



# 5 Questions for Joy Buolamwini

Why AI should move slow and fix things

**J**oy Buolamwini, an AI researcher, first made waves with a TED talk in which Buolamwini, who is Black, showed that facial detection systems didn't detect her face unless she put on a white mask. She's also the founder of the Algorithmic Justice League. Buolamwini's new book, *Unmasking AI*, reminds engineers that default settings are not neutral, that convenient datasets may be rife with ethical and legal problems, and that benchmarks aren't always assessing the right things.

**You've worked hard to find data-collection methods that feel ethical to you. Can you imagine a world in which every AI researcher is so scrupulous?**

**Joy Buolamwini:** When I was earning my academic degrees and learning to code, I did not have examples of ethical data collection. Basically if the data was available online, it was there for the taking. It can be difficult to imagine another way of doing things if you never see an alternative pathway. I do believe there is a world where more AI researchers exercise more caution with data-collection activities, because of the engineers who reach out to the Algorithmic Justice League looking for a better way.

*Joy Buolamwini earned her Ph.D. from MIT Media Lab and founded the Algorithmic Justice League to fight for people who have experienced algorithmic discrimination. Her work was the subject of the documentary Coded Bias, and her new book is Unmasking AI: My Mission to Preserve What Is Human in a World of Machines (Random House, 2023).*

*For a longer version of this interview, see [spectrum.ieee.org/joy-buolamwini](https://spectrum.ieee.org/joy-buolamwini)*

**What can concerned engineers at big tech companies do about algorithmic bias and other AI ethics issues?**

**Buolamwini:** I cannot stress enough the importance of documentation. In conducting algorithmic audits and approaching well-known tech companies with the results, one issue that came up time and time again was the lack of internal awareness about the limitations of the AI systems that were being deployed. [Documentation] that provides an opportunity to see the data used to train AI models and the performance of those AI models in various contexts is an important starting point. Then the question becomes: Is the company willing to release a system with the limitations documented or are they willing to go back and make improvements?

**How can AI researchers convince their companies to take AI ethics seriously?**

**Buolamwini:** It can be helpful to not view AI ethics separately from developing robust and resilient AI systems. If your tool doesn't work as well on women or people of color, you are at a disadvantage compared to companies who create tools that work well for a variety of demographics. If your AI tools generate harmful stereotypes or hate speech, you are at risk for reputational damage that can impede a company's ability to recruit necessary talent, secure future customers, or gain follow-on investment.

**You write that "the choice to stop is a viable and necessary option," even when tools have already been adopted. Would you like to see a course reversal on today's popular generative AI tools?**

**Buolamwini:** Facebook (now Meta) deleted a billion faceprints around the time of a [US] \$650 million settlement after they faced allegations of collecting face data to train AI models without the expressed consent of users. [Such] actions show that when there is resistance and scrutiny, there can be change.

**What's your opinion on superintelligent AI posing an existential risk to our species?**

**Buolamwini:** The "x-risk" I am concerned about is the x-risk of being excoded—that is, being harmed by AI systems. I am concerned with lethal autonomous weapons and giving AI systems the ability to make kill decisions. I am concerned with the ways in which AI systems can be used to kill people slowly through lack of access to adequate health care, housing, and economic opportunity. ■