Attrition of Workers with Minoritized Identities on AI Teams

Jeffrey Brown*
San Diego State University
jbrown4@sdsu.edu

Tina Park
Partnership on AI
tina@partnershiponai.org

Jiyoo Chang
Partnership on AI
jiyoo@partnershiponai.org

Mckane Andrus
University of Washington HCDE
mkandrus@uw.edu

Alice Xiang
Sony AI
alice.xiang@sony.com

Christine Custis
Partnership on AI
christine@partnershiponai.org

ABSTRACT

The effects of AI systems are far-reaching and affect diverse communities all over the world. The demographics of AI teams, however, do not reflect this diversity. Instead, these teams, particularly at big tech companies, are dominated by Western, White, and male workers. Strategies for preventing harms done by AI must also include making these teams more representative of the diverse communities that these technologies affect. The pipeline of students from K-12 and university level contributes to this - those with minoritized identities are underrepresented or excluded from pursuing computer science careers. However there has been relatively little attention given to how the culture at tech companies, let alone AI teams, contribute to attrition of minoritized people in the workplace. The current study uses semi-structured interviews with minoritized workers on AI teams, managers of AI teams, and leaders working on diversity, equity, and inclusion (DEI) in the tech field (N = 43), to investigate the reasons why these workers leave these AI teams. The themes from these interviews describe how the culture and climate of these teams may contribute to attrition of minoritized workers, and strategies for making these teams more inclusive and representative of the diverse communities affected by technologies developed by these AI teams. Specifically, the current study found that AI teams in which minoritized workers thrive tend to foster a strong sense of interdisciplinary collaboration, support professional career development, and are run by diverse leaders who understand the importance of undoing the traditional White, Eurocentric, and male workplace norms. These go beyond the “quick fixes” that are prevalent in DEI practices.

KEYWORDS

Diversity, DEI, Racism, Sexism, Attrition

ACM Reference Format:
Jeffrey Brown, Tina Park, Jiyoo Chang, Mckane Andrus, Alice Xiang, and Christine Custis. 2022. Attrition of Workers with Minoritized Identities on AI Teams. In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22), October 06–09, 2022, Arlington, VA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3551624.3555304

*Corresponding author.

EAAMO '22, October 06–09, 2022, Arlington, VA, USA

ACM ISBN 978-1-4503-9477-2/22/10…$15.00
https://doi.org/10.1145/3551624.3555304

1 INTRODUCTION

AI teams have failed to reflect the diverse communities their technologies ultimately affect and, in some cases, harmed. Although AI as a field has begun to reckon with the harms done to [1]minoritized or marginalized communities, 2020 saw an unprecedented increase in the number of organizations speaking out against racial justice. The murders of George Floyd, Breonna Taylor, Daunte Wright, and Ahmaud Arbery forced the United States and other countries all over the world to come to terms with the legacy of historic harms meted on entire communities of people based on race and intersections therein with other axes of identity such as ability, gender identity, and sexual orientation. In response to these murders, subsequent public outcry, and demands from their employees, organizations released statements and pledged to increase the diversity of their teams, among other efforts to show an attempt to pursue racial equity [18]. Despite this, AI teams have still reflected the broader tech ecosystem in its makeup of mostly White, male workers [29].

Organizations have long focused on recruiting more diverse candidates for positions, especially those from marginalized or minoritized groups. However, relatively less effort has been focused on how the homogenous demographics of these teams have also influenced minoritized individuals to leave these teams. The current paper will briefly describe the crucial need for improving the inclusivity of AI teams, present the results of an interview-based study with minoritized individuals on AI teams to ask them why they have left or continue to stay on AI teams, and propose recommendations for making these teams more inclusive.

2 RATIONALE

The rationale to study the attrition of minoritized workers in the AI field is three-fold. First, the harms associated with AI are disproportionately borne by historically minoritized communities. Buolamwini and Gebru's seminal work *Gender Shades* powerfully demonstrated how existing societal biases can be encoded in algorithms, in this case, bias in classifying the faces of Black women [4]. Several researchers have since shown how bias can be encoded in other domains including hiring, mortgage approval, and approval for credit lines [23, 24]. Second, the people working on AI do not

[1]This paper uses the term minoritized, coined by Gutnaratnam [15] to emphasize the active minoritization or marginalization process meted upon groups with less institutional power than the dominant groups. This is elaborated upon in the methods section. Terms such as “minority” does not accurately capture this.

represent the communities whom this technology affects. A prime example of this is in surveillance technologies being deployed in a wide range of geographic and cultural contexts. While surveillance and facial recognition technology may exemplify these far-reaching effects, there are countless examples of how places use AI technology for everything from multiple fields including education, health care, criminal justice, and the workplace [12, 14, 19,20, 31]. Third, organizations have insufficiently tried to address attrition, and instead have focused much of their efforts at recruiting a more diverse pipeline of workers.

2.1 Attrition Is Understudied

While both challenges are intertwined and crucial to building better tech companies, the attrition question is under-studied. There is less work on what happens once people belonging to minoritized identities enter these organizations that lack the diverse and inclusive culture necessary for their success. The high profile firings of Dr. Timnit Gebru and Dr. Meg Mitchell from Google, and Ifeoma Ozoma and Aerica Shimizu Banks from Pinterest have highlighted this issue over the past two years, although previously the Kapor Center's comprehensive tech leavers' study gave a thorough account of why minoritized tech workers left [15]. No study has previously investigated attrition of diverse AI workers in general, much less from a qualitative angle to probe these reasons for attrition, and to explore potential ways to make these spaces more inclusive.

3 BACKGROUND

3.1 Problems Due to Lack of Diversity

The scope of this paper includes tech organizations that are both for-profit and nonprofit. While there are no specific statistics available for tech companies, they likely reflect the numbers in broader corporations in the US in which there is very little diversity in the C-suite, including 20% White women, 4% women of color, and 13% men of color [32]. This lack of diversity in leadership does not bode well for the transformation of AI teams to be more diverse and inclusive and may help explain the harms that have been attributed to AI algorithms [34].

There are also psychological bases for innovation brought about by diverse teams. First, increased heterogeneity would bring differences in perspectives. These perspectives may better disrupt the comfort and assumption of the White heterodoxy that ignores the complexities of cultural differences more broadly [5, 9]. Second, this disrupts the threat of groupthink and conformity bias [21]. Beyond and perhaps more important than the business case for it, more diversity and more inclusive practices would be much more consistent with the stated values of tech organizations and AI teams within them.

3.2 Diversity in Tech and AI

Stated efforts to improve diversity, especially in hiring, have generally, slightly improved how diverse the tech workforce is. Companies such as Google, Facebook, Amazon, and Adobe all showed very slight increases in the percentage of workers from minoritized groups over the past few years, but still short of goals, and far short of numbers close to representative of the general population [1, 2, 7, 11, 16]. For instance, in 2020, Amazon's corporate workforce were comprised of 7.2% Black people, 7.5% Latinx people, and .5% Native American people [2]. In 2021, 4.4% of Facebook's workers were Black [11]. More broadly, the National Center for Women in Technology (NCWT) reported that only about 26% of the tech workforce identified as women [3]. No current, accurate data exist to pinpoint exact levels of diversity on AI teams, but Stanford's Human-Centered Artificial Intelligence (HAI) 2021 AI Index Report reported that only 19% of computer science Ph.D. graduates were women, and only about 3% of Computer Science (CS) Ph.D. graduates were Black [30]. Part of the challenge in addressing the lack of diversity in AI teams thus remains finding accurate data to quantify the extent of it. To do this would require a comprehensive set of studies that carefully teases out the very complex nature of diversity, equity, and inclusion.

As previously mentioned, efforts have focused on increasing the pipeline of workers recruited to tech companies to diversify AI teams. Global data on this are especially imprecise due to varying definitions of what would constitute a minoritized identity in societies with different social constructions of identity[2]. A few studies could give us some clues as to why pipeline continues to be an issue. In 2021, Code.org reported that of the students taking AP computer science, 22% identified as female and 13% identified as some racially minoritized identity [7]. In the US, the 2020 Taulbee Survey reported that 27% of new Ph.D. students in CS identified as female, and about 20% identified as some minoritized racial identity and were residents of the US [35]. There is no definitive data on the worldwide tech pipeline, although researchers suggest that the pipeline and workforce are no more representative.

3.2.1 DEI Interventions. It is also difficult to comprehensively characterize the range of DEI interventions that companies have applied. Among the most frequently occurring ones is DEI trainings aimed to increase employee awareness and proactively prevent harm brought about by discriminatory practices. 2020 saw renewed attention placed on DEI trainings along with additional scrutiny. The effectiveness of DEI trainings would vary widely because of contextual variation of workers and how these trainings were applied. Kim and Roberson's [21] review on DEI trainings that focused on implicit bias for instance, found that they were generally effective at educating employees about the concept of bias but not necessarily effective at facilitating deeper discussions about how to dismantle it. Intersecting with the concept of embeddedness, Employee Resource Groups (ERGs) are typically employee-led organizations within wider organizations meant to facilitate social connections between employees, usually centering around some aspect of identity or diverse representation. Like DEI trainings, overly general or broad proclamations about the effectiveness of ERGs would not support the very contextualized and diverse nature of ERGs as they exist between industries and even within organizations. However,

[2]The methods section elaborates on this more, but the very conception of Diversity, Equity, and Inclusion (DEI) as is understood in White and Eurocentric societies such as the UK or the USA centers around the marginalization of identities that have been historically oppressed by European and patriarchal colonialism and imperialism. Thus, even though the intergroup power dynamics between and within groups necessarily differ between and even within some societies, the authors of this paper acknowledge that this frame for DEI used to frame the study is necessarily limiting.

the basic structure of ERGs as they are generally designed, position them well to support minoritized workers [13].

3.3 Factors that Lead to Attrition

Within the wider context of worker attrition, factors such as psychological well-being, autonomy, and fair compensation are associated with workers leaving their jobs [6]. This varies widely by field, however, and building a straightforward model or explanation of why workers stay - sometimes constructed as embeddedness, or worker retention - would necessarily be too simplistic. For instance, Ng et al. [24] found that corporate social responsibility (CSR) of a company was associated with greater levels of pride in an organization and that this was in turn associated with workers wanting to stay. However, broadly, job embeddedness models draw a link between the degree to which a worker feels connected to their workplace and their immediate team and supervisor, how aligned the organization's goals are to their own personal or professional goals, and what the worker stands to lose if they were to leave [10]. For minoritized workers, this sense of connectedness and alignment with the organization becomes negatively affected by perceived prejudice or discrimination, something that has been widely documented for women, racially minoritized individuals, sexually minoritized individuals, and those with disabilities [26, 27, 28].

3.3.1 Previous Work on Attrition. The Kapor Center's 2017 tech leaver's study focused on attrition in tech for minoritized workers and found that unfair treatment was a primary reason that workers gave when they voluntarily left their organization. For instance, 30% of women who identified as either Black, Latinx or indigenous, reported that they were unfairly passed over for a promotion. 10% of women reported receiving unwanted sexual attention, and 20% of LGBTQ+ workers reported bullying [29]. These numbers were significantly higher than for groups who did not belong to minoritized identities. The Turing Institute's Women in Data Science and AI project further found that within AI specifically, women were more likely to occupy lower paying jobs and roles that were less technical in scope [33]. No previous studies have addressed attrition of workers in AI specifically, but efforts to increase inclusion and stem attrition should address the above. That is, organizations should specifically target how to create a workplace that is psychologically safe, and one where workers of all identities are treated fairly. This will help to ensure that once organizations succeed in recruiting more diverse workers, they are entering into environments that do not push them out.

3.4 Individual vs. Systemic Change

One additional question is the "what now?" after identifying potential reasons for attrition of minoritized workers in tech companies and AI teams. Previous work has suggested short-term, individual measures such as implicit bias trainings are of limited efficacy [21]. Other work points to the need for more systemic solutions. Within the scope of this paper, we define systemic solutions as those that address individuals, teams, organizational leadership, and policies, as well as the interdependence of these various levels of operation. Thus, systemic change may consist of hiring policies, along with training opportunities and changes in leadership. Within the context of diversifying tech workforce and leadership, much research has examined this problem but has noted that little has changed [25]. Indeed, the most important individual change might be at the leadership level [15], but this has not yet been thoroughly examined in tech companies.

4 CURRENT STUDY

The purpose of the current study was to use qualitative research to investigate why workers 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. The current study thus consisted of 3 broad research questions each corresponding with a "domain" or broad category that guided the interview. The research questions were:

- Why do diverse workers leave research teams (domain - Attrition)?
- What is the culture like on AI teams and organizations (domain - Culture)?
- What is being done to make these teams and organizations more inclusive (domain - Efforts for Inclusion)?

4.1 Recruitment

The research team created a recruitment document that we then distributed to those in [REDACTED] professional network through our partners and other individual collaborators. This recruitment document contained information about the study such as its purpose, format, compensation (US$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.

4.2 Participants

The 40 participants in this study fell into 3 broad categories:

- Category 1 included people who worked on AI teams and who identified with one or more minoritized identities. 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. The term minoritized also refers to power differentials, whereby dominant members of a group hold more institutional power within a social context. While some of these groups may possess institutionalized power across several contexts (e.g. White cisgender men, cisgender heterosexuals), it is possible for someone to have an identity that only has institutionalized power in some social contexts (e.g. White cisgender women) or have some identities that are part of the dominant social group and some that are minoritized (e.g. a cisgender man with a disability). This is in line with Crenshaw's [8] conception of intersectionality.
- Category 2 included people who managed AI teams, regardless of how they identified. organizations.

Table 1: - Number of Participants by Demographic and Role*

Race	
Asian	11
Black	8
Hispanic	2
Latino	3
Mixed	3
White	12
Gender	
Female	25
Male	12
Gender Queer or Non-Binary	3
Disability Status	
Does Not Have Disability	30
Has Disability	9
Prefer Not to Say	1
Place Where Team is Based	
Europe	6
USA	30
Africa	2
Asia	2
Organization Type	
Governmental Organization	1
Non-Profit	4
For Profit Company (excludes startups)	30
Startup	5
Role	
Engineer or Research Scientist	15
AI Ethics or Policy Role	6
Data Scientist	4
Manager	6
DEI Role	6
Other	6

*Some categories in this table are not mutually exclusive and add up to a total of more than 40

- Category 3 included people who worked with DEI in tech 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 directors or an analyst on AI ethics teams. Interested participants filled out a screening form, and we contacted those eligible for the study via email with a link to the consent form and a time to schedule the interview via zoom. All fields in the screening form were open-ended and prospective participants could self-identify. The demographics and professional roles of the total number of participants interviewed are provided in Table 1.

4.3 Measure

The lead researcher used a semi-structured interview protocol developed through the following process: The lead researcher conducted a review of literature pertaining to attrition and DEI in tech, and more specific literature focused on AI. 2) The lead researcher conducted a series of scoping calls with workers in DEI in tech, discussions among the research team, a series of 3 pilot interviews, and 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. Because the approach was semi-structured, the interviewer sometimes changed the order of questions and follow-up queries to questions depending on how the participant answered previous questions.

4.4 Procedure

After signing the consent form, each participant met with the lead researcher via Zoom. The lead researcher greeted each participant, explained the purpose of the interview and the parameters of confidentiality and privacy, and reminded them of the voluntary nature of the study. The lead researcher then asked for verbal consent and permission to record the interview. 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. 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 to an hour. After the interview, the lead researcher converted the saved audio file to a transcript and redacted the transcript for private or identifiable details. 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.

4.5 Analysis

The researchers used a variation of Hill's [19] consensual qualitative research analysis (CQR) procedure to analyze the interview transcripts. Originally designed for use with research in psychotherapy, the CQR framework encourages the qualitative researcher to use multiple coders (described below as "researchers") to analyze blocks of text from interviews, guided by the overarching domains or research questions. CQR approaches qualitative coding from a constructivist perspective, in that it assumes that each participant holds their own version of truth which guides how they interpret the experiences around them. CQR also consists of multiple coders discussing their codes and several layers of audits, because it acknowledges that each coder approaches the procedure with bias. The current study employs a version of CQR consistent with Hill's [19] approach where there are no pre-existing codes, and thus the codes emerge based on answers the participants give. A full description of the study's analysis procedure follows.

The lead researcher first read through each transcript to redact private, identifying, or sensitive information and to gain basic familiarity with the data. The subsequent analysis consisted of the following steps:

- The lead researcher articulated the 3 research questions to the other 3 researchers
- The lead researcher articulated the 3 domains under research questions
- The lead researcher scanned each participant's transcript, and extracted quotes corresponding to each domain
- The lead researcher and a second researcher both listed core ideas for each participant's quotes, summarizing this in one paragraph.
- 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
- The lead and second researcher grouped these core ideas into broad categories and more specific themes
- The lead and second researcher came to a consensus as to themes with which they disagreed
- Lead researcher did final grouping of themes and broader categories

5 RESULTS

This section will describe the main themes that emerged from the interviews, as determined by how well they answered the research questions and whether these themes tended to recur within in the main ideas of participant responses.

5.1 Overview of Results

The study found that, within these 3 domains, workers generally reported responses that fit into these categories and coded with the following themes. Generally, workers left their teams for reasons characterized by aspects of the teams' culture and climate. Within these broader categories they most mentioned being in a "toxic" environment, their own desire to grow their careers, and a lack of systemic supports available as reasons that they left their teams. When probed further about aspects of team culture and climate, they broadly discussed experiences of prejudice and the team culture surrounding diversity and other DEI practices. When asked about aspects of DEI, they mentioned ERGs, specific DEI trainings, and DEI hiring practices most commonly. Appendix A.4 shows all categories, themes, example main ideas (paraphrased amalgams to be more illustrative) and relative frequencies. The following 3 sections describe how some of the more prevalent themes addressed each of the 3 research questions.

5.2 Attrition - Why Do Minoritized Workers Leave AI Teams?

The three biggest factors leading workers to leave AI teams fell into the broader category of the Culture and climate of their teams or organizations, their own desire to make an impact in the AI field, and the amount of systemic supports for them.

5.2.1 Culture and Climate - A Toxic Work Environment and Instances of Prejudice. Workers discussed a "toxic" or negative work environment, including experiences of prejudice, as especially influential in their decisions to leave AI teams or organizations altogether. Partially these experiences could generalize to other industries but there were some answers that were more specific to tech and AI. For instance, one participant said:

> "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**)

Several participants gave examples of what else contributed to this toxic work environment, and what they did when things became untenable for them. One DEI leader who worked to intervene and encourage positive environments for workers said:

> "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**)

Discussions like this suggest that even when minoritized workers, particularly workers of color, are on product teams at tech companies, they often face behaviors from employees and overall cultures that ultimately proved to be hostile or unwelcoming.

5.2.2 Systemic Supports for Workers - A Need for Growth. Workers discussed one of the primary motivations for joining an AI-focused team being the opportunity to join a constantly evolving field. Participants mentioned that one determining factor for whether they wanted to leave or stay on their teams was whether their team organization and manager facilitated their professional growth. They discussed this growth in skills and autonomy especially within the context of their unique contributions to the field as opposed to doing the job to make more money for their organization. This aligns well with the research on embeddedness. Although the interviews did not seek to deconstruct the reasons why these workers prioritized their own professional development and growth, the general sense that participants in the study gave was that their connectedness to their teams and to a certain extent the field of AI hinged upon their ability to grow with the field.

5.3 Culture – General culture and climate, and DEI practices and attitudes

The interviewer asked workers about the aspects of their work environment to get more specific information about what aspects led to them wanting to stay or leave their teams or for some, leaving their organizations. The themes describing their responses fell within the above 2 general categories, while overlapping with some of the themes discussed under Attrition. Participants described both positive and negative aspects of their teams. The primary positive aspect of team culture that these workers mentioned was a willingness to collaborate. Participants generally attributed these positive or collaborative environments to their supervisors. This also intersected with participants' desire for growth within their career or field, and how they felt their ideas were perceived on AI teams. The following quote from a female member of an AI team was relatively typical of how workers described their positive work environment:

> "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 very proactive, I think with her empathy." (F13)

5.3.1 Experiences of Prejudice. As in the section above, workers elaborated on ways in which instances of prejudice sometimes colored the climate of their team in a negative way. This was particularly common from workers who identified as female. For instance, participant F9 described:

> "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, 'Oh 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."

It was typical for participants to describe ways in which either they or their colleagues had experienced instances of interpersonal or institutional prejudice or discrimination. These were both overt, such as colleagues mentioning harmful racial or gender stereotypes, and covert, which they experienced as somewhat insidious or hidden but still harmful remarks or actions. Some participants described what they experienced as microaggressions, and others described feeling discriminated against because of some parts of their identity, for instance their ethnicity but not their gender expression, or vice versa.

5.3.2 Diverse and Inclusive Teams. Several participants also mentioned the diversity of their teams, both in terms of personal identity (e.g., race ethnicity) and in terms of professional discipline (e.g. engineer, social scientist). Participants discussed both sometimes intersecting, such that they appreciated when managers and teams encouraged them to bring their own diverse perspectives to their everyday work, even if this forced teams to grow beyond the comfort of the norms of the dominant group.

Participants often attributed how diverse and accepting teams were to their managers and/or other leaders in the organization. One participant (F14) expressed:

> "So I think my manager is very proactive and sort of leads by example. In larger group meetings or meetings with other workers she'll point out particular social concerns or point out, you know, who's not speaking and sort of gently prodding for certain workers 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."

One manager interviewed also described how they intentionally tried to create a culture that valued diversity across multiple dimensions on teams. The participant described how diverse their team was in terms of race, ethnicity, and gender, but also discussed how they proactively recruited and nurtured the careers of social scientists who became well integrated within the group and added to its methodological diversity.

5.4 Efforts to Improve Inclusivity

The supports described also include inclusive practices that were either structured and formal or more informal practices. More structured (and structural or systemic) practices included DEI programs, ERGs and DEI trainings, but also reform around hiring practices. More informal practices included informal support networks between employees. ERGs were especially common among practices that employees discussed.

5.4.1 ERGs. Participants frequently mentioned ERGs as one aspect of their organizations that effectively promoted inclusion and belonging, specifically when they acted to build community between workers from minoritized identities. Within the AI field, they sometimes provided a forum for workers to collaborate with each other when they were interested in exploring questions of AI and its intersection with minoritized social identities. Several participants described relying on ERGs for providing support on various DEI issues that they would not necessarily turn to their immediate team or managers to discuss. Some participants also discussed that although their ERGs were mostly grassroots, they had a lot of financial support from their organization which benefitted the growth of these ERGs tremendously.

5.4.2 DEI Trainings and Manager Supports. Many participants discussed formal DEI trainings in their organizations. Although they were commonplace, participants were not confident about their effectiveness. There seemed to be a very wide range of trainings, but usually participants described the ones that were more specific, for instance, ones that spoke to working with specific groups, as more effective than ones that spoke to more general topics such as implicit bias and race in general. As discussed previously, participants tended to attribute whether they stayed in or left positions to how supportive their manager was, especially with regards to DEI issues, and how they intersected with AI. Some managers interviewed spoke to their intentional focus on mentorship. Female managers interviewed characterized mentorship as especially crucial to their success and continual professional development in the field of AI, and the need to protect their reports from much of the sexism they experienced not only in their workplaces but in professional organizations and conferences as the AI field began to grow rapidly and others tried to claim credit for their work or exclude them from the very spaces they helped to build.

5.4.3 DEI Hiring Practices. Although this spoke less directly to attrition of people in AI, participants reported that DEI policies around hiring were very common as an example of systemic, structural change initiated by leadership. These hiring practices were usually enforced by the Human Resource departments and required a certain proportion of candidates invited for interviews or screened into a candidate pool to belong to gender or racial minoritized. In several interviews, participants associated these with part of the solution to creating a more inclusive environment for the team. Examples of DEI hiring practices mentioned by participants included having a minimum number of applicants in finalist positions from minoritized groups, increased recruitment at historically Black Colleges and Universities (HBCUs), and structured mentorship programs designed to recruit and support minoritized workers.

6 DISCUSSION

This interview-based study sought to answer 3 research questions about AI teams and organizations. These questions were 1) Why do minoritized workers leave AI teams? 2) What is the culture or climate on these teams? And 3) How can we make these teams more inclusive? Through interviews with minoritized workers, managers, and DEI leaders, the study identified some underlying themes that could answer these questions.

6.1 Most Common Themes

The most common themes suggest that several aspects of these teams come together to create a toxic environment. Prejudice, and especially sexist behaviors contribute to these toxic environments, in which minoritized workers would have to struggle to get deserved recognition. The desire for organizations to grow rapidly and outcompete other products added to this toxic environment. This climate left little room for norms other than the dominant, male, and White-centric values to grow and thrive. As some participants discussed, these values sometimes went unstated and were assumed by members of the dominant groups as the correct way to work. While this could generally explain why workers would leave tech companies more broadly, the resulting themes from the interviews point to a particular effect on AI teams. Specifically, minoritized workers come to AI teams seeking space to develop their skills and ideas in a way that the interdisciplinary space of AI allows. Minoritized workers expressed a strong desire for personal and professional growth, as they brought their own life experiences and identities into their work on AI teams, as they witnessed AI technologies potentially harming marginalized communities like their own.

6.2 Where Minoritized Workers Thrived

The teams on which minoritized workers tended to thrive emphasized the importance of workers coming together from different backgrounds, especially those from minoritized backgrounds. On these teams, guided by managers who intentionally created these spaces, workers could bring their "full selves" to work, not having to worry about the stresses of struggling to fit in with frustrating experiences brought about by the dominant culture disregarding their own experiences. Along with managers who could either relate to navigating the AI field as a member of a minoritized group or who intentionally made inclusiveness a priority. Teams that did not prioritize these seemed to create environments passively or intentionally where minoritized workers were further marginalized, through sexist behaviors, assumptions of not being "technical enough", disregarded as focusing too much on diversity, or otherwise not getting credit for the work that they are doing. Consistent with the literature on the pervading heterodoxy of Whiteness or other dominant structures, a failure to intentionally carve out these spaces for diverse individuals gives way to the status quo of negative experiences that minoritized workers have faced.

There remains no "quick fix" for teams or organizations to easily change their workplace from toxic to inclusive. However, these results suggest that prioritizing diversity beyond surface-level interventions like general DEI trainings, and instead focusing on shifting values away from the norms that only include the status quo of dominant White and male groups could yield some positive effects. As discussed in the introduction, the most important individual level change may be at the leadership level – the participants discussed their climate as heavily influenced by top-down policies that they ultimately had little say over. The participants' answers support the argument that they must ultimately follow the policies handed to them, or risk losing their job. The results, taken together, suggest that effectively inclusive leaders and managers focus on helping their employees grow and develop in the field of AI rather than a singular focus on profits and competition. Minoritized workers in the field of AI may have a need to find a community that would also help each other to grow personally and professionally as they navigate through a system that does not always recognize or value them.

7 IMPLICATIONS AND RECOMMENDATIONS

7.1 Continued Support of ERGs

The study suggests that there are several measures that organizations and AI teams could take to be more inclusive for minoritized workers. Workers discussed needing a sense of community with other minoritized workers who could also relate to their professional path in AI. The study participants' discussions of their reliance on informal networks of people in their organizations to help them navigate unclear paths to promotion and professional growth underscored this. For this reason, further support of ERGs could serve an especially important function. As described by participants, ERGs seemed to have a balance of an official community to which workers could turn to meet likeminded individuals, but still generally informal enough to where workers could shape the direction of these ERGs.

7.2 Build Diverse Teams Beyond Superficial Representation

Organizations must resist building teams that are only superficially diverse, without any regard for how they expect the norms and values of minoritized workers on these teams to assimilate to the White, ableist, and cisnormative values of the wider organization. Inclusive and supportive AI teams will instead take a much more proactive approach to building team norms that reflect everyone's values, by "leading with empathy" as one participant put it. This may not be possible without a diverse team of leaders who have taken the time to understand the importance of a more thorough approach to building diverse and inclusive teams. However, consistent with the literature on team diversity and innovation, "surface level" diversity is not enough, and instead, leaders must take the time to understand the underlying reasons for building AI teams who not only are more representative, and who can think critically about the types of communities their technologies impact. More specific to AI, teams should take the time to recruit workers from multiple disciplinary backgrounds in addition to diverse identities. Both technical and "non-technical" AI teams would benefit from these interdisciplinary climates that recognize the power structures underlying the differential harm meted out by AI technologies.

7.3 Overhaul DEI Trainings

Instead of general and overly broad DEI trainings that workers may only sometimes remember, organizations should prioritize trainings that name and address more specific populations or challenges specific to these teams. For instance, if there has been resistance to or a failure to meet DEI hiring goals, trainings should attempt to get at the underlying reason why. Typically, trained DEI leaders would be able to identify these issues, but there may be power structures that prevent them or prevent minoritized workers from pointing these issues out. Until there are avenues for these workers to challenge these institutional norms, especially as they affect their work in AI, organizations can only hope for superficial or symbolic change from their DEI trainings.

8 LIMITATIONS AND FUTURE DIRECTIONS

While the qualitative nature of the study got in-depth insights into the 3 main research questions - those of attrition, culture, and efforts to improve inclusion - this sacrificed gaining a broad or representative view. Instead, the results are a specific but highly contextualized snapshot of the field. Future study can employ survey-based methods to attempt to gain a broad sense of how these themes exist in a wider cross-section of minoritized workers. Regardless, the voluntary nature of these studies necessarily garner participation from a subset of workers who would be more motivated to discuss the challenges of the AI field and potential ways to improve it. Still, organizations should aim to increase inclusion for these workers particularly. Because of the limitation of time, any subset of questions - such as those about mentorship programs or the value of interdisciplinary collaboration - could be explored with greater depth in future research. This study also acknowledges that the umbrella of "minoritized worker" cuts across an incredibly large and heterogeneous variety of professional and personal identities. Thus, this study is only a starting point for others that could probe with much greater depth the experiences of workers with specific or intersecting identities.

ACKNOWLEDGMENTS

This study took place primarily through the Partnership on AI's Diversity and Inclusion Research Fellowship. We would like to sincerely thank all our study participants and pilot study participants, without whom this research would not be possible. The lead author would like to especially thank his colleague Jingying Yang for her tremendous vision and work in establishing the study goals and design. We would like to thank Madhu Srikumar, Rebecca Finlay and the PAI staff for their support. We would like to thank Margaret Burnett, Danielle Cass, Susan Dray, Allison Druin, and Dominique Mungin for their help scoping the project. Finally, we would like to thank DeepMind for funding this work through PAI's Diversity and Inclusion Research Fellowship.

REFERENCES

[1] Adobe. (2021). *Adobe Diversity By the Numbers*. Adobe. Retrieved 24 November 2021, from https://www.adobe.com/diversity/data.html.

[2] Amazon Staff. (2020). *Our workforce data*. US About Amazon. Retrieved 24 November 2021, from https://www.aboutamazon.com/news/workplace/our-workforce-data.

[3] Ashcraft, C., McLain, B., & Eger, E. (2017). *Women in Tech: The facts* [Ebook]. National Center for Women in Tech. Retrieved 24 November 2021, from https://wpassets.ncwit.org/wp-content/uploads/2021/05/13193304/ncwit_women-in-it_2016full-report_final-web06012016.pdf.

[4] 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.

[5] Cave, S., & Dihal, K. (2020). The whiteness of AI. *Philosophy & Technology*, *33*(4), 685-703.

[6] 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.

[7] Code.org. (2021). *Code.org's Approach to Diversity & Equity in Computer Science*. Code.org. Retrieved 24 November 2021, from https://code.org/diversity.

[8] Crenshaw, K. (1990). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stan. L. Rev.*, *43*, 1241.

[9] Danaher, K., & Branscombe, N. R. (2010). Maintaining the system with tokenism: Bolstering individual mobility beliefs and identification with a discriminatory organization. *British Journal of Social Psychology*, *49*(2), 343-362.

[10] EhtİYar, R., & Ersoy, A. (2018). Job Embeddedness: A Ten Year Literature Review. *Social Sciences Researches in the Globalizing World*, 657

[11] Facebook. (2021). *Facebook Diversity Update: Increasing Representation in Our Workforce and Supporting MinorityOwned Businesses | Meta*. Meta. Retrieved 24 November 2021, from https://about.fb.com/news/2021/07/facebook-diversityreport-2021/.

[12] 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].

[13] Foldy, E. G. (2019, July). Employee Resource Groups: What We Know about Their Impact on Individuals and

[14] Organizations. In *Academy of Management Proceedings* (Vol. 2019, No. 1, p. 10633). Briarcliff Manor, NY 10510: Academy of Management

[15] 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-investigatedover-claims-it-gives-women-lower-credit-limits/.

[16] Google (2021).*2021 Diversity Annual Report*. Retrieved 24 November 2021, from https://static.googleusercontent.com/media/diversity.google/en//annualreport/static/pdfs/google_2021_diversity_annual_report.pdf?cachebust=2e13d07.

[17] Gunaratnam, Y. (2003). Researching 'race' and ethnicity: Methods, knowledge and power. Sage.

[18] Heater, B. (2020). *Tech companies respond to George Floyd's death, ensuing protests and systemic racism*. Techcrunch.com. Retrieved 24 November 2021, from https://techcrunch.com/2020/06/01/tech-co-protests/.

[19] Hill, C. E. (2012). Consensual qualitative research: A practical resource for investigating social science phenomena. American Psychological Association

[20] Karimi-Haghighi, M., & Castillo, C. (2021, June). Enhancing a recidivism prediction tool with machine learning: effectiveness and algorithmic fairness. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law* (pp. 210-214)

[21] Kim, J. Y., & Roberson, L. (2021). I'm biased and so are you. What should organizations do? A review of organizational implicit-bias training programs. *Consulting Psychology Journal: Practice and Research*.

[22] Margetts, T., & Holland, E. (2017). The Case for Group Heterogeneity. In *Discrimination and Diversity: Concepts, Methodologies, Tools, and Applications* (pp. 1852-1871). IGI Global

[23] 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-mortgages2d3d40d5751f933a88c1e17063657586.

[24] McMahon, B. T., Roessler, R., Rumrill, P. D., Hurley, J. E., West, S. L., Chan, F., & Carlson, L. (2008). Hiring discrimination against people with disabilities under the ADA: Characteristics of charging parties. *Journal of Occupational Rehabilitation*, *18*(2), 122-132.

[25] Ng, T. W., Yam, K. C., & Aguinis, H. (2019). Employee perceptions of corporate social responsibility: Effects on pride, embeddedness, and turnover. *Personnel Psychology*, *72*(1), 107-137.

[26] Rooney, K., & Khorram, Y. (2020). *Tech companies say they value diversity, but reports show little change in last six years*. CNBC. Retrieved from https://www.cnbc.com/2020/06/12/six-years-into-diversity-reports-big-tech-has-made-little-progress.html

[27] Sanz, E. (2014). Exploring Stereotype Threat in the Workplace with Sexual Minorities

[28] Sears, B., & Mallory, C. (2011). Documented evidence of employment discrimination & its effects on LGBT people.

[29] 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.

[30] 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.

[31] 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*

[32] 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.

[33] Women in the Workplace 2021. (2021). Retrieved 23 November 2021, from https://www.mckinsey.com/featuredinsights/diversity-and-inclusion/women-in-the-workplace.

[34] 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*.

[35] 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