

# “I started crying”: Inside Timnit Gebru’s last days at Google— and what happens next

MIT Technology Review, by Karen Hao, December 16, 2020

Find original article here: <https://www.technologyreview.com/2020/12/16/1014634/google-ai-ethics-lead-timnit-gebru-tells-story/>



By now, we’ve all heard some version of the story. On December 2, after a protracted disagreement over the release of a research paper, Google [forced out its ethical AI co-lead, Timnit Gebru](#). The paper was on the [risks of large language models](#), AI models trained on staggering amounts of text data, which are a line of research core to Google’s business. Gebru, a leading voice in AI ethics, was one of the only Black women at Google Research.

The move has since sparked a debate about growing corporate influence over AI, the long-standing lack of diversity in tech, and what it means to do meaningful AI ethics research. As of December 15, over 2,600 Google employees and 4,300 others in academia, industry, and civil society had [signed a petition](#) denouncing the dismissal of Gebru, calling it “unprecedented research censorship” and “an act of retaliation.”

Gebru is known for foundational work in [revealing AI discrimination](#), developing [methods for documenting](#) and [auditing AI models](#), and advocating for greater diversity in research. In 2016, she cofounded [the nonprofit Black in AI](#), which has become a central resource for civil rights activists, labor organizers, and leading AI ethics researchers, cultivating and highlighting Black AI research talent.

Losing her job didn’t slow Gebru down. The following week, she took part in several workshops at NeurIPS, the largest annual AI research conference, which over 20,000 people attended this year. It was “therapeutic,” she says, to see how the community she’d helped build showed up and supported one another. Now, another week later, she’s just winding down and catching her

breath—and trying to make sense of it all.

On Monday, December 14, I caught up with Gebru via Zoom. She recounted what happened during her time at Google, reflected on what it meant for the field and AI ethics research, and gave parting words of advice to those who want to keep holding tech companies accountable. You can also listen to a special episode of our podcast, *In Machines We Trust*, for highlights from the interview. (Google declined a request for comment on the contents of this interview.)

## **I wanted to first check in with how you're doing.**

I feel like I haven't really had the time to process everything that happened and its repercussions emotionally. I'm just sort of going and going and going. So I feel like I'll probably fall apart at some point when there's a little bit of a lull. But right now I'm just highly concerned about my team and the people supporting me, and the types of risks they're taking, and making sure that they're not retaliated against.

## **There have been so many accounts of what has happened. I wanted to start from a much earlier point in this story. What made you originally choose to work at Google, and what was Google like back then?**

I think [Samy [Bengio, a director at Google AI](#)] and [Jeff [Dean, the SVP of Google Research](#)] were at the Black in AI workshop [at NeurIPS in 2017]. They were asking me what I did, and they said, "Oh yeah, you should come work at Google." I wasn't planning on it. I was doing my postdoc at the time at Microsoft Research [MSR]. I hadn't figured out what I was going to do next. But I knew I wanted to go back to the Bay Area, and they were creating an office in Accra, Ghana. I thought it would be good for me to help with that.

I had a lot of reservations. I was in New York City at MSR, and there were a lot of vocal women there—Danah Boyd, Hannah Wallach, Jen [Wortman Vaughan], Kate Crawford, Mary Gray. There weren't really women of color. The only Black women I know out of all of Microsoft Research are Danielle Bellgrave in the UK and Shawndra Hill in New York. But still, even the men were very supportive. I was very hesitant to go to an environment where I knew Google Research was not well known for its advocacy for women. There were a number of issues that I had heard through my whisper networks. In fact, when I said I was going to go to Google Research, a number of people actually sat me down. So I was just already dreading it, like "Oh, man, okay, what am I going into?"

**I was definitely the first Black woman to be a research scientist at Google. After me, we got two more Black women. That's, like, out of so many research scientists. Hundreds and hundreds. Three out of God knows how many.**

They did not disappoint. It was just constant fighting. I was trying to approach it as talking to people, trying to educate them, trying to get them to see a certain point of view. I kept on thinking that they could do better, you know? With Samy, he has become such a huge advocate. People were complaining that this organization [Google Research] hired just 14% women. Samy, my manager, hired 39% women. It wasn't like he had any incentive to do that whatsoever. He was the only reason I feel like this didn't happen to me before. It's probably because he was protecting us. And by protecting us, he would get in trouble himself. If other leaders are tone-policing you, and you're too loud, you're like a troublemaker—we all know that's what happens to people like me—then if someone defends you, they're obviously going to also be a problem for the other leaders.

So that was my two years at Google. I actually thought that maybe we were making progress until the last incident, because our team grew. It went from almost disintegrating—two months into my time at Google, my co-lead, Meg Mitchell, was going to quit. But then we expanded our team, and we are now, like, 12 people. So I thought that we were inching forward.

There was so much talk about diversity and inclusion, but so much hypocrisy. I'm one of 1.6% Black women at Google. In [Google] Research, it's not 1.6%—it's way lower. I was definitely the first Black woman to be a research scientist at Google. After me, we got two more Black women. That's, like, out of so many research scientists. Hundreds and hundreds. Three out of God knows how many.

So at some point I was just like, you know what? I don't even want to talk about diversity. It's just exhausting. They want to have meetings with you, they don't listen to you, and then they want to have meetings with you again. I've written a million documents about a million diversity-related things—about racial literacy and machine learning [ML], ML fairness initiatives, about retention of women, and the issues. So many documents and so many emails.

So it's just been one thing after another. There's not been a single vacation I took inside Google where I wasn't in the middle of some issue or another. It's just never been peace of mind. Imagine somebody's shooting at you with a gun and you're screaming. And instead of trying to stop the person who's shooting at you with a gun, they're trying to stop you from screaming. That's how it felt. It was just so painful to be in that position over and over and over again.

## **You successfully built one of the most diverse teams in the AI industry. What did it actually take to do that?**

We had to battle all sorts of stuff. I had to be a manager, and then people did not want me to be a manager. I was like, "Okay, I've started a very well-known nonprofit. Why do you have 'concerns' about me being a manager?" Samy didn't say this to me, but he had to deliver this message: "Does she know that she can get fired for things? Does she know that if she becomes a manager, then she's going to have to be a representative of Google?" Then people raised concerns about me seeming unhappy at Google. It's not like, "Oh, there's a toxic culture that's making people like her unhappy. So let's fix that culture." No, that was not the conversation. The conversation was "She seems to be unhappy, so let's not make her a manager."

I was livid at that time. I was so angry. I was asking every other person who became a manager at my level what their experience was. I'm like, "This person became a manager and nobody ever asked them if they knew they were going to be fired for X, Y, and Z. This other person became a manager. Nobody had to talk to them. There was no discussion whatsoever." For me it wasn't like that.

**Imagine somebody's shooting at you with a gun and you're screaming. And instead of trying to stop the person who's shooting at you with a gun, they're trying to stop you from screaming.**

Another thing: we wanted to hire social scientists. A lot of times researchers just hire their friends, and we didn't want to do that. We put out a call. We got 300 applications. We looked through them by hand because we wanted to make sure that recruiters were not filtering out certain groups of people. We had a quick call with 20-something, we had an onsite interview with 10 people, and then we hired two of them.

Why I thought we were making progress was because we were able to get resources to hire these people. So I thought that maybe Jeff was starting to support our team and support the kinds of stuff we were doing. I never imagined—I don't know exactly how this thing happened at all—but I just did not imagine that he would sign off on it [my firing]. I can't imagine him initiating it, but even him signing off on it was just something so surprising to me.

## **During the time that you were building the team, were you able to build relationships with other teams at Google? Were certain parts of the organization becoming more receptive to AI ethics issues?**

Most people on our team are inundated with requests from other teams or other people. And one of our challenges was to not always be in the middle of a fire, because we wanted to have foresight. We wanted to shape what happens in the future and not just react.

The biggest mismatch I see is that there are so many people who respect us, but then there's people at the top, like VPs, who just maybe can't stand us or just don't respect our authority or expertise at all. I've seen that a number of times. But people in Google Cloud, or Cloud AI specifically, some of the senior leadership—they were very supportive. I felt like they really respected our leadership, so they would try to pull us into many things. On the other hand, this latest fiasco was not from them. I have my suspicions of which VPs it was coming from, and they certainly did not respect our expertise or leadership.

## **Could you talk a little bit more about that?**

Well, even if you just see [the email from Jeff](#)—I'm assuming he didn't write this email; I'm assuming somebody wrote it and he sent it—it talks about how our research [paper on large language models] had gaps, it was missing some literature. [The email doesn't] sound like they're talking to people who are experts in their area. This is not peer review. This is not reviewer #2 telling you, "Hey, there's this missing citation." This is a group of people, who we don't know, who are high up because they've been at Google for a long time, who, for some unknown reason and process that I've never seen ever, were given the power to shut down this research.

We had 128 citations [on that paper], and we sent our paper to many of these people that we cited. We were so thorough. I said, okay, I want to bucket the people that we're going to ask feedback from in four buckets. One is the people who have developed large language models themselves, just to get their perspective. One is people who work in the area of understanding and mitigating the bias in these models. One is people who might disagree with our view. One is people who use these large language models for various things. And we have a whole document with all of this feedback that we were supposed to go through to address, and which I want to do still before we release this work.

But the way they [Google leadership] were talking to us, it wasn't like they were talking to world-renowned experts and linguists. Like Emily Bender [a professor at the University of Washington and a coauthor of the paper] is not some random person who just put her name on a random paper out there. I felt like the whole thing was so disrespectful.

## **Prior to this particular paper, were there earlier instances in which you ever felt that Google was restricting or limiting your team's ability to conduct research?**

I felt like there were prior instances where they watered down the research results a lot. People had conversations with PR and policy or whatever, and they would take issue with certain wording or take issue with certain specifics. That's what I thought they might try to do with this paper, too. So I wrote a document, and I kept asking them, "What exactly is your feedback? Is your feedback to add a section? To remove? What does this mean? What are you asking us?"

This was literally my email on Friday after Thanksgiving [*November 27, five days before Gebru's dismissal*] because on Thanksgiving day I had spent my day writing this document instead of having a good time with my family. The next day, on Friday, which is when I was supposed to retract this paper, I wrote: "Okay, I have written this six-page document addressing at a high level and low level whatever feedback I can gather. And I hope that there is at the very least an

openness for further conversation rather than just further orders.” I wrote that email. Like that. How does Megan [Kacholia, the VP of engineering at Google Research] respond to this email? Monday, she responds to it and says, “Can you please confirm that you have either retracted the paper or taken the names of the authors out of this paper. Thank you. And can you please confirm after you’ve done this. Send me an email and confirm.” As if I have no agency.

**That’s not what they do to people who’ve engaged in gross misconduct. They hand them \$80 million, and they give them a nice little exit. They don’t do what they did to me.**

Then in that document, I wrote that this has been extremely disrespectful to the Ethical AI team, and there needs to be a conversation, not just with Jeff and our team, and Megan and our team, but the whole of Research about respect for researchers and how to have these kinds of discussions. Nope. No engagement with that whatsoever.

I cried, by the way. When I had that first meeting, which was Thursday before Thanksgiving, a day before I was going to go on vacation—when Megan told us that you have to retract this paper, I started crying. I was so upset because I said, I’m so tired of constant fighting here. I thought that if I just ignored all of this DEI [diversity, equity, and inclusion] hypocrisy and other stuff, and I just focused on my work, then at least I could get my work done. And now you’re coming for my work. So I literally started crying.

**What do you think it was about this particular paper that touched off these events? Do you think that it was actually about this paper, or were there other factors at play?**

I don’t know. Samy was horrified by the whole thing. He was like, this is not how we treat researchers. I mean, this is not how you treat people in general. People are talking about how, if this happens to somebody this accomplished—it makes me imagine what they do to people, especially people in vulnerable groups.

They probably thought I’d be super quiet about it. I don’t know. I don’t think the end result was just about this paper. Maybe they were surprised that I pushed back in any way whatsoever. I’m still trying to figure out what happened.

**Did you ever suspect, based on the previous events and tensions, that it would end in this way? And did you expect the community’s response?**

I thought that they might make me miserable enough to leave, or something like that. I thought that they would be smarter than doing it in this exact way, because it’s a confluence of so many issues that they’re dealing with: research censorship, ethical AI, labor rights, DEI—all the things that they’ve come under fire for before. So I didn’t expect it to be in that way—like, cut off my corporate account completely. That’s so ruthless. That’s not what they do to people who’ve engaged in gross misconduct. They [hand them \\$80 million](#), and they give them a nice little exit, or maybe they passive-aggressively don’t promote them, or whatever. They don’t do to the people who are actually creating a hostile workplace environment what they did to me.

I found out from my direct reports, you know? Which is so, so sad. They were just so traumatized. I think my team stayed up till like 4 or 5 a.m. together, trying to make sense of what happened. And going around Samy—it was just all so terrible and ruthless.

I thought that if I just...focused on my work, then at least I could get my work done. And now you’re coming for my work. So I literally started crying.



I expected some amount of support, but I definitely did not expect the amount of outpouring that there is. It's been incredible to see. I've never, ever experienced something like this. I mean, random relatives are texting me, "I saw this on the news." That's definitely not something I expected. But people are taking so many risks right now. And that worries me, because I really want to make sure that they're safe.

**You've mentioned that this is not just about you; it's not just about Google. It's a confluence of so many different issues. What does this particular experience say about tech companies' influence on AI in general, and their capacity to actually do meaningful work in AI ethics?**

You know, there were a number of people [comparing Big Tech and Big Tobacco](#), and how they were censoring research even though they knew the issues for a while. I push back on the academia-versus-tech dichotomy, because they both have the same sort of very racist and sexist paradigm. The paradigm that you learn and take to Google or wherever starts in academia. And people move. They go to industry and then they go back to academia, or vice versa. They're all friends; they are all going to the same conferences.

I don't think the lesson is that there should be no AI ethics research in tech companies, but I think the lesson is that a) there needs to be a lot more independent research. We need to have more choices than just DARPA [the Defense Advanced Research Projects Agency] versus corporations. And b) there needs to be oversight of tech companies, obviously. At this point I just don't understand how we can continue to think that they're gonna self-regulate on DEI or ethics or whatever it is. They haven't been doing the right thing, and they're not going to do the right thing.

I think academic institutions and conferences need to rethink their relationships with big corporations and the amount of money they're taking from them. Some people were even wondering, for instance, if some of these conferences should have a "no censorship" code of conduct or something like that. So I think that there is a lot that these conferences and academic institutions can do. There's too much of an imbalance of power right now.

**What role do you think ethics researchers can play if they are at companies? Specifically, if your former team stays at Google, what kind of path do you see for them in terms of their ability to produce impactful and meaningful work?**

I think there needs to be some sort of protection for people like that, or researchers like that. Right now, it's obviously very difficult to imagine how anybody can do any real research within these corporations. But if you had labor protection, if you have whistleblower protection, if you have some more oversight, it might be easier for people to be protected while they're doing this kind of work. It's very dangerous if you have these kinds of researchers doing what my co-lead was calling "fig leaf"—cover-up—work. Like, we're not changing anything, we're just putting a fig leaf on the art. If you're in an environment where the people who have power are not invested in changing anything for real, because they have no incentive whatsoever, obviously having these kinds of researchers embedded there is not going to help at all. But I think if we can create accountability and oversight mechanisms, protection mechanisms, I hope that we can allow researchers like this to continue to exist in corporations. But a lot needs to change for that to happen.

**In addition to having conferences or regulation change the incentive structures, are there other mechanisms or things that you could see being part of this external oversight framework? Do you see any role that the public could play in holding tech companies accountable?**

I have to think more about this. I think that the public needs to be more educated in the role that tech companies, and also AI, are playing in our daily lives. A lot of people have been doing a really good job of that. People have been writing books—[Weapons of Math Destruction](#) was a very good book for me to read. There's [Coded Bias](#), the documentary. I'm also happy that I'm starting to see universities offering some of these classes in computer science. I think they should be required. Nicki Washington [has a class at Duke](#) that I wish I could've taken. Ruha [Benjamin] [has a class [at Princeton](#)]. I think our understanding of how science is developed and how engineering is developed—it just needs to change. The public needs to understand the political nature of this work. Higher education systems, in my opinion, are very behind in this way. They need to do a lot of work.

In terms of the public, I think that if I were to go back to three, four years ago, there was a lot less awareness. Now entities like the ACLU [American Civil Liberties Union] are heavily involved; you have [the Algorithmic Justice League](#), [Data for Black Lives](#), doing a lot of work here. So I think that the public can learn more. But I think this is more on actually the politicians, because we do have a mechanism to involve the public in various things. But the politicians are not giving the public a way to get involved. For instance, [data analytics company] [Palantir was being used in New Orleans](#) for predictive policing and nobody knew. There was no vote. There was no bill. There was no discussion about it. So our government needs to actively give the public a chance to weigh in on these kinds of issues.

**What does it mean to actually do meaningful AI ethics work?**

I think maybe it means that you can question the fundamental premises and fundamental issues of AI. You don't have to just cover things up, or you don't have to just be reactionary. You can have foresight about the future, and you can get in early while products are being thought about and developed, and not just do things after the fact. That's the number one thing I think about. And that the people most impacted by the technology should have a very big say, starting from the very beginning.

**You've mentioned before how ethics at tech companies often comes from the top down, from people who are not necessarily the ones being impacted by the technology. What is the ideal balance of actually having marginalized voices driving the conversation, while also having top leadership buy-in?**

If people at the margins don't have power, then there's no incentive for anyone to listen to them or to have them shape the discussion. So I think that people at the top need to foster an environment where people at the margins are the ones shaping the discussions about these things, about AI ethics, about diversity and inclusion.

**We are taught when we're learning science "objectively," what's called "the view from nowhere," as if you're not supposed to bring your point of view to it. I think people should do the exact opposite.**

That's not what's happening right now. You go to Google and you see all of the high-level people doing AI ethics—they're absolutely not even people from underrepresented groups at all. That was the conversation we were having when I was there, and many of us were super frustrated by it. We had these principles called "Nothing about us without us," and the very first thing we came

up with was psychological safety. Can you imagine? That means don't punish me for speaking up, which is exactly what they did. So you need to seek the input, but then again, don't do it in a predatory way. Before you even think about the products that you're building or the research that you're doing, you need to start imagining: how can you work with people at the margins to shape this technology?

### **Do you in a way see yourself as a martyr?**

I wasn't trying to be one, and I don't think I'm a martyr right now. I'm hoping that I'm more of an agent for change. But yeah, Google kind of made me one, more so than I would have been if they didn't do this to me.

### **What's next for you?**

Oh, man. I have not had any time to think about that. Right now I'm literally trying to make sure that my team is safe, make sure that the narrative around what's going on with me is accurate, and take a little bit of a breather when I can. Hopefully drink some wine, have some food, eat some chocolate, hang out with family. And maybe I'll figure it out after. But my biggest priority is to not be in the kind of environment I was in. That's my biggest priority.

### **What advice do you have for other people who want to build off the work that you've done?**

My advice is to read up on some of the works that I mentioned, especially by [Ruha [Benjamin](#)], [Meredith [Broussard](#)], [Safiya [Noble](#)], [Simone Browne](#), and a number of other people. And bring your lived experience into it. Don't try to take yourself out of this work. It has a lot to do with you. So your angle and your strength will depend on what your lived experience is. I gave a talk called "[The Hierarchy of Knowledge and Machine Learning](#)," and it was sort of about that: How we are taught when we're learning science "objectively," what's called "the view from nowhere," as if you're not supposed to bring your point of view to it. I think people should do the exact opposite.