

# Walking the Walk of AI Ethics in Technology Companies

Sanna J. Ali, Angèle Christin, Andrew Smart, and Riitta Katila

**THE FIELD OF AI ETHICS HAS GROWN RAPIDLY IN INDUSTRY AND ACADEMIA, in large part due to the “techlash” brought about by technology industry scandals such as Cambridge Analytica and growing congressional attention to technology giants’ data privacy and other internal practices. In recent years, technology companies have published AI principles, hired social scientists to conduct research and compliance, and employed engineers to develop technical solutions related to AI ethics and fairness. Despite these new initiatives, many private companies have not yet prioritized the adoption of accountability mechanisms and ethical safeguards in the development of AI. Companies often “talk the talk” of AI ethics but rarely “walk the walk” by adequately resourcing and empowering teams that work on responsible AI.**

In our paper, “Walking the Walk of AI Ethics,” we present one of the first empirical investigations into AI ethics on the ground in a (thus far) fairly unregulated environment within the technology sector. Our interviews with AI ethics workers in the private sector uncovered several significant obstacles to implementing AI ethics initiatives. Practitioners struggle to have their

## Key Takeaways

Technology companies often “talk the talk” of AI ethics without fully “walking the walk.” Many companies have released AI principles, but relatively few have institutionalized meaningful change.

We interviewed 25 AI ethics practitioners and found that there are significant roadblocks to implementing companies’ stated goals regarding AI ethics.

AI ethics and fairness considerations are championed by individuals who lack institutional support, rarely made a priority in product development cycles, disincentivized by metrics, and disrupted by the frequent reorganization of teams.

Government regulation could play a crucial role in helping the AI ethics field move toward formalization by incentivizing leaders to prioritize ethical issues and protecting AI ethics workers.

companies foreground ethics in an environment centered around software product launches. Ethics are difficult to quantify and easy to de-prioritize in a context where company goals are incentivized by metrics. And the frequent reorganization of teams at technology companies makes it challenging for AI ethics workers to access institutional knowledge and maintain relationships central to their work.

Our research highlights the stark gap between company policy and practice when it comes to AI ethics. It captures the difficulties of institutionalizing change within technology companies and illustrates the important role of regulation in incentivizing companies to make AI ethics initiatives a priority.

## Introduction

Previous research has criticized corporate AI ethics principles for being toothless and vague, while questioning some of their underlying assumptions. However, relatively few studies have examined the implementation of AI ethics initiatives on the ground, let alone the organizational dynamics that contribute to the lack of progress.

Our paper builds on existing research by drawing on theories of organizational change to shed light on the ways that AI ethics workers operate in technology companies. In response to outside pressure, such as regulation and public backlash, many organizations develop policies and practices to gain legitimacy; however, these measures often do not achieve their intended outcome as there is a disconnect between means and ends. New practices may also go against the organization's established rules and procedures.

---

*Many technology companies have released AI principles, but relatively few have made significant adjustments to their operations as a result.*

---

AI ethics initiatives suffer from the same dynamic: Many technology companies have released AI principles, but relatively few have made significant adjustments to their operations as a result. With little buy-in from senior leadership, AI ethics workers take on the responsibility of organizational change by using persuasive strategies and diplomatic skills to convince engineers and product managers to incorporate ethical considerations in product development. Technology companies also seek to move quickly and release products regularly to generate investment and to outpace competitors, meaning that products are often released despite ethical concerns. Responsible AI teams may be siloed within large organizations, preventing their work from becoming integral to the core tasks of the organization.

To better understand the concrete organizational barriers to the implementation of AI ethics initiatives, we conducted a qualitative study of responsible AI initiatives within technology companies. We interviewed 25 AI ethics practitioners, including employees, academics, and consultants—many of whom are currently or were formerly employed

as part of technology companies' responsible AI initiatives—in addition to gathering observations from industry workshops and training programs. Our resulting analysis provides insight into the significant structural risks workers face when they advocate for ethical AI as well as the hurdles they encounter when incorporating AI ethics into product development.

---

*Product managers frequently perceive responsible AI teams' activities as stalling product launches or putting revenue generation at risk.*

---

## Research Outcomes

Our research uncovered a wide array of roadblocks that prevent technology companies from making meaningful progress on AI ethics.

First, AI ethics is usually promoted within companies by individual workers—or, as we call them, “ethics entrepreneurs”—who lack institutional support. Lack of buy-in often prevents them from being able to ensure that AI products are developed responsibly. It can also present considerable risks for these individuals, particularly those from marginalized backgrounds, as they are tasked with speaking up about ethical issues yet lack the authority to direct colleagues to make changes to the product development process. In the words of one AI ethics worker, “Being very loud about putting more brakes on [AI development] was a risky thing to do. It was not built into the process” and required appealing to leadership. Meanwhile, frequent team reorganizations disrupt crucial peer relationships that AI ethic workers rely on to make progress on their projects.

Second, the goal of product innovation supersedes goals related to responsible AI. According to interviewees, while product managers “often bring all the right resources together to drive product development,” AI ethics workers frequently face

resource constraints. They have to worry about finding the time and funding to build new or improve existing datasets, as well as conduct fairness evaluations. Product managers frequently perceive responsible AI teams' activities as stalling product launches or putting revenue generation at risk, which may lead them to allocate inadequate time and resources to responsible AI initiatives.

Product teams are also resistant to ethics reviews because they are often not conducted until a product is ready for launch, meaning that teams have to course-correct just before a launch deadline. The AI ethics workers we spoke with navigated these challenges by reframing ethics-oriented interventions as a matter of product quality and, where possible, working with teams at early stages of the product development cycle. Some have attempted to persuade product teams to use automated ethics tools to reduce friction.

Finally, performance metrics, not ethical considerations, determine a technology company's behavior. Metrics around engagement or the performance of AI models are so highly prioritized that ethics-related recommendations that might

---

*Quantitative metrics of ethics or fairness are hard to come by and challenging to define.*

---

negatively affect those metrics require irrefutable quantitative evidence. Yet quantitative metrics of ethics or fairness are hard to come by and challenging to define given that companies' existing data infrastructures are not tailored to such metrics. As one interviewee noted, "No matter how much people want to do a good job, at the end of the day, quarterly goals will get in the way."

## Policy Discussion

Our research highlights that AI ethics work in technology companies is predominantly conducted by individual AI ethics entrepreneurs who struggle to effect change through informal processes. The many systemic barriers they face severely compromise companies' ability to address AI ethics issues adequately and consistently.

Government regulation could play a crucial role in helping the AI ethics field move toward formalization. AI ethics workers consistently noted that regulation would incentivize leadership to take responsible AI seriously, especially if there were financial penalties for noncompliance. Rules and legal standards would also provide valuable guidance for AI ethics workers

who have to make difficult trade-offs around fairness. Without regulation, individuals are often left to their own devices to navigate thorny decisions such as how to ensure equitable representation of different demographic groups in algorithmic promotion of content.

Given the premium companies place on product innovation, government policies could also empower AI ethics workers to limit the expansion of AI to new use cases that raise major ethical concerns. Technology companies are eager to develop and apply new AI products or features, which makes it hard or even impossible for ethics workers to suggest that the company should not use a new AI feature. Widely accepted standards and norms could change this. Stronger whistleblower protections would help AI ethics workers promote responsible AI without fear of retaliation; such protections should be adopted by both companies and governments. Mandating the integration of ethics initiatives at the early stages of projects could also meaningfully advance responsible AI efforts.

---

*Regulation would incentivize leadership to take responsible AI seriously, especially if there were financial penalties for noncompliance.*

---

At the same time, a one-size-fits-all regulatory approach could make matters worse if, for instance, it requires all technology platforms to collect demographic information about their users. Government policies that aim to standardize or mandate fairness by creating technical toolkits and setting quantitative thresholds could backfire; by formalizing complex processes, such measures could flatten nuance and even introduce new potentials for ethical harms since universal metrics may not be appropriate for every context or application area.

Committed employees have the capacity to act as catalysts for organizational change and help technology companies “walk the walk” on AI ethics. But in order to achieve meaningful implementation of AI ethics initiatives, technology companies will have to grapple with a variety of institutional barriers.

Reference: The original article is accessible at Sanna J. Ali et al., “**Walking the Walk of AI Ethics: Organizational Challenges and the Individualization of Risk among Ethics Entrepreneurs**,” *FACCT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (June 2023): 217–226, <https://dl.acm.org/doi/10.1145/3593013.3593990>.

---

[Stanford University’s Institute on Human-Centered Artificial Intelligence \(HAI\)](#) applies rigorous analysis and research to pressing policy questions on artificial intelligence. A pillar of HAI is to inform policymakers, industry leaders, and civil society by disseminating scholarship to a wide audience. HAI is a nonpartisan research institute, representing a range of voices. The views expressed in this policy brief reflect the views of the authors. For further information, please contact [HAI-Policy@stanford.edu](mailto:HAI-Policy@stanford.edu).



**Sanna J. Ali** is a policy analyst at the Cyber Policy Center and the Regulation, Evaluation, and Governance Lab (RegLab) at Stanford University.



**Angèle Christin** is an associate professor of Communication and the Faculty Director of the Digital Civil Society Lab at Stanford University.



**Andrew Smart** is a sociotechnical researcher in the Responsible AI Impact Lab at Google.



**Riitta Katila** is the W. M. Keck Sr. Professor of Management Science and Engineering, HAI Sabbatical Scholar, and Faculty Director of Stanford Technology Ventures Program (STVP), the Engineering Entrepreneurship Center at Stanford University.



**Stanford HAI:** 353 Jane Stanford Way, Stanford CA 94305-5008

**T** 650.725.4537 **F** 650.123.4567 **E** [HAI-Policy@stanford.edu](mailto:HAI-Policy@stanford.edu) [hai.stanford.edu](http://hai.stanford.edu)