberate alignment of values for an organization to emphasize DEI in its work.

4 Organizations must interrogate their values as practiced and fundamentally alter them to include the perspectives of people who are not White, cis, or male

WHAT MAKES THIS RECOMMENDATION WORK?

Participants frequently reported that a misalignment of values was a primary reason for them leaving (or wanting to leave) their organizations. Participants in this sample discussed joining the AI field to create a positive impact while growing professionally. This led them to feeling disappointed when their organizations did not prioritize these goals (despite them being among their stated values).

WHAT MUST BE IN PLACE?

Participants found it frustrating when organizations stated that they valued diversity and then failed to live up to this value with hiring, promotion, and day-to-day operations, ignoring the voices of minoritized individuals. If diversity is truly a value, organizations may have to investigate their systems of norms and expectations that are fundamentally male, Eurocentric, and do not make space for those from diverse backgrounds. They then must take additional steps to consider how such systems influence their work in AI.

WHAT WON'T THIS SOLVE?

Because achieving a fundamental re-alignment like this is a more comprehensive solution, it cannot satisfy the most immediate and urgent needs for reform. Short-term, organizations must work with DEI professionals to recognize how they are perpetuating potentially harmful norms of the dominant group and work to create policies that are more equitable. Longer-term fixes may not, for instance, satisfy the immediate and urgent need for more diversity in leadership and teams in general.

Introduction

The AI field is rife with examples of harm meted out on communities of color. The populations these technologies affect are not demographically represented on the teams that design AI. While part of the challenge is the candidate pipeline, or recruiting diverse candidates to tech organizations, an overlooked challenge is attrition, or why diverse workers leave AI teams or organizations once there. The following study is the first of its kind in investigating some of the reasons why minoritized folks on AI teams leave these teams and organizations, what this has to do with the culture of these teams and organizations, and what can be done to reduce this attrition and make these teams and organizations more inclusive. When these organizations become more inclusive it results in AI that is designed with a more representative population in mind.

Why Study Attrition of Minoritized Workers in AI?

1 AI has repeatedly exhibited bias in several areas, and part of the problem could be the predominantly White and male makeup of those working in the field.

AI algorithms have long faced criticisms for propagating the biases of the people who design them. Dr. Joy Buolamwini and Dr. Timnit Gebru's seminal work on AI bias in facial recognition showed that facial recognition algorithms were most likely to misclassify the faces of Black women.¹ Work by Dora Zhao and colleagues also recently showed that algorithms showed bias in captioning faces of lighter-skinned individuals compared to darker-skinned individuals, and that this bias seemed to be starker than older classification models.²

Beyond the harmful applications of this technology in things such as surveillance,³ biased financial practices,⁴ and hiring, the people designing these algorithms are homogeneous, especially in terms of race and gender.⁵ Given the increasingly pervasive role of AI technologies in the lives of people all around the world, particularly marginalized people, there has rightfully been increasing concern about the people who are designing this technology and whether they represent the people who are being affected.⁶

2 AI algorithms have had measurable effects on how we conduct day-to-day business and have far-reaching effects, including on how we recruit people to companies, how we conduct educational practices, and how we use technology in several other facets of our daily life.

The pandemic has accelerated the widespread integration of AI technology into our lives, with such systems seeing use in remote education⁷ and work environments.⁸ AI algorithms have also been increasingly used in the banking industry and other high-impact decision-making contexts.

As the adoption of AI technologies has quickly increased over the past decade, it is both urgent and crucial to ensure that the people designing these technologies do not also create products with bias that inflict harm upon minoritized communities.

3 The specific reasons minoritized folks leave AI teams requires deeper investigation.

People on AI teams, like those in the tech industry more generally, are predominantly White and male. Although organizations have often blamed recruitment and the “pipeline problem” for the lack of diversity on teams, this does not fully explain the reasons why people who belong to minoritized identities often leave their teams once they are working for these organizations. This is a cyclic process where teams that are not diverse do not attract and keep minoritized individuals, so these teams do not become more diverse and accepting places.

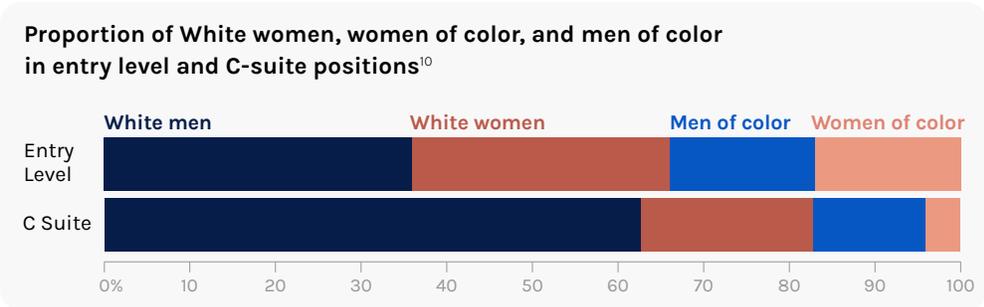
A report from the Kapor Center⁹ outlined some of these issues in the tech ecosystem more broadly, citing unfair and opaque promotion practices, incidents of bias, and lack of clear paths for growth and promotion as several reasons why minoritized folks leave the tech field at a higher rate than those belonging to dominant groups.

This report will probe these issues more deeply, and investigate what specifically about the culture of AI teams may contribute to people staying or leaving. Finally, this study investigates potential ways to make these teams more inclusive.

Background

Problems Due to Lack of Diversity of AI Teams

AI algorithms have caused well-documented problems due to a lack of consideration about the nuance and sensitivity of human diversity in race, gender, sexual orientation, ability, religion, and other identity categories. Categories within these identities and the intersections thereof are subject to power imbalances based on historical marginalization and oppression which contribute to contemporary systemic inequality. One striking example is differences in gender representation as employees move up the corporate ladder.



A 2021 report by McKinsey found that the proportion of White women, women of color, and men of color went from about 30%, 17%, and 17% at the entry level to about 20%, 4%, and 13%, respectively, in the C-suite.¹⁰ In contrast, White men comprise about 35% of entry-level positions and 62% of C-Suite positions. A survey of entrepreneurs by Silicon Valley Bank suggested similar numbers among tech startups, with less than half of respondents reporting even one female or ethnic minority C-suite executive on their teams.¹¹ Tech teams thus lack representation by the people affected by their products, although a tokenistic representation is not enough to undo the systemic harms caused by these products.

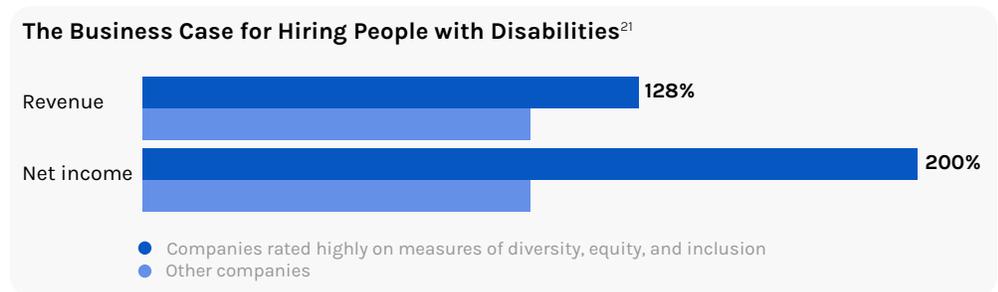
There are several examples of AI that have harmed or had the potential to harm diverse populations all over the world. Some examples include:

- An algorithm used to determine credit lines being biased against female applicants¹²
- The exclusion or erasure of trans people in health tech apps and other medical tech where outdated binary conceptions of gender are used¹³
- Discrimination against home loan applicants by mortgage-approval algorithms, which denied applications from people of color at significantly higher rates than those of similar White people¹⁴

AI ethics researchers have recommended several steps to mitigate algorithmic bias, including training them on data from diverse individuals or ensuring that humans oversee the deployment of these algorithms and overrule biased decisions if necessary.¹⁵ However, this may not be possible without a team of workers who have a deep and comprehensive understanding of how these algorithms can negatively impact affected populations, deeper than a superficially created team that is diverse but lacks understanding of systemic prejudice.

More Diverse Teams Yield Better Outcomes

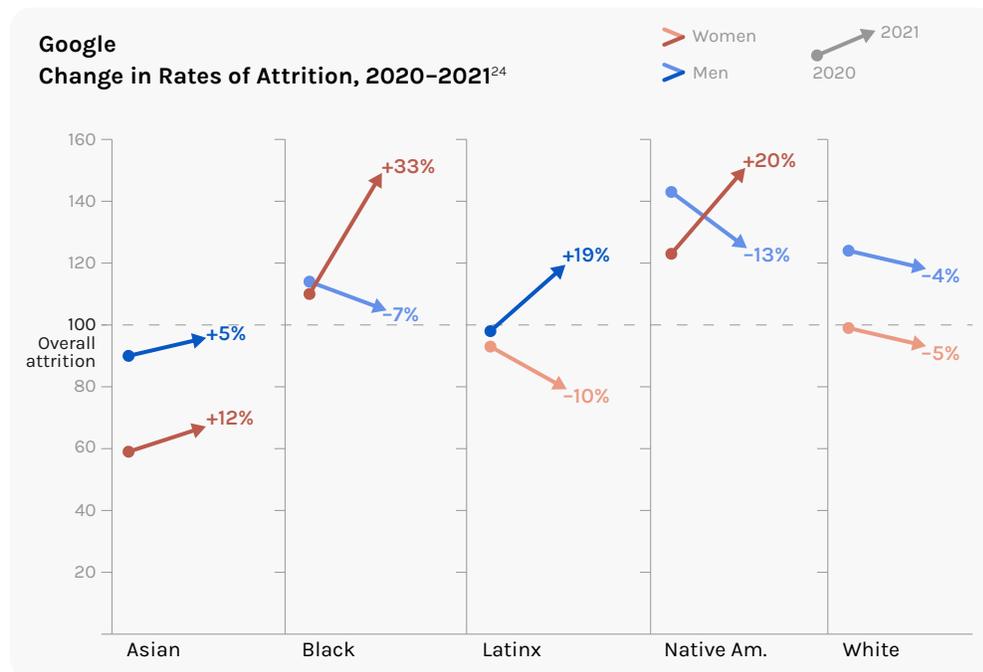
Research has repeatedly shown that diversity on a team serves a greater purpose than good optics. Several studies suggest that diverse teams make more accurate decisions¹⁶ and help to drive innovation.¹⁷ Organizational diversity is associated with higher profits.¹⁸ Internationally, companies with higher numbers of women in executive positions were more successful.¹⁹ Many experts agree with the general conclusions of studies on corporate diversity more broadly and how they could apply to tech teams.²⁰



A joint report from Accenture, the American Association of People with Disabilities, and Disability:IN emphasized the business case for hiring employees with disabilities, reporting a 28% advantage in revenue and twice the net income among other benefits, such as increased returns for shareholders, among companies rated highly for their disability programs and initiatives.²¹ This is crucial for AI teams specifically because of the inherent interdisciplinary nature of the issues that these teams tackle. For instance, AI systems must take into account disability status as an axis of diversity in order to avoid the marginalization of these populations.²²

Current Level of Diversity in Tech

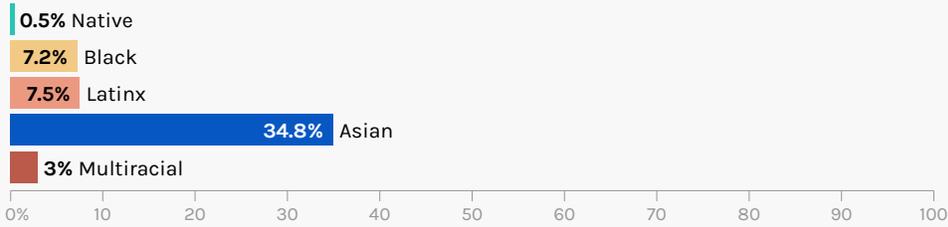
After the murder of George Floyd, the summer of 2020 brought the appearance of a reckoning around anti-Black racism across the US and the world.²³ Despite several pledges to increase diversity of staff and build more inclusive work environments, the tech industry continues to have problems with both. The available data reveal a pattern of increasing diversity within the past few years, but still far from having a workforce that is diverse and representative of the population as a whole. Numbers are most readily available for workers in the US, but some of the same patterns apply to the tech workforce in European countries. This report avoids a direct comparison between the racial demographics in the workforce of other countries since there is a difference in how these places construct race. Here are some examples from recent diversity reports:



Google’s 2021 Diversity Report showed that two groups of women of color left the company at particularly high rates. The company’s “attrition index,” anchored at around 100, went from 110 to 146 for Black women and from 123 to 148 for Native American women. Overall, attrition indices rose for all racially minoritized groups, despite recruiting more racially minoritized individuals in the US in 2020, and more women globally in 2021 than in 2020.²⁴

Between 2020 and 2021, the number of Black people at Facebook rose from 3.9% to 4.4%.²⁵

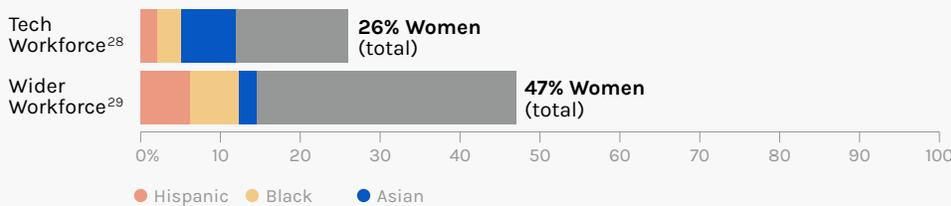
Amazon Corporate Workforce Racial Diversity in 2020²⁶



Amazon’s corporate workforce showed increases in Black and Latinx workers from 2019 to 2020. The percentage of women decreased, however, from 37% to 36.7% that year.²⁶

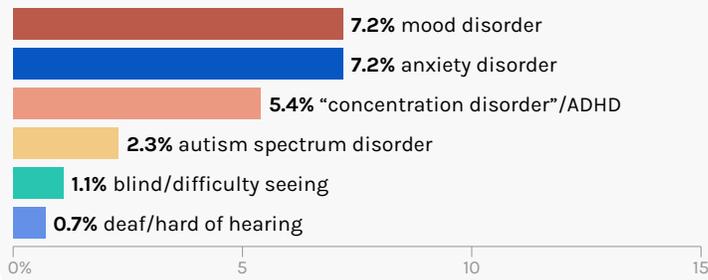
Similar numbers were reported by other companies, such as Adobe, which reported an increase in “underrepresented minorities” from 10.2% to 10.8% between 2019 and 2020.²⁷

Women in the Tech Workforce vs. the Wider Workforce



Although they account for 47% of the workforce in the US, the National Center for Women in Tech reports that only about 26% of tech jobs belong to women. Only about 2% of the tech workforce were Hispanic women, 6% Asian women, and 3% Black women.²⁸ This is compared with the wider US workforce, where about 6% of workers are Hispanic women, 2% are Asian women, and 6% are Black women.²⁹ In 2020, just 22% of CIO positions at Fortune 500 companies were reportedly held by women.³⁰

Developers Reporting a Mental or Physical Difference³¹



Data on tech workers with disabilities is difficult to come by. However, of nearly 65,000 responses to Stack Overflow’s worldwide survey³¹ of developers, 9,532 responses identified as having a mental difference (such as an anxiety, mood, concentration, or autism spectrum disorder) and 1,284 responses identified as having a physical difference (such as difficulty hearing, seeing, standing, walking, or typing).

Diversity in AI

Numbers concerning diversity in AI are generally more difficult to pin down because people working in AI work at both corporations and non-profit organizations across a variety of different product teams as well as in non-technical roles. Among the previous diversity reports mentioned, there are no data specific to their AI teams. Stanford's Human-Centered Artificial Intelligence's (HAI) 2021 AI Index report³² gave some clues as to the percentage of minoritized folks working in AI. For instance, very few computer science (CS) PhDs and tenure-track faculty members are Black (3% and .6% respectively), but the organization Black in AI* has continued to grow. According to the report, fewer than 23% of AI PhD graduates are women. The few specific studies available also suggest that the AI field has much farther to go to reflect the diversity of the members of the population that it affects. About 25% of Queer in AI members describe themselves as "junior industry" and 81% of members characterize the "lack of role models" in the industry as a barrier.

* Founded in 2017, Black in AI aims to "[increase] the presence and inclusion of Black people in the field of AI by creating space for sharing ideas, fostering collaborations, mentorship and advocacy."

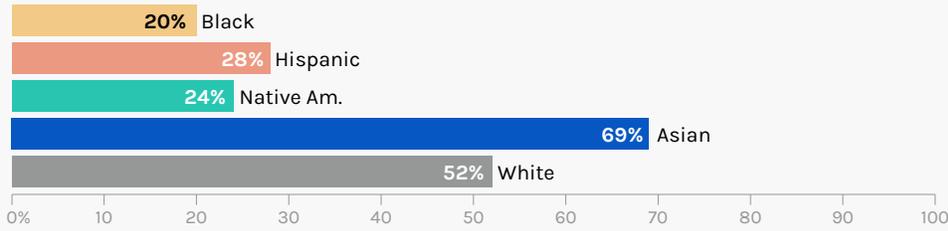
What Has Been Done

Researchers and activists have been pushing for more representation in the field to tackle AI ethics issues for which solutions require diverse perspectives. Some researchers have observed that AI ethics practices were divorced from broader DEI work.³³ Organizations have often kept their DEI teams separate from their AI team areas, with little interaction between the two. This represents a fairly siloed approach to AI ethics, in which content area experts do not weigh in on issues like algorithmic bias and whether technology should be used to automate everything from loan applications to screening applicants for job interviews. To tackle this problem, organizations must move beyond a more tokenistic approach to diversity, doing more than just increasing the pipeline of diverse candidates and increasing the number of minoritized folks interviewed for positions. Instead they must fundamentally audit internal systems to make them more supportive for everyone.

Looking at the Pipeline

Tech companies have long blamed their lack of racial and gender diversity on a lack of diverse candidates from which to recruit.³⁴ Available data in the US show a lack of diversity at all levels between K-12 and college, but do not account for the continued lack of diversity in corporate-level jobs:

Percentage of Children in 4th grade Who Achieved Proficiency in Math³⁵



The National Association for Educational Progress' 2019 Nation's Report Card³⁵ found racial disparities in the percentage of children in the 4th grade who received proficiency in math. 69% of Asian children, for instance, achieved proficiency, compared to just 20% of Black children. These numbers showed a slight increase from previous years. While this is only a proxy to predict future engagement in computer science and tech, available data suggest similar patterns.

In US states that reported these numbers, 12.9% of high school students had disabilities but only 7.6% of high school students who took computer science courses had disabilities.³⁶ Accurate figures are difficult to come by for undergraduate and graduate students in computer science with disabilities, but the Data Buddies Survey reported that about 7.6% of undergraduate CS students and 11.4% of graduate CS students had disabilities.³⁷

High School AP Computer Science Students³⁸



New Ph.D. enrollments in Computer Science³⁹



A report from Code.org reported that only 22% of students taking AP computer science were female and about 13% identified as some racially minoritized identity.³⁸ The report further stated that the higher percentage of minoritized students that comprised a high school, the less likely the school was to offer computer science.

The most recent Taulbee survey from the Computing Research Association (CRA) found that about 23% of new PhD enrollments in computer science identified as female and about 14%

belonged to some minoritized identity and were residents of the US.*³⁹ The CRA's Generation CS report found growth in the number of enrollments in computer science courses at the undergraduate and graduate level between 2006 and 2015.⁴⁰

Figures in the UK mirror those in the US, with female students making up 18% of those enrolled in computer science at the undergraduate level.⁴¹ Data from both BCS⁴² and Inclusive Boards⁴³ indicate about 17% of the broader tech workforce in the UK is female. This is similar to figures of women in tech in Europe overall. No accurate data exist for Europe overall, but estimates point to a similar lack of diversity as that in the US.⁴⁴

Computer Science Undergraduate Students in the UK⁴¹



Tech workers in the UK^{42 43}

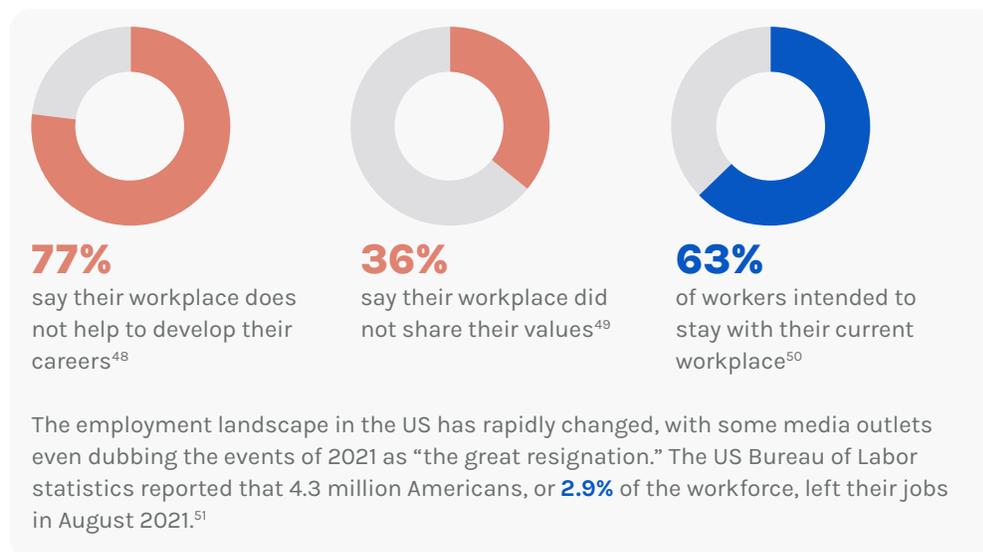


* These percentages belie the true number of minoritized students because they leave out race/ethnicity data for foreign students in the US. However, this recent growth in diversity cannot be seen in the recent numbers in diversity reports from tech companies.

These data suggest that a lack of diversity in the tech pipeline accounts for some lack of diversity in the tech workplace, but there is little evidence that the lack of growth in diversity in big tech companies is mostly due to continued pipeline problems.

What Leads to Attrition?

While no large-scale studies have been done on general reasons that workers leave tech organizations, there are some potential clues in studies of general workplace attrition. These studies point to overall job satisfaction, which is linked to autonomy, psychological well-being, and fair compensation.⁴⁵ Similar results have been found in the IT sector.⁴⁶ Employees also look to characteristics of the workplace such as fair and transparent systems for promotion, and effective managers.⁴⁷ Several recent studies collected data relevant to this:



These findings cannot readily generalize to tech, AI, or other workplaces outside the US, but they might give some clues as to how to approach attrition in the AI workplace.

Attrition in Tech

The Kapur Center's 2017 Tech Leaver's Study⁵² gives the most comprehensive account of attrition focused on tech in the US. The study found that:



Unfairness was the main driver of workers who voluntarily left tech jobs, with **37%** of respondents citing this as the main reason they left their previous employers.



About **78%** of respondents reported that they experienced unfair treatment in their organization.



Different gender and racial/ethnic groups experienced this unfairness differently. For instance, **30%** of Black, Latinx, and Indigenous women (categorized in the report as underrepresented women of color) reported being passed over for a promotion, a difference that was significantly higher than underrepresented men of color and White and Asian women.



Almost **25%** of underrepresented people of color experienced stereotyping in their previous positions and 1 in 10 women reported receiving unwanted sexual attention in their previous jobs.



At **20%**, LGBTQ+ employees were also most likely to report bullying.

Findings from the Women in Data Science and AI project by the Turing Institute provide further support for these findings. The report found that there were gender differences in the types of AI jobs occupied by men and women.⁵³ For instance, women were more likely to occupy jobs that were lower in status and pay and less technical than those occupied by men. It also reported that women left their jobs at higher rates than men left their jobs. Several high-profile cases over the last year, including the firing of Dr. Timnit Gebru and Dr. Margaret Mitchell at Google AI,⁵⁴ and Ifeoma Ozoma and Aerica Shimizu Banks at Pinterest,⁵⁵ have demonstrated many of the hallmarks of these research findings. While there is no comprehensive study about the large-scale attrition of people with disabilities in tech, there is some evidence to suggest a lack of belonging among people with disabilities.⁵⁶

For AI, a field that aims to attract talent that draws from diverse perspectives and backgrounds, it is especially important to intentionally build a positive climate that is welcoming to all identities, since the status quo is likely to prove unwelcome or even hostile for workers belonging to minoritized identities.

Current Study & Methodology

The purpose of the current study was to use qualitative research* to investigate why folks from minoritized backgrounds leave AI teams or organizations, the role of organizational culture or climate in this, and what can be done to stem attrition and make AI teams more inclusive for diverse individuals. It is important to note that recruitment took place primarily in the fall of 2020 and the spring of 2021 during the COVID-19 pandemic.

* For further information on qualitative methods in social science research, see:

[APA Qualitative and Mixed Methods Research What is Qualitative Research? \(PDF\)](#)

Recruitment

The research team created a recruitment document that we then distributed to those in the Partnership on AI's (PAI) professional network through our Partners and other individual collaborators. This recruitment document contained information about the study such as its purpose, format, compensation (\$75) and brief information about data privacy and protections for participants. The full recruitment document can be viewed in Appendix 1. We chose this targeted sampling method because of the specialized nature of the participants.^{57,58}

Participants

The participants in this study fell into 3 broad categories:

1

People who worked on AI teams and who identified with one or more minoritized identities*

2

People who managed AI teams, regardless of how they identified

3

People who worked with DEI in tech organizations

We defined "AI team" broadly, including both people working in "technical" roles such as engineers and data scientists, but also those working in non-technical roles or on AI teams that were non-technical or interdisciplinary such as AI policy or AI ethics teams. Interested participants filled out a screening form. We then contacted those eligible for the study via email with a link to the consent form and a time to schedule the interview via Zoom video teleconferencing. The full privacy document can be viewed in **Appendix 2**.

* We defined minoritized identities to be members from the non-dominant group in their country or within a global context. These minoritized identities were along the lines of race, ethnicity, gender, sexual orientation, and ability.

Measure

The lead researcher used a semi-structured interview protocol developed through the following process:

1. A review of literature pertaining to attrition and DEI in tech and more specific literature focused on AI
2. Scoping calls with workers in DEI in tech
3. Discussions among the research team at PAI
4. A series of 3 pilot interviews
5. Further feedback from the research team

The final research protocol (**Appendix 3**) consisted of 8 questions that focused on 3 research questions, each focusing on one domain:

1. Why do diverse folks leave research teams? (Domain: attrition)
2. What is the culture like on AI teams and organizations? (Domain: culture)
3. What is being done to make these teams and organizations more inclusive? (Domain: efforts for inclusion)

Procedure

1. After signing the consent form, each participant met with the lead researcher via Zoom video conferencing.
2. The lead researcher greeted each participant and explained the purpose of the interview and the parameters of confidentiality and privacy, as well as reminding them of the voluntary nature of the study.
3. The lead researcher then asked for verbal consent and permission to record the interview.
4. The lead researcher proceeded with the questions in the interview protocol, following up on some answers to questions for clarification, asking for additional information or examples, or additional questions based on the responses given.
5. At the end of the interview, the lead researcher thanked the participant and discussed follow-up procedures after the interview. Most interviews lasted between 45 minutes and an hour.
6. After the interview, the lead researcher converted the saved audio file to a transcript and then redacted private or identifiable details from the transcript.
7. Along with payment information, the lead researcher sent this redacted transcript to the participant via encrypted email and asked if they would like to redact or clarify additional information.

Analysis

The researchers used a variation of the consensual qualitative research (CQR) analysis⁵⁹ procedure to analyze the interview transcripts. The lead researcher read through each transcript to redact private, identifying, or sensitive information and to gain basic familiarity with the data. The analysis consisted of the following steps:

1. The lead researcher articulated the 3 research questions to 3 other researchers.
2. The lead researcher articulated the 3 domains under research questions.
3. The lead researcher scanned each participant's transcript and extracted quotes corresponding to each domain.
4. The lead researcher and a second researcher both listed core ideas for each participant's quotes, summarizing this in one paragraph.
5. The third and fourth researchers scanned the summarized core ideas and alerted the lead researcher as to any inconsistencies, while suggesting ways for each core idea to be reconciled.
6. The lead and second researcher grouped these core ideas into broad categories and more specific themes.
7. The lead and second researcher came to a consensus as to themes with which they disagreed.
8. The lead researcher created the final grouping of themes.

Results

DOMAIN: ATTRITION

Managers heavily influenced whether workers stayed or left due to positive or toxic environments

RESEARCH QUESTION 1

Why do women/minoritized folks leave AI teams?

The predominant themes related to attrition that emerged from participant responses were:

1 A toxic work environment

Several factors worked together to create a toxic environment on the AI teams of participants that we interviewed. Some of these factors were more generalizable to other tech or non-tech workplaces, while some uniquely fit the AI ecosystem. For instance, several participants described not getting proper credit for their work and others described a negative sense of competition within teams. One participant said, “It felt like a war zone. Like you weren’t working together to build a product. You were trying to tear other people down to get yourself ahead, which is not, not my mantra.” (F1)* Some participants highlighted the highly competitive nature of AI teams specifically. According to one, “So I hear people aren’t exactly happy on that team and it works in AI because it’s a product team, but [for a] conversational agent. So it’s high pressure. The stakes are high because they have to compete against you know, [AI product name] and other teams like that. But I know that multiple people on that team have taken on medical leave for work, work-related stress and burnout and well.” (F8) Workers stressed that managers heavily influenced whether they stayed or left due to these positive or toxic environments.

Participants also described managers’ approach to building and leading AI teams having a strong impact on whether they stayed or left. Some were on teams that fostered a sense of interdisciplinary collaboration where people from both technical and non-technical backgrounds worked together to build products, explore the implications of AI, and use AI to tackle social problems such as inequality.

Participant F21 said:

“It was really just like one of the big reasons why I left. It was like people making, not allowing me to be fluid and very much being like, ‘Hey, you need to pick a role.’ And that was mostly because of organizational structure. [Because] I was no longer in an interdisciplinary organization.”

* This report denotes each participant with a unique identifier consisting of a letter and a number. The letter corresponds with the type of participant.

F: General participants

M: Managers

D: Those working in DEI roles

Conversely, participant **F19** described reasons for staying:

“What makes me want to stay longer? I think I really resonate with the mission the team is working on. I think I have a good manager and I have a great team. I think it’s, it’s the dynamic. Although, like, there, there are things that can be improved, in terms of team culture and team dynamics. But... I think this is probably above the industry average. I wouldn’t leave at this point and then there are things I want to do. There are researchers that are really, really interesting, really fascinating. They don’t want to continue on, [but] I do, I do think what I’m doing is going to have a positive impact on people and people I care about.”

Environments in which diverse participants thrived tended to encourage exploration of an interdisciplinary approach to AI that incorporated social or sociotechnical issues.

2 Experiences of prejudice

As in other work environments, AI teams were not immune to the effects of prejudice. Participants described microaggressions, overt interpersonal examples of prejudice, and more systemic ways in which their teams and organizations discriminated against them due to their identities.

As one participant who worked within a DEI role stated:

“I think a lot of people burn out and leave as a result, I think that’s one layer. I think the other layer is just, even if you’re not involved in DEI work, being in a product area or even at a company that isn’t particularly diverse means that you’re surrounded by people who in a lot of ways, just aren’t like you, they haven’t had similar life experiences. They don’t know what it means to be you, and they often have a lot of ideas about who you are. And so you’re constantly trying to prove yourself that you’re not, like, a diversity hire that you’re actually meant to be in the room and et cetera.” (D2)

One female participant of color, **F17**, described a case in which:

“It took a fight to have my manager [to] be okay with letting go of a partner because it literally took me and my female coworker for him to sit in on three different meetings and realize how sexist he was. And it was, you know, comments... like, ‘You guys are women. You don’t understand how the space works.’ Or like, ‘Women are better at planning dinners than men, like, you guys should just worry about the menu and like the food.’”

3 A need for growth

Many participants expressed a need for continual growth within the field of AI and left teams or organizations that did not provide pathways for this or those that hindered their professional growth. Managers and AI teams that encouraged this development tended to encourage workers to stay. According to **F28**:

“So I would say the managers and leaders in that previous [REDACTED] company, they were very much about bottom line and customers and delivering good products. They did not pay too much attention to the culture of developing people or retaining them.”

Key Takeaway From This Domain

When taken together, the data suggest that these participants belonging to minoritized identities either left or intended to leave organizations that did not support their continual career growth or had values that did not align with their own. Consistent with this, participants described examples of their organizations not valuing the content of their work. Participants also tied their desire to leave with instances of prejudice or discrimination, which may also be related to a toxic work environment. Some participants reported examples of being tokenized or being subject to negative stereotypes about their identity groups.

Participants belonging to minoritized identities either left or intended to leave organizations that did not support their continual career growth or had values that did not align with their own

RESEARCH QUESTION 2

What are the characteristics of teams where folks do/don't thrive?

The predominant themes related to culture that emerged from participant responses were:

1 Collaborative work environments

Many participants described their current work environments as generally collaborative and collegial. They generally characterized their teammates as willing to help each other. Typically, participants said their previous teams started out this way but eventually started to exhibit toxic aspects to their culture that made them leave. The culture of the teams seemed to depend on the tone set by the manager. According to participants, the managers heavily influenced whether AI teams were inclusive or toxic. **F13** described their own team thusly:

“It’s a pretty personable team. The word ‘family-oriented’ is coming to mind, not so much that my team’s a family, but everyone on the team has families and they’re pretty accommodating of that, which I feel has been not always the case in positions or teams I’ve had, so that stands out... I think my manager has influenced the climate on my team by overdoing empathy and making sure that from the first day that I met her, she tried to get to know me as a person and not an employee and kind of asked, ‘What do you want your work-life balance to look like? What is important to you to work on? Like why do you want to work on this team?’ From then she sometimes will check in, and just ask if she noticed [something unusual] like in a meeting. I’m thinking of a recent time, I kinda got spoken over in a meeting and my manager immediately messaged me after. And she was like, ‘Hey, like, you know, that was a great point you were making and, I heard you.’ So she’s, she’s very proactive, I think with her empathy.”

Other participants emphasized that the collaborative and positive environment felt restricted to their own “bubble” that had been carefully curated by their managers. For instance, **F16** said that:

“I think people see our group as very close together and cohesive because we work on projects together and we communicate all day, and we’re quite separate to the rest of the organization, like just the [REDACTED]. So yeah, it kind of feels like we’re all sometimes in our own little bubble.”

2 Experiences of Prejudice

Several participants described ways in which their AI teams exhibited prejudice, especially female participants discussing specific examples of sexism or ways in which others were allowed to undermine their technical expertise as women.

F9 relayed an experience that several other participants in AI mentioned, specifically:

“There have been times where people, and this is actually not just like, you know, like non-diverse crowd, like a broad spectrum will be like, ‘Oh, are you technical?’ And then I think there has been times when I’ve had to say – and not so much like gender or ethnicity or race or like sexual orientation – but more from like diversity of talent that contributes to the design of AI, which is an area passion of mine where people say like, ‘Ohm so you’re not technical’ or like, ‘Oh, so you’re just the program manager’ in the AI world, and it’s like males and females will say that. And I think there’s definitely this kind of belief that this is a world only for technical people.”

3 Diverse and Inclusive Teams (or lack thereof)

Many participants described the importance of interdisciplinary and identity diversity for fostering a positive team environment. For instance, F14 said:

“So I think my manager is very proactive and sort of leads by example. In larger group meetings or meetings with other folks she’ll point out particular social concerns or point out, you know, who’s not speaking and sort of gently prodding for certain folks to chime in and, or really gracefully getting other people to stop talking. I think also in one-on-one meetings, she’s very upfront about saying, ‘Hey, if this thing is an issue or if anything is an issue, let’s talk about it.’ And for me personally, it helps that she’s a person of color.”

Beyond diverse managers, participants emphasized the importance of managers intentionally creating environments where minoritized workers felt empowered to contribute to their teams using their technical and interdisciplinary skills. As F1 said:

“My old team people would try to steal credit from each other all the time. I had a couple of big things technically stolen from me and someone else got the credit for it. On this team, the managers and the people make sure in every presentation, like, ‘Hey, X, Y, and Z, this person contributed, they did amazing. Here’s all this great work we did because we all did it together.”

One manager described taking time to intentionally create a supportive and inclusive team culture where people with diverse backgrounds could thrive, drawing from diverse perspectives and interdisciplinary expertise to cultivate these qualities. The value of this approach to work was echoed by several comments from participants who found positive AI teams to work on.

Key Takeaway From This Domain

Participants consistently expressed that managers who belonged to minoritized identities or who took the time to learn about working with diverse identities were more supportive of their needs and career goals. This filtered into teams that were also more diverse and inclusive, engendering a positive team culture/climate where folks belonging to minoritized identities thrived. In contrast, there was a recurrent theme of work environments that were sexist or ones where participants experienced acts of prejudice such as microaggressions. This could manifest as decision-making from leaders who did not take folks who were not White into account, or times where credit was misattributed to those in power, instead of female or minoritized folks who did large amounts of the work. While collaborative or positive work environments were also a common theme, these did not negate predominant cultures which deprioritized “DEI-focused” work or work that did not serve the dominant group.

RESEARCH QUESTION 3

What are things that can be done to make organizations/teams more inclusive?

The predominant themes related to efforts to improve inclusivity that emerged from participant responses were:

1 DEI trainings

Participants described a broad range of DEI trainings but few spoke confidently about their effectiveness. In general, trainings that they described as specific received more praise than those that spoke to very general “DEI” topics such as race, culture, or implicit bias.

2 Employee Resource Groups (ERGs)

Many participants mentioned Employee Resource Groups as one aspect of their organizations that effectively promoted inclusivity. These ERGs seemed to serve two purposes as described by the participants. They provided an opportunity for workers from minoritized identities to find fellowship and build community, and they provided a space for workers to find others who practiced interdisciplinary work that incorporated AI with tackling social issues that affected diverse communities.

One participant said that they benefitted daily from ERGs, which acted as valuable spaces for camaraderie, knowledge-sharing, and guidance that foregrounded minoritized identities inside larger organizations that did not. Given their value, the participant thought ERGs should have greater institutional support.

3 Bias against DEI efforts

There still seemed to be a stigma attached to top-down DEI efforts at all levels. Participants described efforts around diversity as not being taken as seriously as more technical topics within AI teams. **F8** said:

“I think that’s just a general, like I’ve noticed that [REDACTED] pretends to a certain values about inclusion and diversity, but at the end of the day, the types of behaviors that get rewarded tend to be the behaviors of the dominant culture.”

On a systemic level, D5 described:

“There are a ton of factors. Certainly there is on the most sort of individual-level interpersonal interactions that are laced with micro- or macro-aggressions. And then all way up to the highest most institutional systemic level. The tech industry is one that has a culture of white supremacy and not in the sense of explicit racism and interpersonal hostility, but one where the communication styles, decision-making styles, resource allocation decisions are all fueled by the predominantly white and Asian male people who have led in the industry for decades, and is not accommodating of a lot of other perspectives, worldviews approaches, needs identities, et cetera. And so there’s a perpetual conflict intention that an individual can feel operating in these spaces. And then, you know, certainly more acute instances of implicit biases leading to impacting hiring decisions or promotion decisions who gets to work on what project, how teams are organized and supported, what work is and isn’t valid, and that may or may not align with the goals, priorities and experiences of people from marginalized identities. And so, yeah, that perpetual tension is one that prevails.”

4 Manager Supports

Ultimately, managers and leadership played a huge role in making their teams more inclusive. Participants tended to attribute much of the success of their teams in creating a diverse and inclusive climate to the intentional actions of their managers. In contrast, participants that complained of toxic environments that failed to foster supports for diverse workers tended to attribute this to the lack of concern from their managers. The managers in the study said that it was at times challenging to maintain a diverse and inclusive culture when the leadership at higher levels was not supportive. Several managers interviewed also referred to the importance of mentoring their reports, especially those who belonged to minoritized identities in the AI field. They reported it was professionally and personally important to them to continue to build an AI community that draws from a variety of disciplines and encourages minoritized workers (both with technical skills and those without) to find their place in this growing field.

5 DEI Hiring practices

Participants typically said that intentional DEI initiatives around hiring were effective. Although they acknowledged that the pipeline into their companies was a separate issue, they mentioned it as being relevant to increasing the inclusive and diverse culture of AI teams in general. F22 said:

“It’s very unfair in the sense that you can work with people, for example, the data science space, but AI space is quite fully occupied by people with masters and PhDs, but you work in corporates whereby you, you get a 25-year-old, 23-year-old woman, who’s got no experience, no degree, no masters. And only because they’re White and they’re well-connected so it’s still an issue. And I must say it’s systematic, it’s, it’s more systematic than the recruitment, I think the recruitment. I mean, there’s a way, or we can use artificial intelligence and machine learning to, deal with these issues.”

Key Takeaway From This Domain

Although many participants reported having to do DEI trainings, these varied widely in terms of how effective they were perceived, with some participants discussing them as having little to no effect and others describing real and observable changes due to these trainings. For instance, several participants discussed DEI trainings that were very specific to some groups (e.g., gender diverse folks, Black people) being the most effective. Participants who mentioned ERGs uniformly praised them, discussing the huge positive impact they have had on them, forming the basis for their social support networks in their organizations. Although this study did not focus on the pipeline, DEI hiring practices, such as recruiting at Historically Black Colleges and Universities, and intentionally diversifying the interview pool of applications, were recurrent themes, potentially related to increasing the diversity within organizations and thus improving climates in general for minoritized folks. Equally recurrent was bias against DEI efforts, with participants discussing that other workers in their organizations would dismiss or discredit these efforts as frivolous or unimportant. Managers also had a huge role to play in the execution of these DEI efforts. Minoritized folks often looked to managers as mentors. The predominant types of mentorship that emerged were:

1. Formal training for mentors,
2. Informal mentorship, often involving minoritized folks looking out for each other, and
3. Mentors supporting more junior workers who belonged to similar identities.

Key Takeaways From Across All Domains

- AI teams which fostered an interdisciplinary and diverse environment, supported by managers and senior leadership, tended to have positive and healthy cultures for workers. These diverse workers – who had both technical and non-technical roles – went into the AI field for a variety of reasons, but often expressed a desire to use their interdisciplinary backgrounds to contribute to social good, especially for minoritized communities, and to grow within their own careers, building products and working on research that advanced innovative and ethical AI methods.
- Other AI teams failed to build this type of environment, instead allowing cultures where:
 1. Microaggressions, sexism, racism, and other prejudice went unchecked,
 2. Teams in which women with technical skills in AI were undermined due to gender, and/or
 3. There was little regard for diversity or interdisciplinary backgrounds, which tended to create the perfect storm for minoritized workers to want to leave the moment their compensation no longer became worth it.
- While organizations and teams have tried numerous efforts to make workplaces more inclusive, such as diversity trainings, the key to building more inclusive spaces seemed to come from diverse leadership that valued a diverse and interdisciplinary AI team. Other efforts such as diversity statements or ERGs seemed to be valuable to minoritized workers, but these alone were not enough to undo systems which did not value them or leadership that did not look like them or understand their perspectives.



The key to building more inclusive spaces seemed to come from diverse leadership that valued a diverse and interdisciplinary AI team

Summary & the Path Forward

This study used interviews with minoritized workers on AI teams, managers of AI teams, and DEI leaders to investigate 3 main questions about workers in AI:

1

Why do minoritized folks leave AI teams?

2

What influence does the culture of these teams have on whether these workers stay or leave?

3

What can be done to make these teams more inclusive?

Approaching these questions using interviews and qualitative analysis created the potential to discover in-depth themes and ask follow-up questions about a topic that has yet to be studied with such specificity. The themes that emerged demonstrated the importance of a strong alignment of an organization's values with the values of the workers on AI teams. Teams which fostered a strong sense of interdisciplinary collaboration, and respect for diverse professional and personal identities, tended to attract and retain diverse individuals. Managers and senior leadership were crucial in maintaining the organizational policies necessary for these teams to thrive. They tended to emphasize the importance of DEI beyond tokenism or superficial trainings mandated by HR.

No easy solution exists to transform an AI team from homogeneous and toxic to diverse and inclusive. Systemic change is difficult and often goes against a company's profit motive. Teams must fundamentally question their norms and values, something which White and middle-class tech workers from prestigious institutions are not used to doing. However, this is necessary to truly make room for diverse perspectives which are necessary to avoid further harm to minoritized populations around the globe.

Since the summer of 2020, companies have redoubled both their internal and public-facing commitments to DEI. The interviews revealed that several of their efforts have been effective, such as supporting ERGs. These are not enough to transform systemically racist or sexist cultures and can in fact backfire if taken over by leaders who do not understand the fundamental mission of DEI.

This interview study is just a starting point from which much more work needs to be done to determine the nuances behind building more inclusive workplaces. In particular, studies specific to more local contexts can better understand some of the specific nuances behind DEI all over the world. The limitations of the study must also be acknowledged. Considering a probable selection bias, the study cannot make claims or generalize towards the experiences of all minoritized individuals on AI teams. Rather, the study gives deep insight into common themes that emerged across a selection of workers who opted into the study.

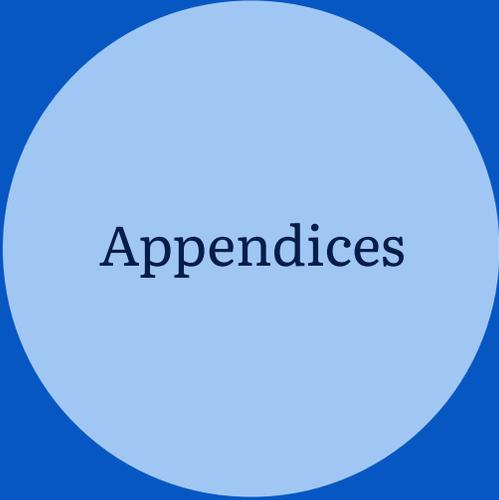
Acknowledgements

This research paper draws its observations from in-depth interviews with more than 40 managers, people working in DEI, and folks who identified as belonging to historically excluded identities working on AI teams. We thank them for their time and insight. Without them, this paper would not be possible.

We would like to thank Margaret Burnett, Danielle Cass, Susan Dray, Allison Druin, and Dominique Mungin for their help scoping the project.

We would also like to thank the many current and former PAI staff members who contributed directly and indirectly to this work. This includes Alice Xiang and Dr. Christine Custis, who oversaw this research as leaders of PAI's Fairness, Transparency, and Accountability Program. It also includes McKane Andrus, Jiyoo Chang, Hudson Hongo, Dr. Tina Park, Madhulika Srikumar, and Jingying Yang, who provided valuable input and assistance throughout the research study and paper-writing process.

Finally, we would like to thank DeepMind, which financially supported this work by providing funding for PAI's Diversity and Inclusion Fellowship.



Appendices

Diversity Equity and Inclusion in AI Interview Study

Overview

You are invited to take part in a study looking at the experiences of diverse professionals working in the AI space. The purpose of the study is to gain insight into the various factors that contribute to an inclusive climate in tech more broadly, and AI teams more specifically. This work is crucial to understanding more about attrition in the field of AI.

Researchers

The study is being conducted by [Jeff Brown](#), Diversity and Inclusion Research Fellow at the [Partnership on AI](#) (PAI), and supervised by [Christine Custis](#), Head of Fairness, Accountability, Transparency and Safety at PAI. Equity and social justice are at the core of PAI's mission, which uses a multistakeholder model to advance these efforts as best practices in the AI field.

How will the Results of the Study Be Used?

The data will be analyzed for themes present, and used to write a research paper, executive summary, and will inform products which aim to increase the inclusivity of teams working in AI. The recommendations made from this study will be disseminated to PAI's partner organizations, who represent a wide array of tech companies and civil society organizations. You can find out more about PAI's mission and Partner organizations here. The researchers will keep the participant informed of progress in the research study.

Eligibility

If you have worked in a tech organization in any of the below roles either currently or in the past 5 years. The study is looking specifically for people who fit any of the following criteria:

- Worked in Diversity, Equity, and Inclusion (or Diversity & Inclusion)
- Managed teams working on projects involving AI
- Worked on a team working on AI projects, and identify as female, non-binary, LGBTQ+, Black, Indigenous, Latinx, Asian, person of color, person with a disability, and/or belonging to minoritized/marginalized identities.

If you are unsure of your eligibility, feel free to email Jeff at jeff@partnershiponai.org

Study Procedure

The study consists of a 45-60 minute interview over Zoom. You will be asked several questions about your experiences working in tech and AI teams, including questions about the climate of teams you have worked on and the organization as a whole. Any personally identifying information gathered from the interview will be redacted from the study.

Privacy and Confidentiality

Data gathered from the screening questionnaire or interviews will not be shared outside the team at PAI. The interviews will be recorded and transcribed. Once the transcribed, recorded video files will be deleted. These files will be password protected, stored on the researcher's PAI issued laptop, and inaccessible to those outside the FTA research team at PAI. Each participant will be sent a copy of their transcript to add to, clarify, or redact any additional information from the interview. Any personally identifiable information, including company names, garnered from interviews will be redacted from transcripts. The final research products will not contain any personally identifiable information, but will rather discuss general themes, and de-identified illustrative examples and quotes. The researcher will contact the participant for any permission to use a quote. Due to the interviewer's status as a mandated reporter, he has a legal duty to report to the appropriate authorities any plans harm to self, others, or instances of child abuse occurring in the last 3 years.

Compensation

Participants will be compensated \$75 for their participation in the interview.

Risks and Benefits

There are no direct risks to participating in this study. However, the interviewer will be asking questions about your experiences working in the tech industry, and may involve topics that are difficult to discuss. Participation is completely voluntary, and you may avoid discussing any topics that you would rather not. There are no direct benefits to participating in this study. However, the aims of the project are to give some insight into the experiences of underrepresented groups in tech and the AI space more specifically.

How to Sign Up

Please fill out [this brief questionnaire](#). You will be contacted by someone at PAI to schedule an interview if you meet the study criteria.

Data Privacy Plan: Diversity, Equity, and Inclusion Study

Researchers

Primary: Jeff Brown, Diversity & Inclusion Research Fellow
Supervisor: Christine Custis, Head of ABOUT ML and FTA
Email: jeff@partnershiponai.org
Phone: [REDACTED]

Communication with Researchers

The researchers and support staff at PAI will communicate with participants primarily through email. You may contact either of the researchers at any time.

Signup, Screening Tool, and Consent Form

Participants will sign up for this study via Qualtrics. First, you will fill out a screening tool to ensure your eligibility for this study. If you are eligible to participate, you will be contacted by a staff member at PAI who will send you a link to a consent form and time to sign up for the study. The screening form will ask for your name, demographic information, and an email address. Read more about Qualtrics data privacy practices [here](#). You may enter a name or pseudonym and email address of your choosing. Any files containing your name or contact information will be stored on the researchers' PAI issued laptops and protected with a password. This information will not be shared with anyone other than the researchers.

Interview

The interview will be conducted on Zoom. Only the participant and researcher(s) will be present. The room will be password protected.

Recording and Transcription

With the consent of the participant, the interview will be recorded so that it can be transcribed using services from temi.com. The recording will be stored on the primary researcher's PAI issued laptop in a password protected folder. Once transcribed, the recording will be deleted. Any personally identifying information on the transcript will be redacted. The redacted transcripts will not be shared beyond the PAI research team.

Research Products

No personally identifiable information, such as individual or company names, will appear on any product stemming from this research, including but not limited to conference papers, journal articles, press releases, executive summaries, or training materials.

Compensation

Compensation will be set up using Bill.com. You may read more about Bill.com's privacy practices [here](#).

APPENDIX 3: RESEARCH PROTOCOL

OVERVIEW

- Welcome participant and introduce self
- Review purpose of the interview/study
- Review how data will be used and safeguards for privacy
- Encourage participant to follow up if they have any questions

DAY BEFORE THE INTERVIEW

- Send reminder email to participant with day/time and zoom link

PROTOCOL

After participant signs on:

Greet participants, and say:

Again, I want to thank you for agreeing to participate in this study. My name is Jeff Brown, a Diversity and Inclusion Research fellow at PAI. The purpose of this study is to look at some of the issues impacting diverse populations in the field of AI. I'll be taking some notes during the interview. Any names or other personally identifying information will be redacted. The transcript won't be shared beyond the research team. This will be kept private and confidential, however, because I'm a mandated reporter, I am legally and ethically obligated to report any instances of potential harm to self or others, or child abuse within the past 3 years. Additionally I'll make sure to give you a copy of the transcript of the interview in case you'd like to clarify any additional details. The video file will be deleted after the transcription is made. It should take about 45 minutes but may be shorter or longer depending on the number of follow up questions that I ask.

I also know that we might be speaking about difficult themes or personal stories, and so you can take a break at any time.

Are there any questions before we begin?

Do I have your permission to record? This will be deleted after it is transcribed.

If yes, start recording. If no, do not begin recording.

Do you consent to taking part in this study?

If yes, begin recording. If no, thank the participant for their time and invite them to follow up by email if they would like to take part in this study or a questionnaire.

Okay, great. Today is [Date/time/location]

Then I'll jump right into the first question:

QUESTIONS

Participant	Interview Questions
DEI Leader	<ul style="list-style-type: none">• What is your position and where? How long have you been working in this position? What are some of your main responsibilities?• From your perspective as [position], what are some of the factors that have led to workers of marginalized identities to leave your organization?<ul style="list-style-type: none">• Query: do you have any examples of this?• Have you worked with people on AI teams specifically? If so, are there any issues that are prevalent in those teams?• What has your organization tried to mitigate this attrition?<ul style="list-style-type: none">• [if they do not know what you mean, query]:• How has the leadership in your organization supported or failed to support your efforts?• What were the results of your intervention?• What are some barriers to implementing efforts to reduce attrition?• Is there anything else that you would like to add?
Manager	<ul style="list-style-type: none">• Tell me about your position. How many employees do you manage? How long? Do you work on projects involving AI?• How would you describe yourself in terms of...• How would you describe your teams in terms of the race/gender/sexual orientation/ or any other known identities?• What are some common challenges in managing a team of people?• What is your experience with mentorship?• Have you received any training specifically in working with or managing people who are of different cultural identities than you?• Has the organization made any steps to make the workplace more inclusive?• Have you taken any steps to make your team more inclusive?• What role has DEI initiatives organization wide played with this?• How effective have these DEI initiatives been?• Have you been treated differently because of any of your identities?• Are there any ways in which you have been made to feel uncomfortable by others at your workplace?<ul style="list-style-type: none">• By your manager?• Other team members?• Can you think of a time when this has happened?• Has any of your reports discussed this with you?• Is there anything else that you would like to add?
Folks	<ul style="list-style-type: none">• Tell me about your position? How long have you been in this position?• How do you identify? You may mention race, sexual orientation, gender, or any other identities you find salient• How would you describe the culture or climate on your team?

- How has the manager influenced this?
 - Has there been anyone in your organization that has helped you to navigate your current role or the organization in general?
 - Are there any ways in which you have been made to feel uncomfortable by others at your workplace/in AI more generally?
 - By your manager?
 - Other team members?
 - Can you think of a time when this has happened?
 - Have you considered leaving your team your team/company/AI? Why?
 - What made you want to stay longer or delay your departure?
 - Have you experienced microaggressions? Other instances of discrimination?
 - Has the wider organization done anything to mitigate these?
 - Is there anything else that you would like to add?
-

Stop recording.

Okay, I have stopped the recording. Thank you again for taking the time to take part in this interview. I will send an email following up with instructions for how to get the compensation for this, as well as an email in the next few days with a transcript of your interview, so you can clarify anything that you would like to. Please let me know via email me if you have any more questions or concerns.

After the interview

Send the participant with an email with instructions on how to obtain the incentive for participating in the study.

Transcribe interview, going over the script to look for any typos etc. Send the transcript to the participant, inviting them to clarify any parts of the transcript or ask additional questions. [create secure system to do this]

APPENDIX 4: IMPORTANT TERMS

The following section contains a clarification of terms as used in this report.

Minoritized Workers	We define minoritized workers as those employees who belong to a minoritized identity. This report used minoritized, as conceptualized by Yasmine Guanaratnam ⁶⁰ instead of the more commonly used “minority” to broadly refer to groups that do not fall under members of the dominant group for a given identity axis. This emphasizes the institutional power dynamics of the dominant or “majority” group regardless of actual proportional representation.
Artificial Intelligence Teams	The study broadly construed AI teams to include work focused on a broad range of techniques under the umbrella of AI, whether it is focused on engineering, research, policy, ethics, or other aspects of AI. These roles may thus focus on specific AI techniques like machine learning or natural language processing, or may focus on broader implications of AI. The study focuses on AI organizations outside of university settings, although these include both for-profit and non-profit organizations.
Race	Race commonly refers to the “[physical characteristics or] differences that groups determine to be significant”. ⁶¹ This has been commonly conceived as related to phenotypical skin color, and commonly linked with ethnicity. Racial categories are thus defined by different societies in different ways, and thus someone’s racial category may change depending on their broader social context. This report defines racial categories similar to how they are defined in the US census, which is necessarily a limited framework. To begin to account for this, the study asked open-ended questions for participants to self-identify, so that they could define their own racial categories.
Racism	The study treats racism as the systemic and institutionalized, race based form of oppression perpetrated by dominant racial groups to minoritized racial groups. The report acknowledges however that the common definition of racism includes or emphasizes interpersonal racial discrimination regardless of power dynamics. Because racialized institutions and racial dynamics differ between and sometimes even within societies, some racial groups may benefit from or be subject to racism depending on their wider social context. However, racial institutions such as White supremacy sometimes span different societies due to a more globalized world.
Ethnicity	This report uses the definition of ethnicity as “shared social, cultural, and historical experiences, stemming from common national or regional backgrounds.” ⁶² This is sometimes but not always synonymous with race (e.g. Asian-American). This study asked participants to self-identify, and some chose to write their ethnicity, while others did not.
Microaggression	This report uses the definition of microaggression as “brief and commonplace daily verbal, behavioral, or environmental indignities, whether intentional or unintentional, that communicate hostile, derogatory, or negative racial slights and insults.” ⁶³ This has also been broadened to include slights or insults based on other axes of identity.

DEI Focused Work This is work that prioritizes principles of diversity, equity, and inclusion within an organization, including but not limited to how socially marginalized groups can access full participation and benefits traditionally privileged to members of the dominant group. For instance, DEI work has focused on reducing gender and race based discrimination, affording equal opportunity to minoritized groups, and increasing the representation of minoritized groups since the status quo privileges the dominant groups.

Ability The report uses the definition of disabled as stated within the Americans with Disabilities Act. That is “A physical or mental impairment that substantially limits one or more major life activities, a record of such an impairment, or being regarded as having such an impairment.” As with other identity categories, participants were asked to self-identify ability status.

Notes

- 1 Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.
- 2 Zhao, D., Wang, A., & Russakovsky, O. (2021). Understanding and Evaluating Racial Biases in Image Captioning. *arXiv preprint arXiv:2106.08503*.
- 3 Feldstein, S., 2021. *The Global Expansion of AI Surveillance*. [online] Carnegie Endowment for International Peace. Available at: <<https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847>> [Accessed 17 September 2019].
- 4 Firth, N. (2021). *Apple Card is being investigated over claims it gives women lower credit limits*. MIT Technology Review. Retrieved 23 November 2021, from <https://www.technologyreview.com/2019/11/11/131983/apple-card-is-being-investigated-over-claims-it-gives-women-lower-credit-limits/>.
- 5 Howard, A., & Isbell, C. (2021). *Diversity in AI: The Invisible Men and Women* | Ayanna Howard and Charles Isbell. MIT Sloan Management Review. Retrieved 21 September 2020, from <https://sloanreview.mit.edu/article/diversity-in-ai-the-invisible-men-and-women/>.
- 6 AI Now. (2019). *Discriminating Systems: Gender, Race, and Power in AI* [Ebook]. Retrieved 23 November 2021.
- 7 Swauger, S. (2021). *Opinion | What's worse than remote school? Remote test-taking with AI proctors*. NBC News. Retrieved 7 November 2020, from <https://www.nbcnews.com/think/opinion/remote-testing-monitored-ai-failing-students-forced-undergo-it-ncna1246769>.
- 8 Belani, G. (2021). *AI Paving the Way for Remote Work* | IEEE Computer Society. Computer.org. Retrieved 26 July 2021, from <https://www.computer.org/publications/tech-news/trends/remote-working-easier-with-ai>.
- 9 Scott, A., Kapor Klein, F., & Onovakpuri, U. (2017). *Tech Leavers Study* [Ebook]. Retrieved 24 November 2021, from <https://www.kaporcenter.org/wp-content/uploads/2017/08/TechLeavers2017.pdf>.
- 10 *Women in the Workplace 2021*. (2021). Retrieved 23 November 2021, from <https://www.mckinsey.com/featured-insights/diversity-and-inclusion/women-in-the-workplace>.
- 11 Silicon Valley Bank. (2021). *2020 Global Startup Outlook: Key insights from the Silicon Valley Bank startup outlook survey* [Ebook]. Retrieved 23 November 2021, from https://www.svb.com/globalassets/library/uploadedfiles/content/trends_and_insights/reports/startup_outlook_report/suo_global_report_2020-final.pdf.
- 12 Firth, N. (2021). *Apple Card is being investigated over claims it gives women lower credit limits*. MIT Technology Review. Retrieved 23 November 2021, from <https://www.technologyreview.com/2019/11/11/131983/apple-card-is-being-investigated-over-claims-it-gives-women-lower-credit-limits/>.
- 13 Tomasev, N., McKee, K.R., Kay, J., & Mohamed, S. (2021). Fairness for Unobserved Characteristics: Insights from technological impacts on queer communities. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21)*, Retrieved October 1, 2021 from <https://doi.org/10.1145/3461702.3462540>.
- 14 Martinez, E., & Kirchner, L. (2021). *The secret bias hidden in mortgage-approval algorithms* | AP News. AP News. Retrieved 24 November 2021, from <https://apnews.com/article/lifestyle-technology-business-race-and-ethnicity-mortgages-2d3d40d5751f933a88c1e17063657586>.
- 15 Turner Lee, N., Resnick, P., & Barton, G. (2021). *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*. Brookings. Retrieved 24 November 2021, from <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>.
- 16 Rock, D., & Grant, H. (2016). Why diverse teams are smarter. *Harvard Business Review*, 4(4), 2-5.
- 17 Wang, J., Cheng, G. H. L., Chen, T., & Leung, K. (2019). Team creativity/innovation in culturally diverse teams: A meta-analysis. *Journal of Organizational Behavior*, 40(6), 693-708.
- 18 Lorenzo, R., Voigt, N., Tsusaka, M., Krentz, M., & Abouzahr, K. (2018). *How Diverse Leadership Teams Boost Innovation*. BCG Global. Retrieved 24 November 2021, from <https://www.bcg.com/publications/2018/how-diverse-leadership-teams-boost-innovation>.
- 19 Hoobler, J. M., Masterson, C. R., Nkomo, S. M., & Michel, E. J. (2018). The business case for women leaders: Meta-analysis, research critique, and path forward. *Journal of Management*, 44(6), 2473-2499.
- 20 Chakravorti, B. (2020). *To Increase Diversity, U.S. Tech Companies Need to Follow the Talent*. Harvard Business Review. Retrieved 24 November 2021, from <https://hbr.org/2020/12/to-increase-diversity-u-s-tech-companies-need-to-follow-the-talent>.
- 21 Accenture. (2018). *Getting to Equal 2018: The Disability Inclusion Advantage*. Retrieved from https://www.accenture.com/_acnmedia/pdf-89/accenture-disability-inclusion-research-report.pdf
- 22 Whittaker, M., Alper, M., Bennett, C. L., Hendren, S., Kaziunas, L., Mills, M., ... & West, S. M. (2019). Disability, bias, and AI. *AI Now Institute*.
- 23 Heater, B. (2020). *Tech companies respond to George Floyd's death, ensuing protests and systemic racism*. Techcrunch. Retrieved 24 November 2021, from <https://techcrunch.com/2020/06/01/tech-co-protests/>.
- 24 Google (2021). *2021 Diversity Annual Report*. Retrieved 24 November 2021, from https://static.googleusercontent.com/media/diversity.google/en/annual-report/static/pdfs/google_2021_diversity_annual_report.pdf?cachebust=2e13d07.

- 25 Facebook. (2021). *Facebook Diversity Update: Increasing Representation in Our Workforce and Supporting Minority-Owned Businesses* | Meta. Meta. Retrieved 24 November 2021, from <https://about.fb.com/news/2021/07/facebook-diversity-report-2021/>.
- 26 Amazon Staff. (2020). *Our workforce data*. US About Amazon. Retrieved 24 November 2021, from <https://www.aboutamazon.com/news/workplace/our-workforce-data>.
- 27 Adobe. (2021). *Adobe Diversity by the Numbers*. Adobe. Retrieved 24 November 2021, from <https://www.adobe.com/diversity/data.html>.
- 28 National Center for Women in Tech. (2020). *NCWIT Scorecard: The Status of Women in Computing [2020 Update]*. Retrieved <https://ncwit.org/resource/scorecard/>.
- 29 Center for American Progress (2012). *The State of diversity in Today's workforce*. Retrieved from <https://www.americanprogress.org/article/the-state-of-diversity-in-todays-workforce/>.
- 30 Gillenwater, S. (2020). *Meet the CIOs of the Fortune 500 – 2021 edition*. Boardroom Insiders. Retrieved from <https://www.boardroominsiders.com/blog/meet-the-cios-of-the-fortune-500-2021-edition>.
- 31 Stack Overflow. (2020). *2020 Developer Survey*. Retrieved from <https://insights.stackoverflow.com/survey/2020#developer-profile-disability-status-mental-health-and-differences>.
- 32 Stanford HAI. (2021). *The AI Index Report: Measuring Trends in Artificial Intelligence* [Ebook]. Retrieved 24 November 2021, from https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report-_Chapter-6.pdf.
- 33 Chi, N., Lurie, E., & Mulligan, D. K. (2021). *Reconfiguring Diversity and Inclusion for AI Ethics*. *arXiv preprint arXiv:2105.02407*.
- 34 Selyukh, A. (2016). *Why Some Diversity Thinkers Aren't Buying The Tech Industry's Excuses*. NPR. Retrieved 24 November 2021, from <https://www.npr.org/sections/alltechconsidered/2016/07/19/486511816/why-some-diversity-thinkers-arent-buying-the-tech-industrys-excuses>.
- 35 National Association for Educational Progress. (2020). *NAEP Report Card: Mathematics*. Retrieved from <https://www.nationsreportcard.gov/mathematics/nation/achievement/?grade=4>
- 36 Ladner, R. (2021). *Expanding the pipeline: The status of persons with disabilities in the Computer Science Pipeline*. Retrieved February 1, 2022, from <https://cra.org/cra-wp/expanding-the-pipeline-the-status-of-persons-with-disabilities-in-the-computer-science-pipeline/>.
- 37 Center for Evaluating the Research Pipeline (2021). *"Data Buddies Survey 2019 Annual Report"*. Computing Research Association, Washington, D.C.
- 38 Code.org. (2021). *Code.org's Approach to Diversity & Equity in Computer Science*. Code.org. Retrieved 24 November 2021, from <https://code.org/diversity>.
- 39 Zweben, S., & Bizot, B. (2021). *2020 Taulbee Survey: Bachelor's and Doctoral Degree Production Growth Continues but New Student Enrollment Shows Declines* [Ebook]. Computing Research Association. Retrieved 24 November 2021, from <https://cra.org/wp-content/uploads/2021/05/2020-CRA-Taulbee-Survey.pdf>.
- 40 Computing Research Association (2017). *Generation CS: Computer Science Undergraduate Enrollments Surge Since 2006*.
- 41 The Higher Education Statistics Agency (2021). *Higher Education Student Statistics*. Retrieved from <https://www.hesa.ac.uk/news/16-01-2020/sb255-higher-education-student-statistics/subjects>.
- 42 BCS. (2014). *Women in IT Survey* [Ebook]. BCS: The Chartered Institute for IT. Retrieved 24 November 2021, from <https://www.bcs.org/media/4446/women-it-survey.pdf>.
- 43 Inclusive Boards. (2018). *Inclusive Tech Alliance Report 2018* [Ebook]. Retrieved 24 November 2021, from <https://www.inclusivetechalliance.co.uk/wp-content/uploads/2019/07/inclusive-Tech-Alliance-Report.pdf>.
- 44 Atomico. (2020). *The State of European Tech 2020*. 2020. stateofeuropeantech.com. Retrieved 24 November 2021, from <https://2020.stateofeuropeantech.com/chapter/diversity-inclusion/article/diversity-inclusion/>.
- 45 Chung-Yan, G. A. (2010). The nonlinear effects of job complexity and autonomy on job satisfaction, turnover, and psychological well-being. *Journal of occupational health psychology*, 15(3), 237.
- 46 McKnight, D. H., Phillips, B., & Hardgrave, B. C. (2009). Which reduces IT turnover intention the most: Workplace characteristics or job characteristics?. *Information & Management*, 46(3), 167-174.
- 47 Vaamonde, J. D., Omar, A., & Salessi, S. (2018). From organizational justice perceptions to turnover intentions: The mediating effects of burnout and job satisfaction. *Europe's journal of psychology*, 14(3), 554.
- 48 Instructure (2019). *How to get today's employees to stay and engage? Develop their careers*. PR Newswire. Retrieved from <https://www.prnewswire.com/news-releases/how-to-get-todays-employees-to-stay-and-engage-develop-their-careers-300860067.html>.
- 49 McCarty, E. (2021). *Integral and The Harris Poll Find Employees are giving Employers a Performance Review - Integral*. Integral. Retrieved 24 November 2021, from <https://www.teamintegral.com/2021/news-release-integral-employee-activation-index/>.
- 50 McCarty, E. (2021). *Integral and The Harris Poll Find Employees are giving Employers a Performance Review - Integral*. Integral. Retrieved 24 November 2021, from <https://www.teamintegral.com/2021/news-release-integral-employee-activation-index/>.
- 51 Bureau of Labor Statistics. (2021). *News Release - The Employment Situation - October 2021* [Ebook]. Retrieved 24 November 2021, from <https://www.bls.gov/news.release/pdf/empstat.pdf>.

- 52 Scott, A., Kapor Klein, F., & Onovakpuri, U. (2017). *Tech Leavers Study* [Ebook]. Retrieved 24 November 2021, from <https://www.kaporcenter.org/wp-content/uploads/2017/08/TechLeavers2017.pdf>.
- 53 Young, E., Wajcman, J. and Sprejer, L. (2021). *Where are the Women? Mapping the Gender Job Gap in AI*. Policy Briefing: Full Report. *The Alan Turing Institute*.
- 54 Metz, C. (2021). *A second Google A.I. researcher says the company fired her*. *Nytimes.com*. Retrieved 24 November 2021, from <https://www.nytimes.com/2021/02/19/technology/google-ethical-artificial-intelligence-team.html>.
- 55 Myrow, R. (2021). *Pinterest Sounds A More Contrite Tone After Black Former Employees Speak Out*. *Npr.org*. Retrieved 24 November 2021, from <https://www.npr.org/2020/06/23/881624553/pinterest-sounds-a-more-contrite-tone-after-black-former-employees-speak-out>.
- 56 Scheer, S. (2021). *The Tech Sector's Big Disability Inclusion Problem*. *ERE*. Retrieved from <https://www.ere.net/the-tech-sectors-big-disability-inclusion-problem/>
- 57 Robinson, O. C. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative research in psychology*, 11(1), 25-41.
- 58 Yancey, A. K., Ortega, A. N., & Kumanyika, S. K. (2006). Effective recruitment and retention of minority research participants. *Annu. Rev. Public Health*, 27, 1-28.
- 59 Hill, C. E., Knox, S., Thompson, B. J., Williams, E. N., Hess, S. A., & Ladany, N. (2005). Consensual qualitative research: An update. *Journal of counseling psychology*, 52(2), 196.
- 60 Gunaratnam, Y. (2003). *Researching 'race' and ethnicity: Methods, knowledge and power*. Sage.
- 61 Race and Ethnicity. American Sociological Association. (2022). Retrieved 29 January 2022, from <https://www.asanet.org/topics/race-and-ethnicity>. Archived at <https://web.archive.org/web/20190821170406/https://www.asanet.org/topics/race-and-ethnicity>.
- 62 University of Minnesota Libraries (2022). 10.2 *The Meaning of Race and Ethnicity*. *Open.lib.umn.edu*. Retrieved 29 January 2022, from <https://open.lib.umn.edu/sociology/chapter/10-2-the-meaning-of-race-and-ethnicity/>.
- 63 Sue, Derald Wing, Christina M. Capodilupo, Gina C. Torino, Jennifer M. Bucceri, Aisha Holder, Kevin L. Nadal, and Marta Esquilin. "Racial microaggressions in everyday life: implications for clinical practice." *American psychologist* 62, no. 4 (2007): 271.