

# Tech industry ethics teams lack resources and authority, and their effectiveness is spotty at best, study finds

July 26 2023, by Katharine Miller

---



Credit: Unsplash/CC0 Public Domain

In recent years, AI companies have been publicly chided for generating machine learning algorithms that discriminate against historically

marginalized groups. To quell that criticism, many companies pledged to ensure their products are fair, transparent, and accountable, but these promises are frequently criticized as being mere "ethics washing," says Sanna Ali, who recently received her Ph.D. from the Stanford University Department of Communication in the School of Humanities and Sciences. "There's a concern that these companies talk the talk but don't walk the walk."

To explore whether that's the case, Ali interviewed AI ethics workers from some of the largest companies in the field. The research project, co-authored with Stanford Assistant Professor of Communication Angèle Christin, Google researcher Andrew Smart, and Stanford W.M Keck Professor and Professor of Management Science and Engineering Riitta Katila, was published in the *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*.

The study found that ethics initiatives and interventions were difficult to implement in the tech industry's institutional environment. Specifically, Ali found, teams were largely under-resourced and under-supported by leadership, and they lacked authority to act on problems they identified.

"Without leadership buy-in, individual workers had to employ persuasive skills and interpersonal strategies in order to make any headway," Ali says. The result: They succeed in working with some teams and not with others, and are often called for a consultation too close to product launch dates and without the authority to require important ethics fixes.

Ali's interviews suggest some solutions: Leadership should incentivize product teams to incorporate ethics considerations into product development processes, she says, and there should be bureaucratic support to empower ethics teams in their work and give them the authority to implement necessary ethics fixes before products are released.

"It's unlikely that these companies are going to change their priority of frequently releasing new products," Ali says. "But at least they could provide incentives so that ethics can be part of that conversation early on."

## **Ethics and the AI industry's institutional environment**

Many tech companies have released statements of principles around accountability, transparency, and fairness, Ali says. They have also developed toolkits for evaluating algorithmic fairness; held seminars about how to implement responsible AI; and hired ethics teams, which go by various names such as "Trust and Safety" or "Responsible AI."

In theory, these ethics teams provide expert support for such things as addressing problems with training data; identifying and implementing various fairness fixes to machine learning models after they've been trained; or evaluating whether a model is sufficiently explainable for its intended use.

That's all well and good, Ali says, "but we wanted to look at the challenges of implementing those initiatives and interventions on the ground."

To do that, she interviewed 25 tech industry employees, including 21 who either currently work or have worked as part of responsible AI initiatives—many of them at more than one company—at businesses with 6,000 to hundreds of thousands of employees.

Prior research has identified some important characteristics of the tech industry, including businesses' tendency to be informal and nonhierarchical; to value rapid product innovation over all other

concerns; and to think that tech can fix tech.

Based on this background, Ali hypothesized that upper-level leadership's responsible AI principles were likely to be "decoupled" from the horizontally distributed product teams where they would be implemented. She compares it to a school principal who states an [educational policy](#) but often has little control over what happens in individual teachers' classrooms.

This would, she further hypothesized, leave ethics teams in the position of acting as "ethics entrepreneurs" who would have to "sell" their services to individual product teams. In essence, they would be left to their own resources to build relationships with product managers in hopes of gaining their cooperation in an ethics review of their products.

## **Ethics teams lack support, resources, and authority**

Based on her interviews, Ali says, many of her predictions about the institutional environment and its expected impact on ethics teams panned out. Company policies were indeed decoupled from implementation by distributed product development teams, and ethics teams had to carefully cultivate relationships with product teams to get anything done.

Ali found that the implementation of responsible AI policies in the tech industry is inconsistent at best. Indeed, [products](#) might be released without ethics team input for multiple reasons. Sometimes a product team wouldn't want to work with the ethics team at all and ethics personnel lacked authority to mandate an ethics review. Sometimes the ethics team would be invited to provide input too close to a product launch date when there was only enough support or authority to implement a few of the necessary fixes before launch. Sometimes product teams believed that ethics workers' fairness goals would conflict with other important goals such as user engagement. Sometimes the

ethics team would ask management to delay a product launch so they could implement an ethics fix, but the manager in charge declined the request.

"It just takes a person with more authority than the ethics worker to speak up," Ali says. "But that's not happening because all of the incentives are around launching the product immediately."

In some cases, companies have implemented a more formal ethics review process where, early in the development of a new product idea, the product team completes an impact assessment that the ethics team reviews. If the product relates to a sensitive use—bail or sentencing, for example—the ethics and product teams then work together to determine the product's potential to treat certain demographic groups unfairly.

If indeed there exists such a potential, then the ethics team will be included in the development of the product from start to finish. "In that setting, the team might have more resources and authority to do something to make sure the AI is deployed responsibly," Ali says.

While Ali favors this more bureaucratic approach, there's a risk of it becoming a box-checking exercise, she says. "The product team might, for example, check the boxes identifying steps they are willing to take while ignoring those that would require deeply thinking about the real ethical issues." And, because responsible AI is a relatively new field, there is a need for such deep thinking, she says.

For example, there are still debates about what fairness means, how to measure it, and how fair is fair enough. Yet ethics workers are tasked with navigating that uncertainty without support, resources, and authority to act, all while functioning inside a business context where fast innovation is prioritized. It's a daunting task, Ali says.

## **Solutions: Incentives, bureaucracy, authority**

Diplomatically approaching one product team after another in hopes of collaborating only gets ethics workers so far. They need some formal authority to require that problems be addressed, Ali says. "An ethics worker who approaches product teams on an equal footing can simply be ignored," she says.

And if ethics teams are going to implement that authority in the horizontal, nonhierarchical world of the tech industry, there need to be formal bureaucratic structures requiring ethics reviews at the very beginning of the product development process, Ali says. "Bureaucracy can set rules and requirements so that ethics workers don't have to convince people of the value of their work."

Product teams also need to be incentivized to work with ethics teams, Ali says. "Right now, they are very much incentivized by moving fast, which can be directly counter to slowly, carefully, and responsibly examining the effects of your technology," Ali says. Some interviewees suggested rewarding teams by giving them "ethics champion" bonuses when a product is made less biased or when the plug is pulled on a product that has a serious problem.

"It would be good to acknowledge the ethical stance that people are taking within the company by rewarding it in some way," Ali says.

By creating some bureaucracy, empowering ethics teams, and incentivizing other employees to work with the ethics teams, [tech companies](#)' promises of fairness will no longer be decoupled from work on the ground. Then, Ali says, "real institutional change may be possible."

**More information:** Sanna J. Ali et al, Walking the Walk of AI Ethics:

Organizational Challenges and the Individualization of Risk among Ethics Entrepreneurs, *2023 ACM Conference on Fairness, Accountability, and Transparency* (2023). [DOI: 10.1145/3593013.3593990](https://doi.org/10.1145/3593013.3593990)

Provided by Stanford University

Citation: Tech industry ethics teams lack resources and authority, and their effectiveness is spotty at best, study finds (2023, July 26) retrieved 3 May 2024 from <https://techxplore.com/news/2023-07-tech-industry-ethics-teams-lack.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.