World view

By Mona Sloane

Here's what's missing in the quest to make AI fair

Developers of artificial intelligence must learn to collaborate with social scientists and the people affected by its applications.

Beginning in 2013, the Dutch government used an algorithm to wreak havoc in the lives of 25,000 parents. The software was meant to predict which people were most likely to commit childcare-benefit fraud, but the government did not wait for proof before penalizing families and demanding that they pay back years of allowances. Families were flagged on the basis of 'risk factors' such as having a low income or dual nationality. As a result, tens of thousands were needlessly impoverished, and more than 1,000 children were placed in foster care.

From New York City to California and the European Union, many artificial intelligence (AI) regulations are in the works. The intent is to promote equity, accountability and transparency, and to avoid tragedies similar to the Dutch childcare-benefits scandal.

But these won't be enough to make AI equitable. There must be practical know-how on how to build AI so that it does not exacerbate social inequality. In my view, that means setting out clear ways for social scientists, affected communities and developers to work together.

Right now, developers who design AI work in different realms from the social scientists who can anticipate what might go wrong. As a sociologist focusing on inequality and technology, I rarely get to have a productive conversation with a technologist, or with my fellow social scientists, that moves beyond flagging problems. When I look through conference proceedings, I see the same: very few projects integrate social needs with engineering innovation.

To spur fruitful collaborations, mandates and approaches need to be designed more effectively. Here are three principles that technologists, social scientists and affected communities can apply together to yield AI applications that are less likely to warp society.

Include lived experience. Vague calls for broader participation in AI systems miss the point. Nearly everyone interacting online — using Zoom or clicking reCAPTCHA boxes — is feeding into AI training data. The goal should be to get input from the most relevant participants.

Otherwise, we risk participation-washing: superficial engagement that perpetuates inequality and exclusion. One example is the EU AI Alliance: an online forum, open to anyone, designed to provide democratic feedback to the European Commission's appointed expert group on AI. When I joined in 2018, it was an unmoderated echo chamber of mostly men exchanging opinions, not representative of the population of the EU, the AI industry or relevant experts.

> "AI technologies are typically built at the request of people in power — which makes users vulnerable."

By contrast, social-work researcher Desmond Patton at Columbia University in New York City has built a machine-learning algorithm to help identify Twitter posts related to gang violence that relies on the expertise of Black people who have experience with gangs in Chicago, Illinois. These experts review and correct notes underlying the algorithm. Patton calls his approach Contextual Analysis of Social Media (see go.nature.com/3vnkdq7).

Shift power. AI technologies are typically built at the request of people in power — employers, governments, commerce brokers — which makes job applicants, parole candidates, customers and other users vulnerable. To fix this, the power must shift. Those affected by AI should not simply be consulted from the very beginning; they should select what problems to address and guide the process.

Disability activists have already pioneered this type of equitable innovation. Their mantra 'Nothing about us without us' means that those who are affected take a leading role in crafting technology, regulating it and implementing it. For example, activist Liz Jackson developed the transcription app Thisten when she saw her community's need for real-time captions at the SXSW film festival in Austin, Texas.

Check AI's assumptions. Regulations, such as New York City's December 2021 law that regulates the sale of AI used in hiring, are increasingly requiring that AI pass audits meant to flag bias. But some of the guidelines are so broad that audits could end up validating oppression.

For example, pymetrics in New York is a company that uses neuroscience-based games to assess job candidates by measuring their "cognitive, social and behavioral attributes". An audit found that the firm did not violate US anti-discrimination law. But it did not consider whether such games are a reasonable way to examine suitability for a job, or what other dynamics of inequity could be introduced. This is not the kind of audit we need to make AI more just.

We need AI audits to weed out harmful tech. For example, with two colleagues, I developed a framework in which qualitative work inspects the assumptions that an AI is built on, and uses them as a basis for the technical part of an AI audit. This has informed an audit of Humantic AI and Crystal, two AI-driven personality tools used in hiring.

Each of these principles can be applied intuitively and will be self-reinforcing as technologists, social scientists, and members of the public learn how to implement them. Vague mandates won't work, but with clear frameworks, we can weed out AI that perpetuates discrimination against the most vulnerable people, and focus on building AI that makes society better.

Mona Sloane is a sociologist at New York University in New York City.
e-mail: mona.sloane@nyu.edu

The author declares competing interests; see go.nature.com/3kqbftz for details.

CATARINA HEECKT

© 2022 Springer Nature Limited. All rights reserved.