AI industry is influencing the world. Mozilla adviser Abeba Birhane is challenging its core values

July 22 2024, by Matt O'brien

Credit: AP Illustration/Jenni Sohn

"Scaling up" is a catchphrase in the artificial intelligence industry as tech companies rush to improve their AI systems with ever-bigger sets of

internet data.

It's also a red flag for Mozilla's Abeba Birhane, an AI expert who for years has challenged the values and practices of her field and the influence it's having on the world.

[Her latest research](link) finds that scaling up on online data used to train popular AI image-generator tools is disproportionately resulting in racist outputs, especially against Black men.

Birhane is a senior adviser in AI accountability at the Mozilla Foundation, the nonprofit parent organization of the free software company that runs the Firefox web browser. Raised in Ethiopia and living in Ireland, she's also an adjunct assistant professor at Trinity College Dublin.

Her interview with The Associated Press has been edited for length and clarity.

Q: How did you get started in the AI field?

A: I'm a cognitive scientist by training. Cog sci doesn't have its own department wherever you are studying it. So where I studied, it was under computer science. I was placed in a lab full of machine learners. They were doing so much amazing stuff and nobody was paying attention to the data. I found that very amusing and also very interesting because I thought data was one of the most important components to the success of your model. But I found it weird that people don't pay that much attention or time asking, 'What's in my dataset?' That's how I got interested in this space. And then eventually, I started doing audits of large scale datasets.

Q: Can you talk about your work on the ethical foundations of AI?

A: Everybody has a view about what machine learning is about. So machine learners—people from the AI community—tell you that it doesn't have a value. It's just maths, it's objective, it's neutral and so on. Whereas scholars in the social sciences tell you that, just like any technology, machine learning encodes the values of those that are fueling it. So what we did was we systematically studied a hundred of the most influential machine learning papers to actually find out what the field cares about and to do it in a very rigorous way.

A: And one of those values was scaling up?

Q: Scale is considered the holy grail of success. You have researchers coming from big companies like DeepMind, Google and Meta, claiming that scale beats noise and scale cancels noise. The idea is that as you scale up, everything in your dataset should kind of even out, should kind of balance itself out. And you should end up with something like a normal distribution or something closer to the ground truth. That's the idea.

Q: But your research has explored how scaling up can lead to harm. What are some of them?

A: At least when it comes to hateful content or toxicity and so on, scaling these datasets also scales the problems that they contain. More specifically, in the context of our study, scaling datasets also scales up hateful content in the dataset. We measured the amount of hateful content in two datasets. Hateful content, targeted content and aggressive content increased as the dataset was scaled from 400 million to 2 billion. That was a very conclusive finding that shows that scaling laws don't really hold up when it comes to training data. (In another paper) we found that darker-skinned women, and men in particular, tend to be allocated the labels of suspicious person or criminal at a much higher rate.

Q: How hopeful or confident are you that the AI industry will make the changes you've proposed?

A: These are not just pure mathematical, technical outputs. They're also tools that shape society, that influence society. The recommendations are that we also incentivize and pay attention to values such as justice, fairness, privacy and so on. My honest answer is that I have zero confidence that the industry will take our recommendations. They have never taken any recommendations like this that actually encourage them to take these societal issues seriously. They probably never will. Corporations and big companies tend to act when it's legally required. We need a very strong, enforceable regulation. They also react to public outrage and public awareness. If it gets to a state where their reputation is damaged, they tend to make change.