

A feeling for the algorithm: Diversity, expertise, and artificial intelligence

Catherine Stinson and Sofie Vlaad

Published in *Big Data & Society*, January – March 2024. DOI: 10.1177/20539517231224247

Abstract

Diversity is often announced as a solution to ethical problems in artificial intelligence (AI), but what exactly is meant by diversity and how it can solve those problems is seldom spelled out. This lack of clarity is one hurdle to motivating diversity in AI. Another hurdle is that while the most common perceptions about what diversity is are too weak to do the work set out for them, stronger notions of diversity are often defended on normative grounds that fail to connect to the values that are important to decision-makers in AI. However, there is a long history of research in feminist philosophy of science and a recent body of work in social epistemology that taken together provide the foundation for a notion of diversity that is both strong enough to do the work demanded of it, and can be defended on epistemic grounds that connect with the values that are important to decision-makers in AI. We clarify and defend that notion here by introducing emergent expertise as a network phenomenon wherein groups of workers with expertise of different types can gain knowledge not available to any individual alone, as long as they have ways of communicating across types of expertise. We illustrate the connected epistemic and ethical benefits of designing technology with diverse groups of workers using the examples of an infamous racist soap dispenser, and the millimeter wave scanners used in US airport security.

Keywords

Philosophy of artificial intelligence, philosophy of science, diversity, expertise, feminist philosophy, emergent expertise

1. Introduction

After decades of activism by marginalized workers pushing for workplace equity in tech companies and computer science departments, the message that a more diverse workforce might be beneficial has started to gain some traction. In Artificial Intelligence (AI) in particular, the suggestion that more diversity could help solve *ethical* problems has also become commonplace (Beard, 2018; Paul, 2019). There remains, however, considerable skepticism about diversity initiatives, both from supporters of greater equity in the workplace, who worry that diversity initiatives are empty promises (Fussell, 2021), and from people in AI and tech who see diversity as a threat to scientific excellence or engineering goals (Gershgorn, 2019). Diversity remains a contested goal, particularly in AI.

In this paper we examine arguments for and against increasing diversity in AI, with an eye to finding arguments for diversity that both point the way to social change and speak to the concerns of AI practitioners. There are notions of diversity robust enough to spur change while also supporting scientific excellence, or so we will argue here. The practical reality remains that the people holding the balance of power in AI may have financial or other motivations to resist diversity initiatives. What we offer here is an analysis of diversity, and arguments for its benefits that reckon with the concerns motivating this resistance. Concrete advice on how to implement policy changes within business or research settings is beyond the scope of this paper.

Section 2 outlines the state of conversation about diversity in AI. We begin with a public conversation about the value of diversity in technology workplaces found in a Twitter thread about what has been dubbed a “racist soap dispenser”. We then examine documents expressing support for or backlash against diversity from within technology companies, and finally review the academic literature about diversity. One thing that quickly becomes clear is that normative arguments for diversity, based on values like fairness or the intrinsic value of all people, are a hard sell among AI practitioners.

In Section 3 we frame the case for diversity in AI instead as an epistemic argument, drawing on work in feminist philosophy of science, and social epistemology. In feminist philosophy of science, we find the insight that including people with marginalized backgrounds and experiences in a scientific community expands the variety of cognitive resources available for noticing unexamined assumptions and coming up with novel ideas. We consider how intersectional feminism complicates this picture by understanding the intersection of structures of domination as non-additive. In social epistemology we look to agent-based simulations showing that diverse groups of problem-solvers outperform groups

of high-performing individuals under a variety of conditions. Together these suggest that networks of people with diverse kinds of expertise should have access to greater epistemic benefits than networks with more homogeneous expertise. When diversity is understood as a relationship that applies across whole teams of scientists, rather than applying to particular individuals, we might avoid the stigma of labelling some workers as “diversity hires”.

In Section 4 we illustrate the dynamic, networked nature of scientific expertise with examples from community involvement in AIDS research and psychiatry, where networks composed of people with technical knowledge and people with lived experience work together to create knowledge that individuals could not have produced working in isolation. We then apply our notion of emergent expertise to the “racist soap dispenser” example and airport screening technologies, showing how a lack of standpoint expertise within design teams leads to technologies that fail to do the jobs for which they were designed. What we reject when we embrace diversity is not scientific excellence, but rather the myth of the lone genius as the prototype of success in AI.

2. The State of Diversity in AI

That AI needs more diversity is frequently announced in headlines, but the discussion under these headlines often fails to specify what is meant by diversity, or to offer arguments for the claim that diversity improves outcomes (whether ethical or scientific).¹

What is well established is that AI has “a white guy problem” (Crawford, 2016). Despite decades of investment in women’s participation in computing, the gender gap has

¹ What is meant by AI is equally fraught. Here we follow contemporary usage by including as part of AI any technological products that include machine learning algorithms, or products thereof, as well as research (both academic and industry) that draws on topics one might find in an AI textbook. When discussing workplace dynamics we broaden the discussion to include the academic field of computer science and the technology sector more generally.

grown in recent decades, not shrunk, unlike in other natural science fields (Cheryan, et al., 2016). According to a recent report by West et al. (2019), engineering degrees going to Black women declined 11% from 2010 to 2015, only 10-15% of AI researchers at major firms are women, and women's attrition rate from technology careers is close to 50%. Furthermore, only 2-4% of researchers at major AI firms are Black, and 3-6% are Latinx. 80% of AI professors are men (West et al., 2019). A former diversity recruiter for Google alleged (see Curley, 2020) that until 2014 Google had never hired workers from Historically Black Colleges and Universities for technical roles (Alcorn, 2021). The culture of AI (which broadly generalizes to academic computer science and the tech industry) lionizes lone genius figures, devalues non-technical expertise (Forsythe, 1993), and has earned a reputation for being unsafe for both women (Hicks, 2013) and racial minorities (Birhane and Guest, 2020). There is no data available on participation by gender minorities or disabled people.

Whether these facts point to a lack of diversity depends on what we mean by diversity. Steel et al. (2018) distinguish several notions of diversity. What they call *egalitarian* diversity is maximized when an attribute is uniformly distributed across a group. If the AI workforce included equal numbers of each race, gender, religion, etc., egalitarian diversity would be maximized.

What Steel et al. (2018) call *normic* diversity is present to the extent that the members of a group diverge from a "non-diverse norm". If that norm were defined, for example, as white men, the AI workforce would be maximally diverse if all its members were not white, or not men. We are not aware of diversity initiatives that seek to exclude any normic group from the workforce, though some reactions against diversity seem to assume this as a goal. Note that a group high in normic diversity could simultaneously be low in egalitarian diversity, for example if the group consisted exclusively of Asian women.

A third notion Steel et al. (2018) describe is *representational* diversity. A group is representationally diverse to the extent that the distribution of relevant attributes in the group matches the distribution in a reference population, like the residents of a country, or people with computer science degrees. Female, Black, and Latinx are all attributes that have been documented to be underrepresented in AI compared both to their proportions in the general US population, and to their proportions among computer science graduates.

In evaluating arguments for and against diversity it is important to keep these differences in mind. For example, complaints that women's representation in the field is too low could be interpreted in terms of any of these three notions of diversity. It could be interpreted as meaning that less than 50% women is too little (assuming egalitarian diversity and erasure of non-binary identities), that less than 100% is too little (assuming normic diversity, where men are taken as the norm), that less than 52% is too little (assuming representational diversity, where the general population is taken as the reference population), or that less than 37% is too little (assuming representational diversity, where the reference population is the highest historical rate of women's representation among computer science undergraduate degrees earners (ComputerScience.org, 2022). Fazelpour and De-Arteaga (2022) illustrate in detail how mismatches between rationales for diversity, mechanisms of its functional benefits, and measures of diversity can be clarified.

This ambiguity in the term diversity allows for misunderstanding and leaves the end goals of diversity initiatives unclear. Egalitarian diversity would be difficult to achieve where some attributes are rare in the general population. This could make it seem like an unreasonable or unfair goal to aim for. However, egalitarian diversity might be useful as a goal worth working closer toward, even if there is no expectation of reaching it as a destination. Normic diversity, like egalitarian diversity, might be used more as a course

corrective, than as a desired destination. For instance, normic diversity initiatives might be a way to create safe spaces that shift power relations, while working toward a world where those safe spaces are no longer needed.

Below we look to online discourse about diversity, internal documents leaked from tech companies, and testimony from diversity workers in AI for perceptions of diversity among AI practitioners. From these sources, we can see that some arguments in favour of diversity have not been well articulated, and that uptake of the idea has been poor in AI and related fields.

2.1 Normative Diversity and Diversity as Representation

A viral video posted to Twitter (Afigbo, 2017) has become a popular talking point in discussions of diversity in technology. Media coverage of the video calls its subject a “racist soap dispenser” (Hale, 2017). In the video, a white man gets soap from an automatic soap dispenser, then a Black man tries to do the same. He moves his hand beneath the machine for some time, to no effect. But if he puts a piece of white paper towel under the soap dispenser, soap is immediately dispensed. The caption reads, “If you have ever had a problem grasping the importance of diversity in tech and its impact on society, watch this video” (Afigbo, 2017).

However, the many replies show that the importance of diversity is not clear from this video. Some claim that the video is a hoax, that the man is holding his hand at the wrong angle, or that lasers are interfering with the sensor. Many assume it is an isolated case of a badly calibrated machine, ignoring the replies by people of colour (POC) saying that they have noticed the same thing, and by engineers saying that it is a known problem with these devices.

The connection between faulty design and diverse representation in the workplace is made more explicit in replies claiming that if the company employed dark-skinned people, the problem would have been discovered before the product was released. Others retort that anyone who knew to test on diverse skin tones could discover the problem. Some commenters admit that diversity might help in this particular case, but don't accept the generalization to the entire field, or disagree on *how* diversity ought to help. Many see the problem purely as a technical failure. Several people suggest that talk of diversity is a politically motivated attack on the status quo, or express exasperation at the suggestion that an inanimate object could be racist (contra Liao and Huebner, 2020; Stinson 2022; Winner, 1980).

Two main types of arguments for diversity seem to be on the table. First, the group objecting that the problem is a technical one, not an ethical one, see engineering problems as possessing purely technical solutions, for which the social identities of engineers are irrelevant. The argument for diversity is thus seen as an appeal for parity in hiring that is at best unconnected to technical outcomes, and at worst a plea to promote unqualified people. As Fehr (2011) notes, in science and technology fields, making a successful case for diversity typically requires spelling out benefits for the field itself (and the people entrenched in it). If diversity is presented as a normative appeal for fairness, and there is no engineering case to be made for it, diversity initiatives can look like pandering to political correctness. While arguments can be made for a rights-based approach to hiring where groups ought to be represented regardless of what impact it makes on technical outcomes, AI practitioners generally do not share these values, so normative arguments fall flat.

Second, many of the sympathetic replies can be interpreted as saying that the point of diversity is to place representatives of marginalized communities in a position to advocate

for their own concerns in the design of technologies, like noticing that a soap dispenser doesn't work for them. For applications like the soap dispenser and facial recognition technology, where skin tone is highly relevant, having POC on the engineering teams is a plausible way of avoiding error. It is common practice in computer science labs for researchers to first test their programs on themselves. The three most cited research papers about face recognition according to a GoogleScholar search all feature the authors' faces in the test images (Ahonen et al., 2006; Turk and Pentland, 1991; Wright et al., 2008). Through mechanisms like this, less diverse workplaces lead directly to less diverse datasets and products that don't work well for some users. This is a common phenomenon, from language systems that don't recognize regional accents (Wheatley and Picone, 1991) or minority dialects (Blodgett and O'Connor, 2017), to kitchen counters installed too high for the people doing the cooking (Avakian, 2005). As West et al. (2019: 6) argue, "issues of discrimination in the workforce and in system building are deeply intertwined."

However, if acting as a token representative were the only defense of diversity, it would be a weak one. The connection between identity markers and the functioning of AI is not always as direct as in a device that reflects light off skin. Most AI does not involve visual detection of humans.² Furthermore, identities can be diced very finely. A small startup company may not have enough employees to cover even the groups protected in civil rights legislation. It would be impractical to employ children. Employment status is a protected category in some places, but employing unemployed people poses a logical difficulty, and involving unemployed people in tech development without remunerating them would be

² It is worth noting that AI-powered biometric surveillance is increasingly being used, so dark skin failing to be detected or accurately recognized by digital photography and facial recognition software is a significant issue, which women of color have been instrumental in highlighting. See Browne (2015), Buolamwini and Gebru (2018), Benjamin (2018), Rhue (2018), Chun (2021).

unethical, though this did not stop Google from tricking people into allowing their pictures to be used to train facial recognition systems in exchange for \$5 gift cards (Nicas, 2019). In these cases egalitarian and representational diversity are impractical goals. While there are cases where diversity as token representation can be helpful (people with uteruses should be involved in developing period trackers, and people with disabilities should be involved in developing accessibility tools) this type of diversity is limited in its reach. An additional problem on the side of a person hired for the sake of token representation, is that they would be expected to take on an impossible task (since no individual has the life experience to represent all diversity), often in addition to their technical work. Being invited to be the token woman, POC, or disabled person on a project might serve to deflect criticism of the institution, but if that token person does not have any allies to work with, and is not given any power to change the institution, their presence has little effect. It can even make the people put in this position leave the field due to “diversity fatigue” (Lam, 2018). If that person is not aware of the concerns of marginalized groups, their presence may be counter-productive.

Online discussion about diversity in AI, as represented in the responses to the “racist soap dispenser”, assume either a normative defense of diversity that fails to connect with cultural values in the field, or a defense in terms of representation that is too weak to do the work demanded of it. **2.2 Diversity as Code for Women**

Diversity is often used as a code word for hiring more women (for example, Al-Heeti, 2021; St. Louis, 2017). Diversity initiatives in computer science have historically focused on women. Programs aimed at helping women often end up in fact being for the benefit of cisgender, middle-class, white women, rather than supporting all women, or prioritizing people who are multiply marginalized (Crenshaw, 1991). Such programs can

harm other equity-seeking groups by taking up limited resources earmarked for equity, diversity, and inclusion, reducing trust among sub-groups of women, or by promoting policies that target some women with discrimination while claiming to help others, like bathroom bills that exclude trans women.

Diversity as code for women has been the target of significant backlash within AI. It is the main target of a notorious leaked memo written by an engineer no longer at Google that assumes a biological basis for essentialized gender roles and argues that increasing diversity would mean hiring unqualified people, making diversity initiatives bad for the bottom line. The memo seems to assume a notion of diversity where the goal is to hire anyone except white men, regardless of qualifications for technical roles. This could be read as a confusion over what the reference group for representational diversity is—the general population, or qualified candidates for jobs at Google. This reaction is also consistent with the assumption of normic diversity, taken as an end goal to be achieved, rather than a corrective measure to be worked toward.

Despite this backlash, hiring more people with stereotypically feminine soft skills could in fact be beneficial to AI, however, equating soft skills with women's work has several drawbacks. It casts women in tech in the role of soft skill specialists, regardless of their actual skills and strengths, potentially making it more difficult for women with technical skills to be promoted in technical roles. Gendering work as feminine tends to de-value that work (Hicks, 2017), leading to lower status in the workplace and lower wages, and it could exacerbate existing problems where women in tech are expected to take on unpaid clerical and care work.

The infamous memo also argues that white, heterosexual, conservative men are marginalized and discriminated against at Google, making them a group whose viewpoints

need to be foregrounded. A similar sentiment was expressed by a computer science department faculty member where one of the authors was a graduate student, who made the national news for replying-all to a campus-wide announcement about a Montreal Massacre memorial, complaining that as an anti-feminist he felt personally targeted by the event (CBC, 2000). These men may be naïve to think that marginalization is the same thing as the discomfort they experience when asked not to say things that are offensive to their colleagues, however, the implied claim that diversity programs should increase the numbers of people who are truly under-represented has some merit. Focusing on women under the banner of diversity employs such a weak proxy for under-represented workers that it has the potential to do as much damage as it does good.

2.3 The Business Case for Diversity

The opening to Ahmed's (2012) study of diversity workers in higher education, asks, "What does diversity do? What are we doing when we use the language of diversity?" She goes on to detail how the term "diversity" has displaced others like "equal opportunity" and "antiracism" as equity initiatives have been taken over by marketing departments more interested in the "business case" for diversity than normative reasons for seeking equity (52–53).

The business case for diversity depends on correlations between measures like profits and the proportion of women or POC in executive positions. For example, as the conservative management consultancy McKinsey & Company says in a report about the value of diversity, "The most diverse companies are now more likely than ever to outperform less diverse peers on profitability" (McKinsey, 2020). This report offers details about how profitability tracks diversity, and how to deal with negative feelings about diversity in the workplace, but does not mention reasons why equitable hiring could translate

into better functioning workplaces. That there happens to be this profitable correlation is all that matters.

The pattern Ahmed points to, where the equity goals of diversity initiatives get whittled away and eventually replaced with diversity theatre has been playing out in AI, where calls to fix an institutional culture that pushes out minorities and produces discriminatory products has led to an explosion of official statements, codes of ethics, and other pro-diversity speech acts, but very little meaningful change. Empirical evidence suggests that ethical codes are not very effective at changing behaviour (McNamara, Smith, and Murphy-Hill, 2018).

Equity-seeking groups in AI are also among the critics of diversity, on the grounds that investments in diversity go to these empty PR rituals, while the people hired to increase diversity are prevented from doing their work (Fussell, 2021), fired if they prove too effective (Curley, 2020), and denied resources, autonomy, and a voice (Whittaker et al., 2018). These criticisms are targeted at the superficial and ineffective ways diversity programs have been designed and rolled out, and at the lack of efforts to change the culture of AI in which diversity recruiting efforts operate and place workers. Treating diversity as a means to increased profits can too easily be coopted into an empty gesture, wherein neither the normative nor the epistemic aims are achieved.

2.4 Diversity as Critical Perspective

As Noble (2018: 163) writes, if tech companies really want to deal with their discrimination problems, they could hire graduates in “Black studies, ethnic studies, American Indian studies, gender and women’s studies, and Asian American studies”. Instead, they are hiring machine learning experts to build fairness algorithms. Note that Noble’s suggestion is not to hire representatives of Black, ethnic, etc., communities, but

rather people who have studied oppression and difference, are aware of power dynamics, and have social science and humanities skills. Timnit Gebru has also been an outspoken supporter of this perspective, however, there has also been considerable pushback against these ideas from influential people in AI (Cai 2020).

Fehr (2011), responding to the challenge of communicating to people in STEM fields the reasons they should take diversity more seriously, argues that “diversity development work” like hiring and supporting “those with underrepresented theoretical perspectives, social locations” brings “epistemic benefits to academic communities” (145), not in virtue of including more minorities per se, but in virtue of the critical value of their perspectives. She distinguishes between the ethical problem of “situational diversity” and the cognitive problem of “epistemic diversity” (146), but citing Lorraine Code, explains that “our cognitive problems and ethical problems are often intertwined” (147). Failing to hire and support a more diverse workforce is both an ethical and an epistemic failure. Including people with diverse standpoints can simultaneously bring scientific excellence and relief for AI’s recent string of ethical blunders. That the normative and epistemic benefits of diversity often coincide in this way suggests a route through which one might defend diversity on STEM-friendly epistemic grounds and gain the normative benefits as a welcome side-effect.

3. The Epistemic Grounds of Diversity

As we learned from the “racist soap dispenser” and the leaked Google memo, the case for diversity is a hard sell in tech and AI. Calls for more representation of some demographics in the workforce are received as threats to the status quo and the bottom line. While we take seriously normative arguments for diversity in AI, an alternative framing is needed to reach the intended audience of AI and tech workers and decision-makers, and provide diversity advocates with another set of arguments. In this section we explore several resources for

supporting diversity on epistemic grounds. We first turn to feminist philosophy of science, giving a brief overview of both feminist empiricism and feminist standpoint theory. We then consider how intersectional feminist approaches complicate these frameworks before turning to empirical research suggesting that diverse groups tend to outperform groups of experts. Taken together, these insights suggest an epistemic grounding of diversity projects in AI.

3.1 Insights from Feminist Philosophy of Science

Early work in feminist philosophy of science was largely concerned with the underrepresentation of women in science. Whether increased diversity can mitigate issues of bias and masculinist perspectives in science, and how we might conceptually ground such initiatives, remains a central topic of inquiry. Feminist philosophers of science have argued that increasing diversity in scientific roles is justified on both ethical and epistemic grounds, as it serves to produce both more equitable hiring practices and better scientific outcomes. Feminist philosophers of science have expanded their concept of diversity to include members of other marginalized groups as well. Early examples of attempts to understand the role of diversity in science include feminist empiricism (Anderson, 1995; Longino, 1987; Nelson, 1990), situated knowledges (Haraway 1991), and feminist standpoint theory (Harding, 1986; Hill Collins, 1991; Smith, 1974).

Early iterations of feminist empiricism introduced the notion that diversity in scientific roles leads to better scientific outcomes. Helen Longino (1987) explores what a feminist practice of science might entail. Rather than conflating the feminine with the feminist, Longino turns her attention to science as a process. Longino argues that science is comprised of both constitutive and contextual values that cannot be disentangled. While

constitutive values are internal to the sciences—rules which distinguish “good” science from “bad”—contextual values concern the social and cultural context.

Longino points out that constitutive values are not as separable from contextual values as sometimes believed. Consider for example what counts as scientific evidence. Contextual values might seem absent here, but background metaphysical assumptions often inform what types of evidence are taken seriously. For example, as Keller (1983) describes in *A Feeling for the Organism*, it took several decades for mainstream genetics to accept Barbara McClintock’s evidence that environment affects gene expression, because this meant upending the “central dogma” wherein genetic information is statically and linearly encoded in DNA. Scientific paradigms are structured by their social and political conditions.

Although Longino argues that science is value-laden, she assures us this need not be a problem. If it is possible to do “good” science that is value laden, we just need to figure out which values should be used in which contexts. While feminist empiricism has been revised many times since it first appeared, Intemann (2010) notes that three key features have been consistent. Science is context-sensitive, normative, and social; context and values matter, and objectivity emerges from scientific communities.

In contrast, feminist standpoint theorists argue that systemically marginalized groups may possess greater epistemic insight in virtue of having different experiences than dominant groups. Such knowledge is “socially situated” (Haraway, 1991) and may be different in both kind and degree. Situated knowledge is also importantly partial and differs between various individuals within social groups. In this account, knowledge is socially situated and marginalized groups are often epistemically privileged.

Wylie (2003) identifies another commonality between strands of feminist standpoint theory: a rejection of standpoint essentialism, and what she calls “automatic epistemic

privilege” (28). On Wylie’s account, one does not gain epistemic insight automatically. Rather, standpoints are developed through experience. It is not simply in virtue of being a woman, a POC, etc., that one gains epistemic privilege, but rather reflection on that experience of difference that may bring insights. It is a form of expertise.

Feminist standpoint theory differs most obviously on the question of epistemic privilege. Harding (1995) has advocated for “strong objectivity” meaning that marginalized standpoints ought to be taken as the starting point for undergoing scientific inquiry. While other standpoints are still necessary for painting a full picture, starting with marginalized standpoints ensures that scientific inquiry will not be ignorant of the effects of systemic oppression on science. A certain kind of objectivity may be obtained through this method. Harding’s suggestion that we ought to start our inquiry with the most marginalized can be read as a defense of normic diversity, used as an interim strategy, given that dominant social positions would be excluded from such projects until all marginalized voices are represented.

Both views agree that diversity is essential to mitigating bias in science, but differ on the kind of diversity that contributes most to success in science. Feminist empiricists hold that diversity *of values* support better scientific results, whereas standpoint theorists hold that diversity *of social positions* is what matters most. Intemann (2010) argues for social positions as the most plausible form of diversity to consider; diversity of social positions produces a broader range of empirical data because of scientists’ wider range of experiences. Individuals across social locations can have the same values, but their experiences are different, by definition.

Intemann suggests a fusion of the two views that she terms “feminist standpoint empiricism.” The proposed view takes feminist standpoint theory’s perspective on diversity

and feminist empiricism's explicit commitments to empiricism. On this view, many people with diverse social locations are required in order to produce stronger scientific results.

3.2. Insights from Intersectional Feminism

The consideration of intersectional feminist approaches complicates this picture. Intersectionality is a framework that examines how structures of domination intersect to produce unique forms of oppression. Crenshaw uses the example of racism and sexism (Crenshaw 1991). The racism experienced by Black men differs in character to the racism experienced by Black women, as racism is entangled with sexism. Likewise, the sexism experienced by white women differs in character to the sexism experienced by Black women, as sexism is entangled with racism. It is not the case that Black women experience the same racism as Black men *plus* the same sexism as white women, as structures of racism and sexism intersect to produce unique forms of oppression.

This has consequences for determining which form of diversity (normic, egalitarian, or representational) would be required to maximize the diversity of standpoints. Contrary to Harding, it would not be enough to recruit "the most marginalized", as each intersection of oppression is unique. There might be insights into how racism functions that can only be understood from the standpoint of Black men, rather than any Black person.

Furthermore, it is not always evident which aspects of one's social location will be relevant. For example, street violence against trans women is often taken to be the result of the intersection of sexism and transphobia. However, some have argued that the relevant factor in the frequent murders of trans women in the United States is oppression against sex workers (Namaste, 2005). If one wanted insight into how this particular form of violence against trans women manifests, it would not be enough to possess the standpoint of a trans woman. Rather, one would need to possess the standpoint of a (possibly trans) sex worker.

To cultivate a diversity of social locations, each intersection acts as its own specific standpoint and the ideal would be egalitarian diversity, or inclusion of members of all standpoints. This might be an impossible end goal to achieve, but can be pursued as an ideal

3.3 Diversity Trumps Individual Expertise

In social epistemology the value of diversity has been subject to empirical examination. Hong and Page (2004) used agent-based simulations to compare the contributions of diversity and ability on problem solving teams. Their surprising finding was that diverse teams (i.e., made up of randomly chosen individuals) performed better on average on problem solving tasks than teams made up of the individuals who had highest ability as independent problem solvers.

This result has been both contested and supported by a large body of follow-up research and cited in a wide range of policy documents. Grim et al. (2019) examine how the original results fare when genuine expertise is distinguished from high ability. If experts are defined as ones who have high ability that generalizes across many related problems, they find that expertise sometimes trumps diversity. Which is most important depends on the problem topography. As in the discussion above over whether values or standpoints are most important, this literature also grapples with whether epistemic virtues attach to “functionally diverse groups” or “identity diverse groups” (Singer, 2019: 179).

Singer (2019) explores the difference between random groups and groups chosen such that their problem-solving heuristics maximize coverage of the space of possibilities. Diverse groups on this definition outperform randomly selected groups. They conclude that the epistemic value of random groups is due to their relative diversity, and that it is worth

spelling out more carefully what exactly one means by diversity. Fazelpour and De-Arteaga (2022) echo the second conclusion.

Cowgill et al. (2020) empirically test how groups of programmers with different demographics and resources for avoiding bias compare in terms of their ability to develop an unbiased algorithm. They find no evidence that female or racial minority engineers exhibit less bias than other engineers, supporting the claim that standpoints are not achieved automatically. However, Cowgill et al. (2020) do find that different demographic groups exhibit *different* biases. What turns out to be important for avoiding biased algorithms is not the inclusion of members of marginalized groups per se, but that teams not be homogeneous, so that the biases within a group do not all coincide. Even if minorities might be more likely to have insight into marginalization, this does not by itself determine which perspectives have a high probability of adding new knowledge to a given project, as this depends on the task and the composition of the team. What seems to matter is the addition of diversity relative to the existing team. This could be seen as support for egalitarian diversity, in that the inclusion of people from all standpoints is equally valuable, or normic diversity, if the norm against which members are chosen is defined as whomever is already on the team.

These results provide validation of previous theoretical work by feminist philosophers of science and social epistemologists, but are limited to particular experimental setups. One thing not captured here is the role of intersectionality. The social epistemology models treat all agents as individuals with unique collections of knowledge. This in effect recognizes intersectional positions while erasing connections across single axes, like between Black men and women. Cowgill et al. analyze gender and race separately, though they mention that the racial makeup differs for men and women in their study. Another thing not

captured here is how the dynamics between subgroups within teams affect outcomes. The difference between egalitarian and representational diversity is the relative numbers of people from different standpoints. A critical mass may be needed for marginalized voices to be heard.

4. The Epistemology of Expertise

To distinguish between the expertise held by individuals and the expertise held by groups, we borrow the term *networks* from Actor Network Theory (Latour, 1987). Networks are made up of individuals working collectively on a given task or problem. In this section we describe technical expertise, standpoint expertise, and introduce *emergent expertise*. People working in AI tend to value technical expertise, while diversity initiatives often appear to focus on standpoint expertise. Here we elaborate a more complex picture of the dynamics that obtain in networks, wherein the technical and standpoint expertise(s) of the various individuals combine to create knowledge that may go beyond what any individuals hold in isolation. We then turn to two examples of glitchy technology to show how more diverse networks of expertise could produce machines that avoid both ethical and functional pitfalls.

4.1. Kinds and Degrees of Expertise

Technical expertise often aligns with our everyday understanding of expertise. It broadly includes specialized knowledge acquired through study or practice. Examples include the surgical expertise of a medical doctor, the literary expertise of an English professor, or the research skills of a social scientist. In the context of AI, engineers and programmers have technical expertise to design and construct models, datasets, and hardware.

On the other hand, *standpoint expertise* is knowledge typically gained through lived experience of one's social location in relation to dominant systems. In the same way that an individual can know-that with respect to a technical skill like driving a car (one ought to keep one's hands at ten and two on the steering wheel and maintain a safe braking distance of at least two seconds), knowing-how to drive is quite different. Analogously, one can have knowledge-that understanding of oppression from taking a gender studies seminar but gain know-how with respect to its operations if one is subject to oppression. Standpoint expertise can help illuminate dynamic situations in ways mere technical expertise cannot.

In *Rethinking Expertise*, Collins and Evans (2007) discuss different levels and gradations of expertise. On their account, specialist expertise ranges from ubiquitous tacit knowledge (popular understanding) to specialist tacit knowledge. The highest level of specialist knowledge is "contributory expertise," (CE) which is necessary to perform a task with competence. Below this is "interactional expertise," (IE) which consists of mastery of the language deployed in a specialist domain. For example, one could read and understand mathematical proofs without having the know-how to write the proofs oneself.

An example of expertise interacting in a network is how AIDS activists in the late 1980s changed the direction of research. The network included immunologists, virologists, physicians, pharmaceutical companies, as well as treatment activists, activist publications, and the gay press (Epstein, 1995: 409). AIDS activists were able to learn medical vocabulary and insert themselves into pre-existing debates within the field. In this case the intersection of gay and male was a relevant standpoint, as gay men were disproportionately affected by AIDS. Having both interactional technical expertise in the form of their newfound medical knowledge and contributory standpoint expertise with respect to how AIDS was spreading in their communities, AIDS activists were able to communicate their interests and work with

medical professionals who possessed contributory medical expertise in order to produce advances in AIDS research that required contributory expertise of both standpoint and medical kinds, which neither group possessed in isolation.

The emergent character of the network's expertise is visible in the development of a "pragmatic" approach to clinical trials, upon the activists' insistence. It did not make sense to AIDS activists for drug trials to operate under a "fastidious" model designed to uphold rigid standards and neatly controlled variables. Since people were dying from AIDS every day, widely available drugs and treatments were the focus rather than answers to abstract theoretical questions. Through this collaboration between individuals with contributory medical expertise and contributory standpoint expertise, a new paradigm for clinical trials was developed.

What previous accounts of expertise often miss is how network expertise is not purely additive, similar to how structures of domination intersect to create unique forms of oppression. The activists' standpoint CE and medical IE combined with the physicians' medical CE to create *emergent expertise* (EE) surpassing both bodies of CE. This EE is possessed by the network as a whole. This new form of expertise expands on previous accounts by considering how different kinds of expertise at multiple levels interact to produce novel insights.

4.2. Emergent Expertise

EE is not a guaranteed outcome when groups with different kinds of expertise meet. A minimal condition for EE is that some individuals in a network will possess relevant technical CE, while others will possess relevant standpoint CE. It is also necessary for some individuals to possess or gain *both* kinds of expertise to some degree (like a POC who is an

entry-level engineer, or a white senior engineer who has read Critical Race Theory) so that conversations across kinds of expertise are possible. Simply putting two sets of experts in a room together is not enough. While the ideal individual is one who possesses both standpoint and technical CE, such "unicorns" may be rare, and still benefit from working within a network.

For example, we might imagine a small network with four individuals:

Individual 1 has standpoint CE

Individual 2 has technical CE

Individual 3 has standpoint CE and technical IE

Individual 4 has tacit standpoint knowledge and technical CE

While Individuals 1 and 2 possess CE of one kind, they lack knowledge of the other. Individuals 3 and 4, who have some of each kind of expertise, play an essential role in initiating conversation across kinds of expertise. Their expertise is in a sense 'intersectional' and as such not equivalent to the sum of their two kinds of expertise taken in isolation. If all the individuals work together, the network might generate EE. In the AIDS example, the doctors did not possess relevant standpoint IE, but through their clinical work and interactions with the activists gained tacit standpoint knowledge. That the AIDS activists acquired technical IE through study and participation in specialist conferences, rather than mere tacit knowledge, seemed to be essential to the creation of EE in this case. In AI, where non-technical knowledge is devalued, standpoint expertise may not be taken seriously unless it is accompanied by sufficient technical expertise.

In a scientific network where laypeople's views are given more voice, it is also possible for EE to emerge in a network in which individuals have only tacit knowledge of their non-specialist area. An example like this, where tacit knowledge is sufficient, is

participatory research in psychiatry. Participatory research is a movement to involve patients and service users in medical research as co-producers or citizen scientists. One example where mental health service users have had an impact on how their symptoms are treated and understood is the Hearing Voices Network (HVN) (<https://www.hearing-voices.org/>), which encourages positive responses to the experience of hallucination.

In this example, psychiatric service users and survivors have been described as being experts-by-experience, who contribute their expertise in numerous ways, including re-focusing clinicians attention on the symptoms and side-effects that they find most interfere with their quality of life, which are not always the symptoms (like hallucinations) that figure in psychiatric diagnoses or are the main targets for pharmaceutical interventions. Noorani (2013) develops a concept of “experiential authority” to describe the expertise service users have in virtue of living through mental distress and working through self-help or support groups. This is contrasted with the “traditional authority” held by medical professionals. Relabeling standpoint knowledge in this was as “expertise” might also be effective in AI.

Some of the documented outcomes of participatory research are noticing novel themes in interview data, better retention in studies, more forthcoming interviewees, greater accessibility of clinical trials, and identification of novel research questions (Friesen et al., 2019). The interplay between service users’ experiential authority (standpoint CE) and scientists’ traditional authority (technical CE) has brought about knowledge that neither group would have been capable of producing in isolation, in the form of EE. For this kind of relationship of mutual respect to obtain, individuals with technical CE need to accept that their kind of expertise is not the only relevant kind. The conditions that aid in the production of EE can be summarized as:

1. Some individuals must possess relevant technical CE.

2. Some individuals must possess relevant standpoint CE.
3. Some individuals must possess both kinds of expertise.
4. The individuals must engage in conversation across kinds of expertise.

Even if these conditions are met, EE is created through a process of working together, not automatically generated.

As there might be multiple kinds of relevant standpoint CE, increasing the number of social locations represented when constructing a network (while still including individuals with technical expertise) is good policy. Multiple types of technical expertise may also be beneficial, including social science expertise, as some scholars have argued (Carman and Rosman, 2020; Crawford and Calo, 2016; Miller 2019). These goals are most consistent with egalitarian diversity, at least as an interim strategy, though it need not be pursued to the point of actually achieving parity among all groups. Normic diversity could have the same effect depending on how the norm is defined. Normic diversity may be more suitable when a technology is positioned in such a way as to consolidate power, since in that situation a critical mass of non-normative individuals may be needed for their voices to be heard. A representational diversity strategy could have the effect of continuing to add people from standpoints already well represented, if larger groups have not yet reached their proportion in the reference population, so would be less effective.

Normative arguments for diversity might suggest representational diversity as the type that best upholds liberal values like fairness. However, the epistemic argument developed here instead supports normic or egalitarian diversity, at least as initial or course correcting strategies, which coincides with more radical political values.

4.3 Emergent Expertise in AI

We now turn our attention to applications of EE in the tech sector, looking first at the suggestion that adding POC to the design team would have prevented the creation of a “racist soap dispenser”. Consider two hypothetical individuals: an engineer with experience building light-sensitive devices but little knowledge of the experiences of POC, and a POC with some knowledge of industrial design, and situated expertise as a user of services that assume a white audience. The engineer possesses the technical CE to design the sensor mechanism, and the POC has both the standpoint CE to observe a problem with the soap dispenser’s operation and tacit knowledge in industrial design. Neither individual possesses the expertise required to solve the problem working in isolation, though in this simple case, one could find a single engineer with both kinds of CE.

When adjudicating between prospective experts, we ought to consider whether new types of expertise are being added to the network. It is unclear what type of expertise a second engineer with a similar standpoint to the first would add to the network. However, someone with standpoint CE in the needs of neurodivergent people or children might also be a useful addition to the team. A single engineer can only go so far on good intentions when attempting to design a universally accessible product. As Green (2019: 2) argues, even in computer science research projects that aim to advocate for the greater social good, “the assumptions and values of dominant groups will tend to win out.”

Consider a more complex example: the millimeter wave scanners that are now a standard part of US airport security screening. Costanza-Chock (2020) details their stressful history of interactions with these machines as a trans person. Most travelers remain oblivious to the fact that before you step into the machine, security staff pushes a pink or blue button based on their judgment of your sex/gender. The image of your body the machine takes is then judged as either passing or failing the automated screening based on how your

configuration of flesh compares to either the female or male binary sexed/gendered model of what are deemed non-suspicious bodies. When the security staff chooses one of the gendered buttons for a trans or gender non-conforming person, fluorescent yellow alerts may show up in the crotch or chest, where unexpected flesh is detected. In the case of many trans people, both buttons may lead to security alerts, as neither presumed body model corresponds to their anatomy. An additional pat-down is all but guaranteed for people with bodies that fit neither model.

. This mirrors the soap dispenser case, but defeats arguments that a simple technological fix could make it work. Any combination of lumps of flesh in crotch and chest are possible, they do not fit into two distinct models, and they are not reliably predictable from gender expression.

What is striking about this example from the perspective of expertise is that the problem was uncovered through informal networks of trans people sharing their negative experiences with airport security and discovering that there was a pattern. This network of disgruntled travelers also needed to connect their standpoint expertise to people who work in airport security and combine it with those workers' procedural knowledge of the pink and blue buttons in order to fully understand the problem. A well-meaning AI developer trying to fulfill the TSA's request for an anomalous flesh detection system without discriminating against trans travelers would not only need to have technical expertise about designing AI systems, they would also need expertise about airport security procedures from the perspective of staff, and detailed knowledge about the many different configurations of flesh one might find under the clothing of trans people, for which one either needs standpoint CE, or to have done enough research on a the subject to have obtained IE. As Browne (2015) details, similar concerns affect Black women who also often face additional airport

security screening because their hair is identified as anomalous. Other groups with relevant standpoint expertise might include people with limb differences, and people who use wearable medical devices. These examples serve to illustrate that networks of individuals with diverse kinds of expertise are necessary to produce technologies that solve complex problems and serve a wide variety of differently socially situated individuals.

What diversity initiatives are aiming for is not, after all, to hire people with no technical expertise but only standpoint expertise. The point of diversity initiatives in AI is to hire people who *do* have technical expertise, but who *also* inhabit different social positions than the majority of existing workers. Diverse candidates qualified for entry-level tech jobs exist in large numbers, and while candidates at higher levels of technical expertise may be rarer, more can be created. So-called diversity hires can have not only the technical expertise relevant to the technical aspects of their job, but also standpoint expertise that differs from other workers already in the field. Network effects and ‘intersectional’ expertise can then work to combine these diverse sources of expertise into EE that can outstrip the sum of the expertise of the individuals in the network. This model of diversity promises to deliver greater knowledge than teams with homogeneous expertise, regardless of the identities of the individuals involved.

This is not to say that identities never matter. In some cases existing workplace power imbalances or the political nature of the technology being designed may call for a normic approach to diversity as a goal unto itself. These conditions may also call for more traditional diversity initiatives like mentoring programs aimed at workers from equity-seeking groups. Part of the value of those programs comes from the fact that mentoring and training people with standpoint CE who are hired at the entry-level can help

them rise to the level of technical CE. Over time this creates more of the ideal individuals who possess both forms of CE.

5. Conclusion

By grounding our diversity projects in an understanding of expertise as an emergent phenomenon, we are able to circumnavigate many of the problems flagged by both critical advocates of diversity, and skeptics of diversity. Diversity advocates can use the framing presented here as a way to ground more effective diversity-based initiatives. Diversity skeptics might find some of the arguments presented here compelling enough that the backlash against diversity does not lead AI companies to wash their hands of it entirely. Furthermore, by understanding expertise as being distributed across networks, we see that high quality science is not produced by lone geniuses or even homogenous groups, but rather through the interaction of different kinds of expertise. Rather than merely advocate for more inclusive technologies on a case-by-case basis, networks of diverse individuals with wide-ranging expertise have the potential to build better, more equitable technology, and spur the growth of more diverse expertise within their field.

References

- Afigbo C (2017) [Twitter]. 22 August. Available at: https://twitter.com/nke_ise/status/897756900753891328 (accessed 3 June 2023).
- Ahmed S (2012) *On Being Included: Racism and Diversity in Institutional Life*. Durham: Duke University Press.
- Ahonen T, Hadid A, and Pietikainen M (2006) Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 12: 2037–2041.
- Al-Heeti A (2021) Melinda Gates, diversity advocates address the long road to equality at SXSW. In: Cnet. Available at: (accessed June 3 2023).

- Alcorn C (2021) Google CEO and HBCU leaders discuss talent pipeline for Black tech workers. *CNN Business*, 1 February. Available at: <https://www.cnn.com/2021/01/31/business/google-hbcus/index.html> (accessed 3 June 2023).
- Anderson E (1995) Feminist epistemology: An interpretation and defense. *Hypatia* 10(3): 50–84.
- Avakian AV (2005) *From Betty Crocker to feminist food studies: Critical perspectives on women and food*. Liverpool: Liverpool University Press.
- Beard C (2018) Diversity and Inclusion at Mozilla. In: Mozilla. Available at: (accessed 3 June 2023).
- Benjamin R (2018) *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity Press.
- Birhane A and Guest O (2020) Towards decolonising computational sciences. *arXiv preprint arXiv:2009.14258*.
- Blodgett SL and O'Connor B (2017) Racial disparity in natural language processing: A case study of social media African-American English. *arXiv preprint arXiv:1707.00061*.
- Browne S (2015) *Dark Matters: on the Surveillance of Blackness*. Durham: Duke University Press.
- Buolamwini J and Gebru, T (2018) Gender Shades: Intersectional accuracy disparities in commercial gender classification. In: *Conference on fairness, accountability and transparency*.
- CBC News (2000) Professor criticizes Montreal massacre memorials. Available at: <https://www.cbc.ca/news/canada/professor-criticizes-montreal-massacre-memorials-1.235652> (accessed 3 June 2023).
- Cai, Fangyu. (2020) Yann LeCun Quits Twitter Amid Acrimonious Exchanges on AI Bias. *Synced*. <https://syncedreview.com/2020/06/30/yann-lecun-quits-twitter-amid-acrimonious-exchanges-on-ai-bias/> (accessed 23 October 2023).
- [Carman M and Rosman B \(2021\) Applying a principle of explicability to AI research in Africa: should we do it? *Ethics and information technology* 23\(2\): 107–117.](#)
- Cheryan S, Ziegler SA, et al. (2016). Why Are Some STEM Fields More Gender Balanced Than Others? *Psychological Bulletin* 143(1): 1–35.
- Chun WHK (2021) *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Boston: MIT Press.
- Collins H and Evans R (2007) *Rethinking Expertise*. Chicago: University of Chicago Press.
- ComputerScience.org Staff. (2022) Women in Computer Science: Getting Involved in STEM. Available at: <https://www.computerscience.org/resources/women-in-computer-science/> (accessed 3 June 2023).
- Costanza-Chock S (2020) *Design Justice: Community-led practices to build the world we need*. Boston: MIT Press.
- Cowgill B, et al. (2020) Biased programmers? or biased data? A field experiment in operationalizing AI ethics. In: *Proceedings of the 21st ACM Conference on Economics and Computation*.
- Crawford K (2016) Artificial Intelligence's White Guy Problem. *The New York Times*. 25 June. Available at: <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html> (accessed 3 June 2023).
- Crawford K and Calo R (2016) There is a blind spot in AI research. *Nature (London)* 538 (7625): 311–313.

- Crenshaw K (1991) Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review* 43: 1241–1299.
- Curley A (2020) [Twitter]. 21 December. Available at: <https://twitter.com/RealAbril/status/1341135819487100928> (accessed 3 June 2023).
- Epstein S (1995) The Construction of Lay Expertise: AIDS Activism and the Forging of Credibility in the Reform of Clinical Trials. *Science, Technology, & Human Values* 20(4): 408–437.
- Fazelpour S, and De-Arteaga M (2022) Diversity in Sociotechnical Machine Learning Systems. *Big Data & Society* 9: 1–14.
- Fehr C (2011) What is in it for me? The benefits of diversity in scientific communities. In: Grasswick H (ed) *Feminist epistemology and philosophy of science*. NYC: Springer, pp. 133–155.
- Forsythe DE (1993) Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence. *Social Studies of Science* 23(3): 445–477.
- Friesen P, Lignou S, et al. (2019) Measuring the impact of participatory research in psychiatry: how the search for epistemic justifications obscures ethical considerations. *Health Expectations* 24(S1): 54–61.
- Fussell S (2021) Black Tech Employees Rebel Against ‘Diversity Theater’. *Wired*, 8 March. Available at: <https://www.wired.com/story/black-tech-employees-rebel-against-diversity-theater/> (accessed 3 June 2023).
- Gershgorin D (2019) Microsoft staff are openly questioning the value of diversity. *Quartz*, 19 April. Available at: <https://qz.com/1598345/microsoft-staff-are-openly-questioning-the-value-of-diversity/> (accessed 3 June 2023).
- Green B (2019) ‘Good’ isn’t good enough. In: *Proceedings of the AI for Social Good workshop at NeurIPS*.
- Grim P, Singer DJ, et al. (2019) Diversity, ability, and expertise in epistemic communities. *Philosophy of Science* 86(1): 98–123.
- Hale J (2017). This Viral Video of a Racist Soap Dispenser Reveals a Much Bigger Problem in Tech. *IFL Science*, 18 August. Available at: <https://www.iflscience.com/technology/this-racist-soap-dispenser-reveals-why-diversity-in-tech-is-muchneeded/> (accessed 3 June 2023).
- Haraway D (1991) *Simians, Cyborgs, and Women*. NYC: Routledge.
- Harding S (1986) *The science question in feminism*. Ithaca: Cornell University Press.
- Harding S (1995) Strong Objectivity: A Response to the New Objectivity Question. *Synthese* 104(3): 331–349.
- Hicks M (2013) De-Programming the History of Computing. *IEEE annals of the history of computing* 35(1): 88.
- Hicks M (2017) *Programmed inequality: How Britain discarded women technologists and lost its edge in computing*. Boston: MIT Press.
- Hill Collins P (1991) Learning from the outsider within. In Fonow MM and Cook JA (eds) *Beyond methodology: Feminist scholarship as lived research*. Bloomington: Indiana University Press, pp. 35–59.
- Hong L and Page SE (2004) Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences* 101(46): 16385–16389.

- Intemann K (2010) 25 Years of Feminist Empiricism and Standpoint Theory: Where Are We Now? *Hypatia* 25(4): 778–796.
- Keller EF (1983) *A feeling for the organism: The life and work of Barbara McClintock*. NYC: Freeman.
- Lam MB (2018) Diversity Fatigue is Real. *The Chronicle of Higher Education*, 23 September. Available at: <https://www.chronicle.com/article/diversity-fatigue-is-real/> (accessed 3 June 2023).
- Latour B (1987) *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, Massachusetts: Harvard University Press.
- Liao S and Huebner B (2020) Oppressive things. *Philosophy and Phenomenological Research* 103(1): 92–113.
- Longino H (1987) Can There be a Feminist Science? *Hypatia* 2(3): 51–64.
- McKinsey and Co (2020) Diversity Wins: How Inclusion Matters. Online report. Available at: <https://www.mckinsey.com/featured-insights/diversity-and-inclusion/diversity-wins-how-inclusion-matters> (accessed 3 June 2023).
- McNamara A, Smith J, and Murphy-Hill E (2018) Does ACM's Code of Ethics Change Ethical Decision Making in Software Development? In: *Proceedings of the 26th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*.
- Miller T (2019) Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* 267: 1–38.
- Namaste V (2011) *Sex Change, Social Change*. Toronto: Canadian Scholars' Press. Nelson LH (1990) *Who knows? From Quine to a feminist empiricism*. Philadelphia: Temple University Press.
- Nicas, J. (2019) Atlanta Asks Google Whether it Targeted Black Homeless People. *The New York Times*, 4 October. Available at: <https://www.nytimes.com/2019/10/04/technology/google-facial-recognition-atlanta-homeless.html> (accessed 3 June 2023).
- Noble S (2018) *Algorithms of Oppression*. Boston: MIT Press.
- Noorani T (2013) Service user involvement, authority and the 'expert-by-experience' in mental health. *Journal of Political Power* 6(1): 49–68.
- Paul K (2019) 'Disastrous' lack of diversity in AI industry perpetuates bias, study finds. *The Guardian*, 17 April. Available at: <https://www.theguardian.com/technology/2019/apr/16/artificial-intelligence-lack-diversity-new-york-university-study> (accessed 3 June 2023).
- Rhue L (2018) Racial influence on automated perceptions of emotions. *Available at SSRN* 3281765.
- Smith D (1974) Women's perspective as a radical critique of sociology. *Sociological Inquiry* 44(1): 50–84.
- Singer DJ (2019) Diversity, not randomness, trumps ability. *Philosophy of Science* 86(1): 178–191.
- St. Louis M (2017). Game-Changing Diversity Lessons from Women in Tech. *Inc*. Available at: <https://www.inc.com/molly-reynolds/game-changing-diversity-lessons-from-women-in-tech.html> (accessed 3 June 2023).
- Steel D, et al. (2018) Multiple Diversity Concepts and Their Ethical-Epistemic Implications. *European journal for philosophy of science* 8(3): 761–780.

- Stinson C (2022) Algorithms are not neutral: Bias in collaborative filtering. *AI and Ethics* 2: 763–770.
- Turk MA and Pentland AP (1991) Face recognition using eigenfaces. In: *1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- West SM, Whittaker M, and Crawford K (2019) *Discriminating Systems: Gender, Race and Power in AI*. AI Now Institute at New York University.
- Wheatley B and Picone J (1991) Voice Across America: Toward robust speaker- independent speech recognition for telecommunications applications. *Digital Signal Processing* 1(2): 45–63.
- Whittaker M, et al. (2018) *AI now report 2018*. AI Now Institute at New York University.
- Winner L (1980) Do artifacts have politics? *Daedalus* 109(1): 121–36.
- Wright J, et al. (2008) Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 31(2): 210–227.
- Wylie, A. (2003). Why Standpoint Matters. In: Figueroa R and Harding S (eds) *Science and Other Cultures: Issues in Philosophies of Science and Technology* NYC: Routledge, pp. 26–48.