THE BATTLE TO EMBED ETHICS IN AI RESEARCH

Bias and the prospect of societal harm increasingly plague the field of artificial intelligence.

By Elizabeth Gibney

Diversity and inclusion took centre stage at one of the world's major artificial-intelligence (AI) conferences in 2018. But at last month's Neural Information Processing Systems (NeurIPS) conference in Vancouver, Canada, attention shifted to another big issue in the field: ethics.

The focus comes as AI research increasingly deals with ethical controversies surrounding the application of its technologies — such as in predictive policing or facial recognition. Problems include tackling biases in algorithms that reflect existing patterns of discrimination in data, and avoiding disproportionately affecting already-vulnerable populations. "There is no such thing as a neutral tech platform," warned Celeste Kidd, a developmental psychologist at the University of California, Berkeley, during her NeurIPS keynote talk about how algorithms can influence human beliefs. At the meeting, which hosted a record 13,000 attendees, researchers grappled with how to meaningfully address the ethical and societal implications of their work.

Ethicists have long debated the impacts of AI and sought ways to use the technology for good, such as in health care. But researchers are now realizing that they need to embed ethics in the formulation of their research and understand the potential harms of algorithmic injustice, says Meredith Whittaker, an AI researcher at New York University and founder of the AI Now Institute, which seeks to understand the social implications of the technology. At the latest NeurIPS, researchers couldn't "write, talk or think" about these systems without considering possible social harms, she says. "The question is, will the change in the conversation result in the structural change we need to actually ensure these systems don't cause harm?"

Meetings matter

Conferences such as NeurIPS, which, together with two other annual meetings, publishes the majority of papers in AI, bear some responsibility, says Whittaker. "The field has blown up so much there aren't enough conferences or reviewers. But everybody wants their paper in. So there is huge leverage there," she says.

But research presented at NeurIPS doesn't face a specific ethics check as part of the review process. The pitfalls of this were encapsulated by the reaction to one paper presented at the conference, in which faces — including aspects of a person's age, gender and ethnicity — were generated on the basis of voices. Machine-learning scientists criticized it on Twitter as transphobic and pseudoscientific.

One solution could be to introduce ethical review at conferences. For the first time, NeurIPS 2019 included a reproducibility checklist for submitted papers. In the future, once accepted, papers could also be checked for responsibility, says Joelle Pineau, a machine-learning scientist at McGill University in Montreal, Canada, and at Facebook, who is on the NeurIPS organizing committee and developed the checklist.

NeurIPS says that an ethics committee is on hand to deal with concerns during the existing review process, but it is considering ways to make its work on ethical and societal impacts more robust. Proposals include asking authors to make a statement about the ethics of their work. The organizers of the annual International Conference on Learning Representations — another of the major AI meetings — said it was discussing the idea of reviewing papers with ethics in mind.

AI Now goes a step further: in a report published last month, it called for all machine-learning research papers to include a section on societal harms, as well as the provenance of their data sets. Such considerations should centre on the perspectives of vulnerable groups, which AI tends to affect disproportionately, Abeba Birhane, a cognitive scientist at University College Dublin, told NeurIPS's Black in AI workshop. Developers should ask not only how their algorithm might be used, but also whether it is necessary, she said.

Business influences

Tech companies — which are responsible for vast amounts of AI research — are also addressing the ethics of their work. But activists say that they must not be allowed to get away with 'ethics-washing'. Tech firms suffer from a lack of diversity, and although some companies have staff and entire boards dedicated to ethics, campaigners warn that these often have too little power. Their technical solutions — which include efforts to 'debias algorithms' — are also often misguided, says Birhane. The approach wrongly suggests that bias-free data sets exist, and fixing algorithms doesn't solve the root problems in underlying data, she says.

Forcing tech companies to include people from affected groups on ethics boards would help, said Fabian Rogers, a community organizer from New York City. Rogers represents a tenants' association that fought to stop its landlord from installing facial-recognition technology without residents' consent. "Context is everything, and we need to keep that in mind when we're talking about technology. It's hard to do that when we don't have necessary people to offer that perspective," he said.

Facial-recognition algorithms have been at the centre of privacy and ethics debates.