

REPROGRAMMING THE PATRIARCHY

Combatting Gender Bias in Machine Learning

By Sara Kassir

Sara is a first-year Master in Public Policy student at the Harvard Kennedy School who is interested in data-driven approaches to navigating public-private partnerships. She previously worked for Deloitte Consulting in Washington, D.C., supporting federal agencies seeking to develop performance measurement capabilities.

The media narrative on artificial intelligence (AI) is not far removed from the plot of a science-fiction movie: robots will one day become smarter than the humans who created them, leading to cataclysmic events we cannot control. While this scenario depicts the risks that AI may pose in the future, a more immediate threat receives far less attention: that robots will become just like us, leading to a world of human prejudices we cannot control.

Ten years ago, readers would have been baffled to see headlines like “Robots are sexist and racist” and “Machines must start to learn without prejudice” in the mainstream news.^{1,2} But today, these headlines are real, and much of our technology has developed a mind of its own as a result of machine learning.

A subset of AI, machine learning refers to computer programs that can improve their performance at a given task over time. Renowned computer scientist Arthur Samuels first demonstrated the technology in 1959 with a game of checkers, in which a computer studied patterns of effective moves and incorporated them into future rounds.³

Over 50 years later, the same basic principle

is used in common programs like image recognition. Because there is an infinite amount of variation in pictures, humans could never write a static algorithm that would allow a computer to correctly identify an image’s subject. However, machine learning allows the computer to draw from an underlying “training” dataset, which it uses to predict the identity of an object or person. Over time, and with more data, the computer “learns” to execute this task with greater accuracy—and without any human intervention.⁴

How can a task like automated image recognition take a biased turn? In 2010, the newly-released Nikon Coolpix S630 digital camera asked Asian users, “Did you blink?” after taking a photo.⁵ Five years later, an engineer at Google reported that the company’s image classifier was confusing his black friends for gorillas.⁶ The problem driving these errors is surprisingly simple: machines can only use the data we give them, and the data we have is often as biased as the society in which we live. According to one study by the National Institute of Standards and Technology, facial recognition algorithms developed in France, Germany, and the US are more likely to

accurately identify Caucasian subjects, while those developed in China, Japan, and South Korea perform better with East Asian faces.⁷

Technology is often praised as a way to address the worst parts of human nature, but mounting evidence suggests that it may be doing the very opposite, institutionalizing stereotypes more than we realize. This is particularly true when it comes to gender.

Some of the gender biases that have emerged in machine learning labs are predictable. For example, speech recognition software must have access to a sampling of recorded sentences to “learn” to identify particular words. If the sampling does not include female voices or accented speakers, the program will not function properly for those populations.

Other oversights are less easily remedied. Software engineers may decide to populate their voice recognition training dataset with a sampling of Hollywood films, unaware that men receive nearly twice as much screen time as women in American cinema.⁸ As more industries have adopted machine learning, these biases have appeared in settings where their consequences are much more severe.

In medicine, machine learning tools use training data from clinical trials to interpret symptoms and make diagnoses. However, symptoms often present differently in men and women, and alarmingly, clinical trials overwhelmingly oversample men. Although cardiovascular disease is the number-one killer of women in the United States, the Texas Heart Institute reports that the patient population for a typical clinical trial is 85 percent male.⁹ Algorithms developed from this unequal data will result in biases that prevent physicians from effectively treating female patients.¹⁰

In another concerning case, an experiment at Carnegie Mellon revealed how machine learning is reinforcing the glass ceiling in female employment. Researchers created

TECHNOLOGY IS OFTEN PRAISED AS A WAY TO ADDRESS THE WORST PARTS OF HUMAN NATURE, BUT MOUNTING EVIDENCE SUGGESTS THAT IT MAY BE DOING THE VERY OPPOSITE, INSTITUTIONALIZING STEREOTYPES MORE THAN WE REALIZE.

two groups of fake user accounts that were identical on every dimension except gender. The study found that Google showed the female-designated users significantly fewer ads for high-paying executive jobs than the male-designated users. In the words of the study’s authors, “We cannot claim that Google has violated its policies...we consider it more likely that Google has lost control over its massive, automated advertising system.”¹¹

It does not require much imagination to think of other instances in which gender imbalances could negatively affect machine learning tools. As technology rapidly advances, society cannot just wait and see if algorithms continue to disadvantage large segments of the population. Instead, programmers need to consider the current cultural and societal context and include a greater diversity of perspectives before developing machines that execute human functions.

As with most scientific fields in the United States, machine learning is dominated by men, who comprise nearly 87 percent of the specialty’s engineers.¹² The fact that so few women study and work in science, technology, engineering, and math (STEM) has been a problem that policymakers have been seeking to solve for decades. Despite concerted efforts by both the public and private sectors to increase women’s participation in these fields, the percentage of computer science majors who were women actually dropped from 37 percent to 18 percent between 1995 and 2014, indicating a need for varied approaches.¹³

Earlier research explaining why women shy away from STEM has focused on a variety of social pressures, from being teased in elementary school to facing an aggressively masculine culture in labs.¹⁴ While these arguments highlight real problems in male-dominated workplaces, they also seem to ignore the reality that women have successfully broken into other professions with notorious gender imbalances. For example, only 10 percent of US medical and law degree holders in 1970 were women compared to 50 percent today.¹⁵

Recent studies have focused less on the societal disincentives for women in STEM fields, and more on what motivates women and girls to choose a given career path. One market research firm, polling a group of teenage girls interested in engineering, found that 74 percent were only drawn to the field after they learned about the economic benefits and impact they could have on the world.¹⁶ Lina Nilsson, who holds a biomedical engineering PhD and teaches at UC Berkeley, investigated enrollment data at a variety of American universities and found further evidence to support this trend. At the MIT D-Lab, which “develops and advances collaborative approaches and practical solutions to global poverty,” 74 percent of students are women, making it one of the only engineering initiatives in the country with a considerable female majority.¹⁷

To put it in the words of C. Dian Matt, executive director of the non-profit Women in Engineering ProActive Network, “Women are drawn to fields where the social relevance is high.”¹⁸ Around the time Sebastian Thrun left his position as vice president at Google to found the online educational organization Udacity, he said, “Nobody phrases it this way, but I think that artificial intelligence is almost a humanities discipline. It’s really an attempt to understand human intelligence and human cognition.”¹⁹ In articulating the

very real character of AI as one based in creativity, empathy, and emotional intelligence, educators and employers can better position women and girls to develop an interest in the field.

Importantly, women and girls would not be the only benefactors of a shift in society’s approach to machine learning. Technology will almost always outpace government’s capacity to regulate it, so messages that highlight the need for responsible decision-making in AI may be the only way to promote accountability. Until we figure out a way to hold elections for robots and referendums on algorithms, ensuring that women have a seat at the lab bench may be our best option for preserving equality.

NOTES

- 1 Henry Bodkin, "AI Robots Are Sexist and Racist, Experts Warn," *The Telegraph*, 24 August 2017, accessed 20 February 2018, <https://www.telegraph.co.uk/news/2017/08/24/ai-robots-sexist-racist-experts-warn/>.
- 2 Helen Lewis, "In 2018, Machines Must Start to Learn without Prejudice," *Wired*, 19 December 2017, accessed 20 February 2018, <http://www.wired.co.uk/article/technology-prejudice-artificial-intelligence-helen-lewis>.
- 3 Bernard Marr, "A Short History of Machine Learning—Every Manager Should Read," *Forbes*, 19 February 2016, accessed 20 February 2018, <https://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning-every-manager-should-read/#42a622e415e7>.
- 4 Bernard Marr, "What Is The Difference Between Artificial Intelligence And Machine Learning?" *Forbes*, 15 September 2017, accessed 20 February 2018, <https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/2/#457781d8483d>.
- 5 Adam Rose, "Are Face-Detection Cameras Racist?" *Time*, 22 January 2010, accessed 20 February 2018, <http://content.time.com/time/business/article/0,8599,1954643,00.html>.
- 6 Tom Simonite, "When It Comes to Gorillas, Google Photos Remains Blind," *Wired*, 11 January 2018, accessed 20 February 2018, <https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/>.
- 7 Clare Garvie and Jonathan Franke, "Facial-Recognition Software Might Have a Racial Bias Problem," *The Atlantic*, 7 April 2016, accessed 20 February 2018, <https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/>.
- 8 Steven Overly, "Women in Hollywood Get Less Screen Time than Men. This Technology Could Help Fix That," *The Washington Post*, 20 September 2016, accessed 20 February 2018, https://www.washingtonpost.com/news/innovations/wp/2016/09/20/women-in-hollywood-get-less-screen-time-than-men-this-technology-could-help-fix-that/?utm_term=.ca4cbb9161c6.
- 9 Anne Hamilton Dougherty, "Gender Balance in Cardiovascular Research," *Texas Heart Institute* 38, no. 2 (2011): 148-50.
- 10 Khader Shameer et al. "Machine Learning in Cardiovascular Medicine: Are We There Yet?" *Heart*, 19 January 2018.
- 11 Kevin Montgomery, "Google's Ad System Has Become Too Big to Control," *Wired*, 8 July 2015, accessed 20 February 2018, <https://www.wired.com/2015/07/googles-ad-system-become-big-control/>.
- 12 Cale Guthrie Weissman, "The Women Changing The Face Of AI," *Fast Company*, 18 August 2016, accessed 20 February 2018, <https://www.fastcompany.com/3062932/ai-is-a-male-dominated-field-but-an-important-group-of-women-is-changing-th>.
- 13 Gaby Galvin, "Study: Middle School Is Key to Girls' Coding Interest," *U.S. News*, 20 October 2016, accessed 20 February 2018, <https://www.usnews.com/news/data-mine/articles/2016-10-20/study-computer-science-gender-gap-widens-despite-increase-in-jobs>.
- 14 Jane Margolis and Allan Fisher, *Unlocking the Clubhouse: Women in Computing* (Cambridge, MA: MIT Press, 2003).
- 15 Philip Cohen, "More Women Are Doctors and Lawyers Than Ever—but Progress Is Stalling," *The Atlantic*, 11 December 2012, accessed 20 February 2018, <https://www.theatlantic.com/saxes/archive/2012/12/more-women-are-doctors-and-lawyers-than-ever-but-progress-is-stalling/266115/>.
- 16 Mark Fidelman, "Here's the Real Reason There Are Not More Women in Technology," *Forbes*, 5 June 2012, accessed 20 February 2018, <https://www.forbes.com/sites/markfidelman/2012/06/05/heres-the-real-reason-there-are-not-more-women-in-technology/>.
- 17 Lina Nilsson, "How to Attract Female Engineers," *The New York Times*, 27 April 2015, accessed 20 February 2018, <https://www.nytimes.com/2015/04/27/opinion/how-to-attract-female-engineers.html>.
- 18 "Engineering for women—making a tangible difference to people's lives: Part 1," *The Telegraph*, 29 November 2013, accessed 20 February 2018, <https://jobs.telegraph.co.uk/article/engineering-for-women-making-a-tangible-difference-to-people-s-lives-part-1/>.
- 19 Bernard Marr, "28 Best Quotes About Artificial Intelligence," *Forbes*, 25 July 2017, accessed 20 February 2018, <https://www.forbes.com/sites/bernardmarr/2017/07/25/28-best-quotes-about-artificial-intelligence/#573d29f4a6fc>.