

World view



By Pratyusha Kalluri

Don't ask if AI is good or fair, ask how it shifts power

Those who could be exploited by artificial intelligence should be shaping its projects.

Law enforcement, marketers, hospitals and other bodies apply artificial intelligence (AI) to decide on matters such as who is profiled as a criminal, who is likely to buy what product at what price, who gets medical treatment and who gets hired. These entities increasingly monitor and predict our behaviour, often motivated by power and profits.

It is not uncommon now for AI experts to ask whether an AI is 'fair' and 'for good'. But 'fair' and 'good' are infinitely spacious words that any AI system can be squeezed into. The question to pose is a deeper one: how is AI shifting power?

From 12 July, thousands of researchers will meet virtually at the week-long International Conference on Machine Learning, one of the largest AI meetings in the world. Many researchers think that AI is neutral and often beneficial, marred only by biased data drawn from an unfair society. In reality, an indifferent field serves the powerful.

In my view, those who work in AI need to elevate those who have been excluded from shaping it, and doing so will require them to restrict relationships with powerful institutions that benefit from monitoring people. Researchers should listen to, amplify, cite and collaborate with communities that have borne the brunt of surveillance: often women, people who are Black, Indigenous, LGBT+, poor or disabled. Conferences and research institutions should cede prominent time slots, spaces, funding and leadership roles to members of these communities. In addition, discussions of how research shifts power should be required and assessed in grant applications and publications.

A year ago, my colleagues and I created the Radical AI Network, building on the work of those who came before us. The group is inspired by Black feminist scholar Angela Davis's observation that "radical simply means 'grasping things at the root'", and that the root problem is that power is distributed unevenly. Our network emphasizes listening to those who are marginalized and impacted by AI, and advocating for anti-oppressive technologies.

Consider an AI that is used to classify images. Experts train the system to find patterns in photographs, perhaps to identify someone's gender or actions, or to find a matching face in a database of people. 'Data subjects' – by which I mean the people who are tracked, often without consent, as well as those who manually classify photographs to train the AI system, usually for meagre pay – are often both exploited and evaluated by the AI system.

Researchers in AI overwhelmingly focus on providing highly accurate information to decision makers. Remarkably little research focuses on serving data subjects.

It is our responsibility to recognize our skewed perspective and listen to those impacted by AI."

Pratyusha Kalluri is a co-creator of the Radical AI Network and an AI researcher at Stanford University in California. e-mail: pkalluri@stanford.edu

What's needed are ways for these people to investigate AI, to contest it, to influence it or to even dismantle it. For example, the advocacy group Our Data Bodies is putting forward ways to protect personal data when interacting with US fair-housing and child-protection services. Such work gets little attention. Meanwhile, mainstream research is creating systems that are extraordinarily expensive to train, further empowering already powerful institutions, from Amazon, Google and Facebook to domestic surveillance and military programmes.

Many researchers have trouble seeing their intellectual work with AI as furthering inequity. Researchers such as me spend our days working on what are, to us, mathematically beautiful and useful systems, and hearing of AI success stories, such as winning Go championships or showing promise in detecting cancer. It is our responsibility to recognize our skewed perspective and listen to those impacted by AI.

Through the lens of power, it's possible to see why accurate, generalizable and efficient AI systems are not good for everyone. In the hands of exploitative companies or oppressive law enforcement, a more accurate facial recognition system is harmful. Organizations have responded with pledges to design 'fair' and 'transparent' systems, but fair and transparent according to whom? These systems sometimes mitigate harm, but are controlled by powerful institutions with their own agendas. At best, they are unreliable; at worst, they masquerade as 'ethics-washing' technologies that still perpetuate inequity.

Already, some researchers are exposing hidden limitations and failures of systems. They braid their research findings with advocacy for AI regulation. Their work includes critiquing inadequate technological 'fixes'. Other researchers are explaining to the public how natural resources, data and human labour are extracted to create AI.

Race-and-technology scholar Ruha Benjamin at Princeton University in New Jersey has encouraged us to "remember to imagine and craft the worlds you cannot live without, just as you dismantle the ones you cannot live within". In this vein, it is time to put marginalized and impacted communities at the centre of AI research – their needs, knowledge and dreams should guide development. This year, for example, my colleagues and I held a workshop for diverse attendees to share dreams for the AI future we desire. We described AI that is faithful to the needs of data subjects and allows them to opt out freely.

When the field of AI believes it is neutral, it both fails to notice biased data and builds systems that sanctify the status quo and advance the interests of the powerful. What is needed is a field that exposes and critiques systems that concentrate power, while co-creating new systems with impacted communities: AI by and for the people.