EMERGING TECH BREW

AI

AI-ethics pioneer Margaret Mitchell on her five-year plan at open-source AI startup Hugging Face

Mitchell, who founded Google's AI ethics group, joined the company in August

OCTOBER 4, 2021 · 8 MIN READ

HAYDEN FIELD
EMERGING TECH REPORTER

Follow

COPY

Margaret Mitchell

Margaret Mitchell has spent her career founding bootstrap-style AI projects inside large tech companies. She helped create Microsoft Research's "Cognition" group, which concentrated on AI advancement, before moving to Google and founding its Ethical AI team and cofounding its ML Fairness group.

Now, she's left Big Tech behind for full-time startup life—leading data governance efforts at Hugging Face, a 60-person AI company founded in 2016 and based in NYC.

It's a big change for Mitchell, following even bigger ones. Over a three-month span beginning in December 2020, Google fired both Mitchell and her Ethical AI team co-lead, Timnit Gebru, after disagreements over their research paper on the dangers of large language models. (Google disputes this version of events.) Though the powerful language algorithms increasingly underpin popular and useful services like Google Search and AutoComplete, their large-scale pattern recognition—trained on vast swaths of the internet—can replicate harmful human biases and multiply harms.

The most powerful language models are also concentrated in the hands of a few powerful companies. Typically, their training and development is limited to the Big Tech sphere, or to FAMGA-funded research groups, both in academia and outside of it (e.g, OpenAI).

Hugging Face wants to bring these powerful tools to more people. Its mission: Help companies build, train, and deploy AI models—specifically natural language processing (NLP) systems—via its open-source tools, like Transformers and Datasets. It also offers pretrained models available for download and customization.

So what does it mean to play a part in "democratizing" these powerful NLP tools? We chatted with Mitchell about the split from Google, her plans for her new role, and her near-future predictions for responsible AI.

This interview has been edited for length and clarity.

After your time at Google ended, did you take some time to think about your must-haves—and must-not-haves—for any future roles?

No, I actually didn't do that at all. I guess that's a strange answer, but in terms of what I was going to do next, I felt pretty frustrated that Google had put up an obstacle for me—another barrier. I've spent my whole career getting through barriers and obstacles, and here was another. So I was immediately looking for a place where I could just continue doing what I was doing, essentially, and not be slowed down too much.

So obviously that means the Big Tech companies, especially now that there's more interest in doing this kind of responsible, ethical AI work. That route at first really appealed to me because I have a ton of friends and people I respect at these companies, but then as time went on, I needed to find a way to quickly make money. It's kind of like having the rug pulled out from you, with no salary or severance or anything like that, so I ended up cobbling together work as a consultant. But that gave me a lot of insight into the possibilities of what I could do and the paths I could take.

Suddenly I was thinking more seriously about jobs in the public sector and regulation space, as well as with smaller companies like the startups I was consulting and contracting with. So it was only once I was starting to talk to the bigger companies while trying to make money in other ways that I realized I could start prioritizing more of what I'm looking for and make a much more informed decision.

You mentioned you've had the same goal in this space for a long time, that you didn't want to be slowed down. How would you put that objective into words?

I started on my path toward working on what's now called AI in 2004, specifically with an interest in aligning AI closer to human behavior. Over time, that's evolved to become less about mimicking humans and more about accounting for human behavior and working with humans in assistive and augmentative ways. So it's still about aligning to humans, but through a perspective of what's best for humans, as opposed to what's—I don't want to say what's best for AI—but I switched from a more tech-focused approach to a person-focused approach. So that's been an evolution of my work over time.

By the time I was at Microsoft Research, I was working really hard to go from initial ideation to a product launch in ways that could fundamentally change both the development cycle and the kind of technology that's created in ways that are more aligned with ethical values.

So once I was at Google, it was pretty clear that my role was to work toward shaping AI toward these more human-informed values and goals that have to do with what can create the most benefit for people. So continuing on that line with leaving Google, I was increasingly interested in companies that had ethical values baked into their core, like part of the initial construction of the company, because I had felt that I had gotten a pretty good expertise at *retrofitting* ethical processes and retrofitting inclusion, essentially, in companies that hadn't been built on that. It's very challenging, obviously, and I think that what happened with Google showed a bit of how intense that can be in a way that people didn't really realize before.

Did you have any concerns about the company itself and what it's working on—making large language models and the tech that powers them more widely accessible? Although concentration of power is fraught in its own right, were you concerned about this kind of work accelerating harms?

There's actually a distinction between Hugging Face, the company, and the BigScience project, which is a collaborative effort with different universities and organizations working together on more ethics-informed language model work. And there were tons of people involved there who I really respect, so I came into some of the meetings and saw that they were thinking a lot about the values at play. And I was invited to be in charge of data governance—and I think the handling of data is one of the hardest problems right now in tech and AI, like how do you consent if you're scraping data.

It seemed to me, given this increasing interest in large language models, this would be a really great opportunity for me to help shape what that interest could look like—in a way that was less the traditional "Just make the numbers go higher" and more like, "What is the social context of the use of these models?" So that was exciting to me.

I mean, I'm still a tech nerd at heart. I love to code; I love building models. So I worry sometimes that it gets lost in the shuffle—that my interest would be in stopping technology. While I definitely have interest in slowing technology to be more intentional and well-informed, and in regulating technology, I still love working on it. It's still something that I see a lot of like positive paths forward on—I'm very tech-positive. So this seemed like a really natural thing to do.

On an individual level, what's your biggest goal for one year from now and for five years from now? What do you personally want to have accomplished in the responsible-AI space?

That's a great question. One year from now, I want it to be a norm that people provide documentation for datasets and for models. I've been working on that for a few years, so maybe four years ago, that was my five-year plan—now we're at one year left. And I think the progress has gone pretty well. We see companies across the board starting to use things like model cards—so Facebook, Salesforce, Nvidia, OpenAI, Allen AI, various papers—we're really seeing a rise in responsible, value-aligned documentation that I like to think that I've helped with.

I'd hope that in a year, that's just what people do—that it will be strange not to do that. And I think we're getting closer and closer. Maybe a year is a reasonable goal, especially if through Hugging Face, I can help provide the tools to make these things move more easily within development workflows.

My five-year plan, though—one thing that I'm really sort of fascinated by and hoping to do something meaningful about within the next five years focuses on people's rights to contestation of how they're represented in models and in data. Currently, no one can contest if they have some text that's been picked up in a training data set and then used to train a model, people can't say no to that. So figuring out how to allow contestation brings with it a whole bunch of research questions: How do you remove training data from a model? How do you even find the data that someone would contest?

That also means we'll also start to have a lot more possibilities around data rights and who actually owns the data. So the current paradigm is that whoever finds it, owns it—not whoever creates it, owns it. So I think within five years, we'll see that paradigm fundamentally shifting—or at least I hope so.

COPY

You might also like...

WORK LIFE

Slack: its pros, cons, and moral reckoning

MALIAH WEST / 10.14.2021