WORK & FAMILY

A Crucial Step for Averting AI Disasters

Tech companies working on artificial intelligence find that a diverse staff can help avoid biased algorithms that cause public embarrassments
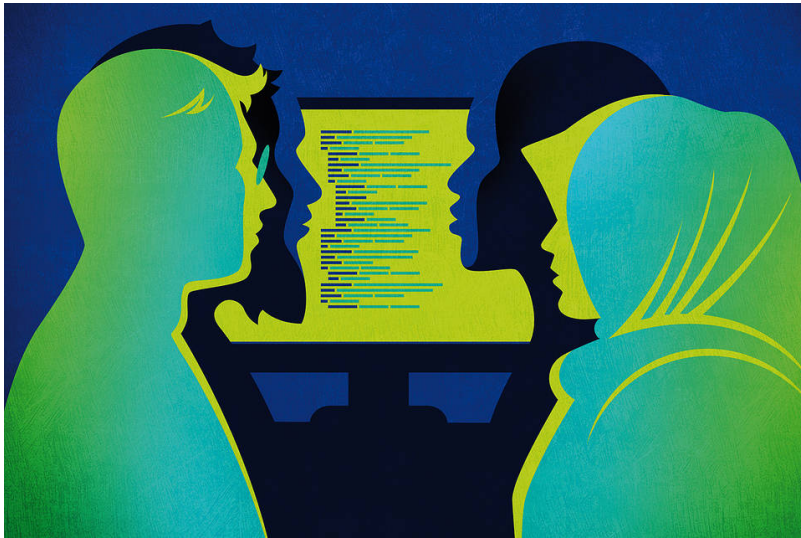

ILLUSTRATION: DARIA KIRPACH

By Sue Shellenbarger
Feb. 13, 2019 9:57 am ET

Artificial intelligence isn't always intelligent enough at the office.

One major company built a job-applicant screening program that automatically rejected most women's résumés. Others developed facial-recognition algorithms that mistook many black women for men.

The expanding use of AI is attracting new attention to the importance of workforce diversity. Although tech companies have stepped up efforts to recruit women and minorities, computer and software professionals who write AI programs are still largely white and male, Bureau of Labor Statistics data show.

Developers testing their products often rely on data sets that lack adequate representation of women or minority groups. One widely used data set is more than 74% male and 83% white, research shows. Thus, when engineers test algorithms on these databases with high numbers of

Deborah Harrison, left, leader of the editorial writing team for Microsoft's Personality Chat project, works with diverse colleagues from various creative, technical and artistic backgrounds to write small talk for bots. **PHOTO:** BARET YAHN

people like themselves, they may work fine.

The risk of building the resulting blind spots or biases into tech products multiplies exponentially with AI, damaging customers' trust and cutting into profit. And the benefits of getting it right expand as well, creating big winners and losers.

Flawed algorithms can cause freakish accidents, usually because they've been tested or trained on flawed or incomplete databases. Google came under fire in 2015 when its photo app tagged two African-American users as gorillas. The company quickly apologized and fixed the problem. And Amazon.com halted work a couple of years ago on an AI screening program for tech-job applicants that systematically rejected résumés mentioning the word "women's," such as the names of women's groups or colleges. (Reuters originally reported this development.) An Amazon spokeswoman says the program was never used to evaluate applicants.

Broader evidence of bias came in a 2018 study of three facial-recognition tools of the kind used by law-enforcement agencies to find criminal suspects or missing children. Analyzing a diverse sample of 1,270 people, the programs misidentified up to 35% of dark-skinned women as men, compared with a top error rate for light-skinned men of only 0.8%. The study was led by Joy Buolamwini, a researcher at the MIT Media Lab in Cambridge, Mass.

The findings have spurred calls for closer scrutiny. Microsoft recently called on governments to regulate facial-recognition technology and to require auditing of systems for accuracy and bias. The AI Now Institute, a research group at New York University, is studying ways to reduce bias in AI systems.

An algorithm can become a black box in the marketplace, however. Algorithms can learn and make predictions on data without being explicitly programmed to do so. This process continues in the background after a program is built, says Douglas Merrill, CEO of ZestFinance, a Los Angeles maker of machine-learning tools for financial-services companies.
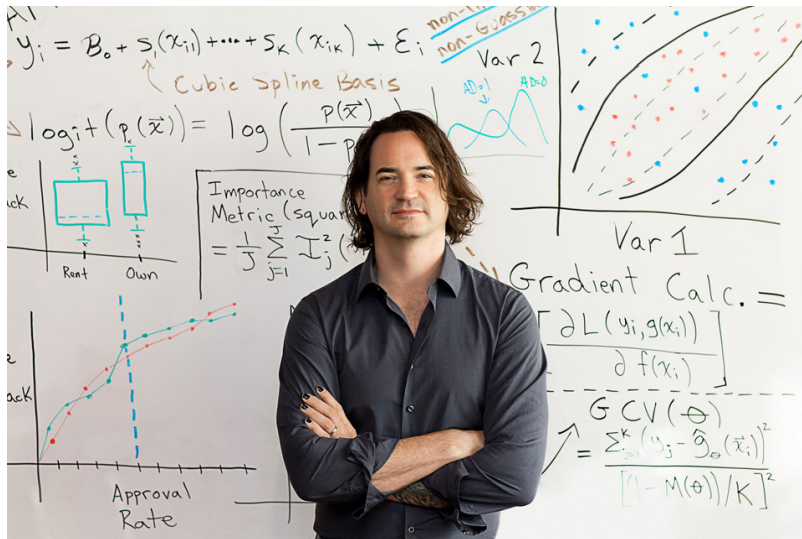
Douglas Merrill, CEO of ZestFinance in Los Angeles, says diverse employee teams may have more conflicts, but they also produce better AI programs. PHOTO: JEFF GALFER/ZESTFINANCE

Any biases in the algorithm can skew companies' decision-making in costly ways. One financial-services company's algorithm noticed that people with high mileage on their cars and those living in a particular state tended to be poor credit risks, Dr. Merrill says. Each factor alone made some sense, but combining the two would have led the company, unintentionally, to reject an undue number of African-American applicants, he says. After ZestFinance rewrote the algorithm and added a large number of additional criteria, many of those same applicants proved creditworthy.

Eliminating bias up front among those who write the code is essential. "That's why we work so hard on building diverse teams," says Dr. Merrill, a former CIO of Google. Asked about the makeup of his 100-person workforce, he ticks off a half-dozen groups his employees represent, including a high percentage of women, as well as military veterans and people with disabilities.

"The biases that are implicit in one team member are clear to, and avoided by, another," Dr. Merrill says. "So it's really key to get people who aren't alike."

Successful AI programs promise to open up new markets for some companies. Ford Motor Credit found in a joint 2017 study with ZestFinance that machine learning may enable it to broaden credit approvals among young adults and other applicants without lowering its underwriting standards.

Younger applicants are often routinely denied loans because they don't have a credit history and their incomes are low, Dr. Merrill says. Machine learning allows lenders to scrutinize a much larger number of decision-making criteria, including whether the applicant has paid rent and cellphone bills on time, made regular deposits into savings accounts and other measures of responsible behavior. This may help identify many more creditworthy young people. "The answer to almost every question in machine learning is more data," Dr. Merrill says.

A spokeswoman for Ford Motor Credit says the company is continuing to work on machine-learning applications.

Affectiva, an AI company based in Boston, has attracted more than 100 corporate customers by amassing a database of 4 billion facial images from 87 countries. It develops technology to read the emotional expressions on those faces accurately, regardless of race, ethnicity or gender. Companies use its software to study consumers' reactions to proposed ads and promotions, and auto makers use it to monitor drivers for drowsiness and distraction.

At one point, Rana el Kaliouby, Affectiva CEO and co-founder says, women working in the company's Cairo office asked, "Are there any people in here who look like us?" Engineers quickly added images of Muslim women wearing hijabs.

"You need diversity in the data, and more important, in the team that's designing the algorithm," Dr. el Kaliouby says. "If you're a 30-year-old white guy who's programming this algorithm, you might not think about, 'Oh, does this data set include a woman wearing a hijab?' "

Beyond racial and gender diversity, Microsoft recruits employees with diverse creative and artistic skills to help write conversational language for its Cortana virtual assistant and Personality Chat, an AI program that handles small talk for bots developed by others. Team members have included a playwright, a poet, a comic-book author, a philosophy major, a songwriter, a screenwriter, an essayist and a novelist, whose professional skills equip them to write upbeat language for the bots and anticipate diverse users' reactions, says Deborah Harrison, a senior manager and team leader. They also teach the bots to avoid, say, misusing ethnic slang or making sexualized remarks.

One team labored over how Cortana should respond to a user who announced, "I'm gay," Ms. Harrison says. Her team came up with a pleasant, nonjudgmental response: "I'm AI." But they weren't satisfied, she says. It was a teenage visitor to their lab who suggested a tweak that finally pleased everyone: "Cool. I'm AI."

Write to Sue Shellenbarger at sue.shellenbarger@wsj.com

MORE FROM WORK & FAMILY

- Find Your Zen, Then Crush Your Job Interview **October 21, 2019**

- You Can Overcome a Long Gap in Your Résumé **October 14, 2019**

- The Dangers of Hiring for Cultural Fit **September 23, 2019**

- When You Think Your Child's Teacher Is a Bad Fit **September 16, 2019**

- Why We Should Teach Kids to Call the Robot 'It' **August 26, 2019**