

## MIT Technology Review

## MIT Technology Review

### ARTIFICIAL INTELLIGENCE

## Responsible AI has a burnout problem

Companies say they want ethical AI. But those working in the field say that ambition comes at their expense.

By Melissa Heikkilä

October 28, 2022



STEPHANIE ARNETT/MITTR | UNSPLASH

**Margaret Mitchell had been working at Google for two years before she realized she needed a break.**

“I started having regular breakdowns,” says Mitchell, who founded and co-led the company’s Ethical AI team. “That was not something that I had ever experienced before.”

## MIT Technology Review

Mitchell, who now works as an AI researcher and chief ethics scientist at the AI startup Hugging Face, is far from alone in her experience. Burnout is becoming increasingly common in responsible-AI teams, says Abhishek Gupta, the founder of the Montreal AI Ethics Institute and a responsible-AI consultant at Boston Consulting Group.

Companies are under increasing pressure from regulators and activists to ensure that their AI products are developed in a way that mitigates any potential harms before they are released. In response, they have invested in teams that evaluate how our lives, societies, and political systems are affected by the way these systems are designed, developed, and deployed.

Tech companies such as Meta have been forced by courts to offer compensation and extra mental-health support for employees such as content moderators, who often have to sift through graphic and violent content that can be traumatizing.

But teams who work on responsible AI are often left to fend for themselves, employees told MIT Technology Review, even though the work can be just as psychologically draining as content moderation. Ultimately, this can leave people in these teams feeling undervalued, which can affect their mental health and lead to burnout.

Rumman Chowdhury, who leads Twitter's Machine Learning Ethics, Transparency, and Accountability team and is another pioneer in applied AI ethics, faced that problem in a previous role.

## MIT Technology Review

All the practitioners MIT Technology Review interviewed spoke enthusiastically about their work: it is fueled by passion, a sense of urgency, and the satisfaction of building solutions for real problems. But that sense of mission can be overwhelming without the right support.

“It almost feels like you can’t take a break,” Chowdhury says. “There is a swath of people who work in tech companies whose job it is to protect people on the platform. And there is this feeling like if I take a vacation, or if I am not paying attention 24/7, something really bad is going to happen.”

Mitchell continues to work in AI ethics, she says, “because there’s such a need for it, and it’s so clear, and so few people see it who are actually in machine learning.”

But there are plenty of challenges. Organizations place huge pressure on individuals to fix big, systemic problems without proper support, while they often face a near-constant barrage of aggressive criticism online.

### Cognitive dissonance

#### Related Story



#### **How Facebook got addicted to spreading misinformation**

The company’s AI algorithms gave it an insatiable habit for lies and hate speech. Now the man who built them can’t fix the problem.

The role of an AI ethicist or someone in a responsible-AI team varies widely, ranging from analyzing the societal effects of AI systems to developing responsible strategies and policies to fixing technical issues. Typically, these workers are also tasked with coming up with ways to mitigate AI harms, from algorithms that spread hate speech to systems that allocate things like housing and benefits in a discriminatory way to the spread of graphic and violent images and language.

## MIT Technology Review

Trying to fix deeply ingrained issues such as racism, sexism, and discrimination in AI systems might, for example, involve analyzing large data sets that include extremely toxic content, such as rape scenes and racial slurs.

AI systems often reflect and exacerbate the worst problems in our societies, such as racism and sexism. The problematic technologies range from facial recognition systems that classify Black people as gorillas to deepfake software used to make porn videos appearing to feature women who have not consented. Dealing with these issues can be especially taxing to women, people of color, and other marginalized groups, who tend to gravitate toward AI ethics jobs.

And while burnout is not unique to people working in responsible AI, all the experts MIT Technology Review spoke to said they face particularly tricky challenges in that area.

“You are working on a thing that you’re very personally harmed by day to day,” Mitchell says. “It makes the reality of discrimination even worse because you can’t ignore it.”

But despite growing mainstream awareness about the risks AI poses, ethicists still find themselves fighting to be recognized by colleagues in the AI field.

Some even disparage the work of AI ethicists. Stability AI’s CEO, Emad Mostaque, whose startup built the open-source text-to-image AI Stable Diffusion, said in a tweet that ethics debates around his technology are “paternalistic.” Neither Mostaque nor Stability AI replied to MIT Technology Review’s request for comment by the time of publishing.

“People working in the AI field are mostly engineers. They’re not really open to humanities,” says Emmanuel Goffi, an AI ethicist and founder of the Global AI Ethics Institute, a think tank.

Companies want a quick technical fix, Goffi says; they want someone to “explain to them how to be ethical through a PowerPoint with three slides and four bullet points.” Ethical thinking needs to go deeper, and it should be applied to how the whole organization functions, Goffi adds.

“Psychologically, the most difficult part is that you have to make compromises every day—every minute—between what you believe in and what you have to do,” he says.

## MIT Technology Review

you have to prove that they're worth working on. So it's completely the opposite of support. It's pushback."

Chowdhury adds, "There are people who think ethics is a worthless field and that we're negative about the progress [of AI]."

Social media also makes it easy for critics to pile on researchers. Chowdhury says there's no point in engaging with people who don't value what they do, "but it's hard not to if you're getting tagged or specifically attacked, or your work is being brought up."

### Breakneck speed

The rapid pace of artificial-intelligence research doesn't help either. New breakthroughs come thick and fast. In the past year alone, tech companies have unveiled AI systems that generate images from text, only to announce—just weeks later—even more impressive AI software that can create videos from text alone too. That's impressive progress, but the harms potentially associated with each new breakthrough can pose a relentless challenge. Text-to-image AI could violate copyrights, and it might be trained on data sets full of toxic material, leading to unsafe outcomes.

"Chasing whatever's really trendy, the hot-button issue on Twitter, is exhausting," Chowdhury says. Ethicists can't be experts on the myriad different problems that every single new breakthrough poses, she says, yet she still feels she has to keep up with every twist and turn of the AI information cycle for fear of missing something important.

## MIT Technology Review

things won't fall apart, because I'm not the only person doing it," she says.

But Chowdhury works at a big tech company with the funds and desire to hire an entire team to work on responsible AI. Not everyone is as lucky.

People at smaller AI startups face a lot of pressure from venture capital investors to grow the business, and the checks that you're written from contracts with investors often don't reflect the extra work that is required to build responsible tech, says Vivek Katial, a data scientist at Multitudes, an Australian startup working on ethical data analytics.

The tech sector should demand more from venture capitalists to "recognize the fact that they need to pay more for technology that's going to be more responsible," Katial says.

The trouble is, many companies can't even see that they have a problem to begin with, according to a [report released by MIT Sloan Management Review and Boston Consulting Group](#) this year. AI was a top strategic priority for 42% of the report's respondents, but only 19% said their organization had implemented a responsible-AI program.

Some may believe they're giving thought to mitigating AI's risks, but they simply aren't hiring the right people into the right roles and then giving them the resources they need to put responsible AI into practice, says Gupta.

"That's where people start to experience frustration and experience burnout," he adds.

### Growing demand

Before long, companies may not have much choice about whether they back up their words on ethical AI with action, because regulators are starting to introduce AI-specific laws.

## MIT Technology Review

on regulation for the use of AI in high-risk sectors such as employment. In early October, the White House unveiled the [AI Bill of Rights](#), which lays out five rights Americans should have when it comes to automated systems. The bill is likely to spur federal agencies to increase their scrutiny of AI systems and companies.

And while the volatile global economy has led many tech companies to freeze hiring and threaten major layoffs, responsible-AI teams have arguably never been more important, because rolling out unsafe or illegal AI systems could expose the company to huge fines or requirements to delete their algorithms. For example, last spring the US Federal Trade Commission forced Weight Watchers to delete its algorithms after the company was found to have illegally collected data on children. Developing AI models and collecting databases are significant investments for companies, and being forced by a regulator to completely delete them is a big blow.

Burnout and a persistent sense of being undervalued could lead people to leave the field entirely, which could harm the field of AI governance and ethics research as a whole. It's especially risky given that those with the most experience in solving and addressing harms caused by an organization's AI may be the most exhausted.

"The loss of just one person has massive ramifications across entire organizations," Mitchell says, because the expertise someone has accumulated is extremely hard to replace. In late 2020, Google sacked its ethical AI co-lead [Timnit Gebru](#), and it fired Mitchell a few months later. [Several other members](#) of its responsible-AI team left in the space of just a few months.

Gupta says this kind of brain drain poses a "severe risk" to progress in AI ethics and makes it harder for companies to adhere to their programs.

Last year, Google [announced](#) it was doubling its research staff devoted to AI ethics, but it has not commented on its progress since. The company told MIT Technology Review it offers training on mental-health resilience, has a peer-to-peer mental-health support initiative, and gives employees access to digital tools to help with mindfulness. It can also connect them with mental-health providers virtually. It did not respond to questions about Mitchell's time at the company.

Meta said it has invested in benefits like a program that gives employees and their families access to 25 free therapy sessions each year. And Twitter said it offers employee counseling and coaching sessions and burnout prevention training. The company also has a peer-support program focused on mental health. None of the companies said they offered support tailored specifically for AI ethics.

## MIT Technology Review

Change starts from the very top. “Executives need to speak with their dollars, their time, their resources, that they’re allocating to this,” he says. Otherwise, people working on ethical AI “are set up for failure.”

Successful responsible-AI teams need enough tools, resources, and people to work on problems, but they also need agency, connections across the organization, and the power to enact the changes they're being asked to make, Gupta adds.

A lot of mental-health resources at tech companies center on time management and work-life balance, but more support is needed for people who work on emotionally and psychologically jarring topics, Chowdhury says. Mental-health resources specifically for people working on responsible tech would also help, she adds.

“There hasn’t been a recognition of the effects of working on this kind of thing, and definitely no support or encouragement for detaching yourself from it,” Mitchell says.

“The only mechanism that big tech companies have to handle the reality of this is to ignore the reality of it.” **T**

by Melissa Heikkilä

## DEEP DIVE

# ARTIFICIAL INTELLIGENCE