

# Technology can't fix this

AI tools used in society often enhance inequality, affecting Black lives disproportionately. Addressing this issue will require more than technological solutions. Researchers and experts in the field are overwhelmingly white and need to engage and listen to those experiencing the harm.

The Black Lives Matter movement has been a wake-up call for many in academia to reflect on the pervasive effects of racial injustice, both in terms of the underrepresentation of Black researchers and the harmful effects of technology, which often plays a discriminatory role in society. While blatant racism may not always be in plain sight, many mechanisms in society are in place that cause racial discrimination, and these mechanisms are often amplified by technology. On Wednesday 10 June, thousands of researchers (and editors) around the world joined in a #Strike4BlackLives initiative (also known as #ShutDownSTEM and #ShutDownAcademia) to spend the day learning about structural racism and finding ways to counter it. This should be the start of a long journey.

Many have raised the alarm in recent years over the potential for racial discrimination inherent in AI algorithms. Black researchers Timnit Gebru and Joy Buolamwini, among others, have worked tirelessly to identify and raise awareness of racial discrimination in AI and facial recognition technologies in particular. The harmful effect of facial recognition technologies on Black people was recently demonstrated with a real-world example when Robert Williams, a Black man, was wrongfully arrested based on his face being [mistakenly matched](#) with a robbery suspect. It is a well-known problem that commercial facial recognition systems have a higher error rate for darker-skinned faces, as they are typically trained on datasets that are insufficiently diverse and phenotypically [imbalanced](#). The solution to fairer AI systems may seem to be training on better, less biased datasets. However, this ignores the deeper problem with technologies like facial recognition, which are not separate from the social structures in which they are built. Emily Denton, a research scientist on Google's Ethical AI team, [tweeted](#) recently "AI research and development cannot be decoupled from the social relations

that structure our world and AI failures are intrinsically linked to asymmetries with respect to who is driving research agendas and shaping incentives". Answers to questions such as 'who is developing these algorithms and why', 'where does the training data (typically [images scraped from the web](#)) come from' and 'who decides on how the tools are used' will reveal numerous underlying structural imbalances.

One fundamental problem is that among those developing state-of-the-art algorithms, in computer science departments and the tech industry, the percentage of Black researchers is very low. While 6% of undergraduates in US computer science programmes are Black, this drops to 1–2% at the PhD and faculty level, while a 2018 [report](#) found that less than 2.5–4% of the workforce across major AI companies are Black. The importance of representation is clear when noting how activism among scientists and engineers can be pivotal in changing a company or organization's course of direction. IBM recently announced that they would no longer sell or develop facial recognition technologies, Amazon announced a one-year moratorium in selling their facial recognition systems to law enforcement, while Microsoft stated they would not sell such technologies to police forces until a national law governing their use is brought into place. This is welcome progress, but must not replace governmental action to ensure that less visible companies do not fill this gap.

Beyond facial recognition, there are many more algorithms used in society that discriminate against Black people, such as credit risk scoring and predictive policing technologies. As Cathy O'Neil discussed in her book *Weapons of Math Destruction*, big data gathering increases inequality and too often lends legitimacy to new forms of social and political oppression. One example is that the gathering of crime data, including for nuisance crime, reinforces the presence of police in poor and mostly Black neighbourhoods. This impacts arrests

in those neighbourhoods, producing more crime data and a harmful feedback loop. Cathy O'Neil mentions that those behind big data models like predictive policing often insist that their model, unlike humans, is blind to race and ethnicity. But it's a mistake to think technology can ever be neutral, as it is simply impossible to abstract away social context. In fact, as Ruha Benjamin argues in her recent book *Race after Technology*, race itself is a tool that is deeply embedded in society.

To fully grasp these issues, it will be essential to work with those in marginalized communities who have experienced the harmful effects of technologies. Conferences like the Association for Computing Machinery's meeting on Fairness, Accountability and Transparency (FAccT), which took place in Barcelona earlier this year, and the upcoming [workshop](#) 'Participatory approaches in machine learning' at the International Conference on Machine Learning seek to address the power imbalance by involving those who are directly impacted by machine learning systems. It seems an important and stimulating way forward for the design of AI tools to seek awareness of the [power dynamics](#) surrounding the application of those tools from the start.

At the same time, it is [exhausting for Black scholars](#) to keep explaining what they are grappling with on a daily basis in the face of a usual white audience who have not lived through the same experiences and struggle to accept structural racism exists. It is up to those with the power to shape the AI community, including science journals, to actively engage with issues around structural racism and ask difficult, often uncomfortable questions about how AI technologies might consolidate, or even intensify, racism in society. □

Published online: 15 July 2020  
<https://doi.org/10.1038/s42256-020-0210-5>