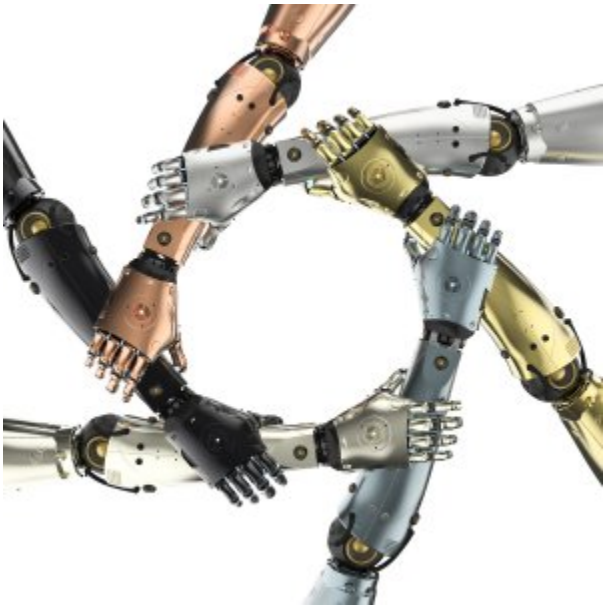
Using Artificial Intelligence to Promote Diversity

PAUL R. DAUGHERTY, H. JAMES WILSON, AND RUMMAN CHOWDHURY

AI can help us overcome biases instead of perpetuating them, with guidance from the humans who design, train, and refine its systems.



Artificial intelligence has had some justifiably bad press recently. Some of the worst stories have been about systems that exhibit racial or gender bias in facial recognition applications or in evaluating people for jobs, loans, or other considerations. [1] One program was routinely recommending longer prison sentences for blacks than for whites on the basis of the flawed use of recidivism data. [2]

But what if instead of perpetuating harmful biases, AI helped us overcome them and make fairer decisions? That could eventually result in a more diverse and inclusive world. What if, for instance, intelligent machines could help organizations recognize all worthy job candidates by avoiding the usual hidden prejudices that derail applicants who don't look or sound like those in power or who don't have the "right" institutions listed on their résumés? What if software programs were able to account for the inequities that have limited the access of minorities to mortgages and other loans? In other words, what if our systems were taught to ignore data about race, gender, sexual orientation, and other characteristics that aren't relevant to the decisions at hand?

AI can do all of this — with guidance from the human experts who create, train, and refine its systems. Specifically, the people working with the technology must do a much better job of building inclusion and diversity into AI design by using the right data to train AI systems to be inclusive and thinking about gender roles and diversity when developing bots and other applications that engage with the public.

Design for Inclusion

Software development remains the province of males — only about one-quarter of computer scientists in the United States are women [3] — and minority racial groups, including blacks and Hispanics, are

Copyright © Massachusetts Institute of Technology, 2018. All rights reserved. | Reprint #60216 | https://mitsmr.com/2DQz2XT

underrepresented in tech work, too. [4] Groups like Girls Who Code and AI4ALL have been founded to help close those gaps. Girls Who Code has reached almost 90,000 girls from various backgrounds in all 50 states, [5] and AI4ALL specifically targets girls in minority communities. Among other activities, AI4ALL sponsors a summer program with visits to the AI departments of universities such as Stanford and Carnegie Mellon so that participants might develop relationships with researchers who could serve as mentors and role models. And fortunately, the AI field has a number of prominent women — including Fei-Fei Li (Stanford), Vivienne Ming (Singularity University), Rana el Kaliouby (Affectiva), and Cynthia Breazeal (MIT) — who could fill such a need.

These relationships don't just open up development opportunities for the mentees — they're also likely to turn the mentors into diversity and inclusion champions, an experience that may affect how they approach algorithm design. Research by sociologists Frank Dobbin of Harvard University and Alexandra Kalev of Tel Aviv University supports this idea: They've found that working with mentees from minority groups actually moves the needle on bias for the managers and professionals doing the mentoring, in a way that forced training does not. [6]

Other organizations have pursued shorter-term solutions for AI-design teams. LivePerson, a company that develops online messaging, marketing, and analytics products, places its customer service staff (a profession that is 65% female in the United States) alongside its coders (usually male) during the development process to achieve a better balance of perspectives. [7] Microsoft has created a framework for assembling "inclusive" design teams, which can be more effective for considering the needs and sensitivities of myriad types of customers, including those with physical disabilities. [8] The Diverse Voices project at the University of Washington has a similar goal of developing technology on the basis of the input from multiple stakeholders to better represent the needs of nonmainstream populations.

Some AI-powered tools are designed to mitigate biases in hiring. Intelligent text editors like Textio can rewrite job descriptions to appeal to candidates from groups that aren't well-represented. Using Textio, software company Atlassian was able to increase the percentage of females among its new recruits from about 10% to 57%. [9] Companies can also use AI technology to help identify biases in their past hiring decisions. Deep neural networks — clusters of algorithms that emulate the human ability to spot patterns in data — can be especially effective in uncovering evidence of hidden preferences. Using this technique, an AI-based service such as Mya can help companies analyze their hiring records and see if they have favored candidates with, for example, light skin.

Train Systems With Better Data

Building AI systems that battle bias is not only a matter of having more diverse and diversity-minded design teams. It also involves training the programs to behave inclusively. Many of the data sets used to train AI systems contain historical artifacts of biases — for example the word *woman* is more associated with *nurse* than with *doctor* — and if those associations aren't identified and removed, they will be perpetuated and reinforced. [10]

While AI programs learn by finding patterns in data, they need guidance from humans to ensure that the software

Copyright © Massachusetts Institute of Technology, 2018. All rights reserved. | Reprint #60216 | https://mitsmr.com/2DQz2XT

doesn't jump to the wrong conclusions. This provides an important opportunity for promoting diversity and inclusion. Microsoft, for example, has set up the Fairness, Accountability, Transparency, and Ethics in AI team, which is responsible for uncovering any biases that have crept into the data used by the company's AI systems.

Sometimes AI systems need to be refined through more inclusive representation in images. Take, for instance, the fact that commercial facial recognition applications struggle with accuracy when dealing with minorities: The error rate for identifying dark-skinned women is 35%, compared with 0.8% for light-skinned men. The problem stems from relying on freely available data sets (which are rife with photos of white faces) for training the systems. It could be corrected by curating a new training data set with better representation of minorities or by applying heavier weights to the underrepresented data points. [11] Another approach — proposed by Microsoft researcher Adam Kalai and his colleagues — is to use different algorithms to analyze different groups. For example, the algorithm for determining which female candidates would be the best salespeople might be different from the algorithm used for assessing males — sort of a digital affirmative action tactic. [12] In that scenario, playing a team sport in college might be a higher predictor of success for women than for men going after a particular sales role at a particular company.

Give Bots a Variety of Voices

Organizations and their AI system developers must also think about how their applications are engaging with customers. To compete in diverse consumer markets, a company needs products and services that can speak to people in ways they prefer.

In tech circles, there has been considerable discussion over why, for instance, the voices that answer calls in help centers or that are programmed into personal assistants like Amazon's Alexa are female. Studies show that both men and women tend to have a preference for a female assistant's voice, which they perceive as warm and nurturing. This preference can change depending on the subject matter: Male voices are generally preferred for information about computers, while female voices are preferred for information about relationships. [13]

But are these female "helpers" perpetuating gender stereotypes? It doesn't help matters that many female bots have subservient, docile voices. That's something that Amazon has begun to address in its recent version of Alexa: The intelligent bot has been reprogrammed to have little patience for harassment, for instance, and now sharply answers sexually explicit questions along the lines of "I'm not going to respond to that" or "I'm not sure what outcome you expected." [14]

Companies might consider offering different versions of their bots to appeal to a diverse customer base. Apple's Siri is now available in a male or female voice and can speak with a British, Indian, Irish, or Australian accent. It can also speak in a variety of languages, including French, German, Spanish, Russian, and Japanese. Although Siri typically defaults to a female voice, the default is male for Arabic, French, Dutch, and British English languages.

Just as important as the way they speak, AI bots must also be able to understand all types of voices. But right now, many don't. [15] To train voice recognition algorithms, companies have relied on speech corpora, or databases of

Copyright © Massachusetts Institute of Technology, 2018. All rights reserved. | Reprint #60216 | https://mitsmr.com/2DQz2XT

audio clips. Marginalized groups in society — low-income, rural, less educated, and non-native speakers — tend to be underrepresented in such data sets. Specialized databases can help correct such deficiencies, but they, too, have their limitations. The Fisher speech corpus, for example, includes speech from non-native speakers of English, but the coverage isn't uniform. Although Spanish and Indian accents are included, there are relatively few British accents. Baidu, the Chinese search-engine company, is taking a different approach by trying to improve the algorithms themselves. It is developing a new "deep speech" algorithm that it says will handle different accents and dialects.

Ultimately, we believe that AI will help create a more diverse and better world if the humans who work with the technology design, train, and modify those systems properly. This shift requires a commitment from the senior executives setting the direction. Business leaders may claim that diversity and inclusivity are core goals, but they then need to follow through in the people they hire and the products their companies develop.

The potential benefits are compelling: access to badly needed talent and the ability to serve a much wider variety of consumers effectively.

Copyright © Massachusetts Institute of Technology, 2018. All rights reserved. | Reprint #60216 | https://mitsmr.com/2DQz2XT

About the Authors

Paul R. Daugherty is Accenture's chief technology and innovation officer. He tweets @pauldaugh. H. James Wilson is managing director of IT and business research at Accenture Research. He tweets @hjameswilson. Rumman Chowdhury is a data scientist and social scientist, and Accenture's global lead for responsible AI. She tweets @ruchowdh.

References

1. L. Hardesty, "Study Finds Gender and Skin-Type Bias in Commercial Artificial Intelligence Systems," MIT News Office, Feb. 11, 2018.

2. E.T. Israni, "When an Algorithm Helps Send You to Prison," The New York Times, Oct. 26, 2017.

3. L. Camera, "Women Can Code — as Long as No One Knows They're Women," U.S. News & World Report, Feb. 18, 2016.

4. M. Muro, A. Berube, and J. Whiton, "Black and Hispanic Underrepresentation in Tech: It's Time to Change the Equation," The Brookings Institution, March 28, 2018.

5. "About Us," girlswhocode.com.

6. F. Dobbin and A. Kalev, "Why Diversity Programs Fail," Harvard Business Review 94, no. 7/8 (July-August 2016).

7. R. Locascio, "Thousands of Sexist AI Bots Could Be Coming. Here's How We Can Stop Them," Fortune, May 10, 2018.

8. "Inclusive Design," Microsoft.com.

9. T. Halloran, "How Atlassian Went From 10% Female Technical Graduates to 57% in Two Years," Textio, Dec. 12, 2017.

10. C. DeBrusk, "The Risk of Machine-Learning Bias (and How to Prevent It)," MIT Sloan Management Review, March 26, 2018.

11. J. Zou and L. Schiebinger, "AI Can Be Sexist and Racist — It's Time to Make It Fair," Nature, July 12, 2018.

12. D. Bass and E. Huet, "Researchers Combat Gender and Racial Bias in Artificial Intelligence," Bloomberg.com, Dec. 4, 2017.

13. B. Lovejoy, "Sexism Rules in Voice Assistant Genders, Show Studies, but Siri Stands Out," 9to5Mac.com, Feb. 22, 2017.

14. J. Elliot, "Let's Stop Talking to Sexist Bots: The Future of Voice for Brands," Fast Company, March 7, 2018.

15. S. Paul, "Voice Is the Next Big Platform, Unless You Have an Accent," Wired, March 20, 2017.

Copyright © Massachusetts Institute of Technology, 2018. All rights reserved. | Reprint #60216 | https://mitsmr.com/2DQz2XT

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.