

https://hbr.org/2018/02/can-we-keep-our-biases-from-creeping-into-ai?utm_campaign=hbr&utm_source=twitter&utm_medium=social

Kriti Sharma

Can We Keep Our Biases from Creeping

- [Kriti Sharma](#)

FEBRUARY 09, 2018

- [Summary](#)
- [Save](#)
- [Share](#)
- [Comment](#)
- [Text Size](#)
- [Print](#)
- [8.95 Buy Copies](#)

RECOMMENDED

HBR'S
10
MUST
READS

BONNIE
“Why B
Organ
an Au
Realit
By Mich
James E

On AI, Analytics and the New Machine Age

If you read nothing else on how intelligent
revolutionizing business, read these defining
from **Harvard Business Review**.

PAPERBACK + EBOOK

• [HBR's 10 Must Reads on AI, Analytics, and the...](#)

BUNDLES

34.95 [Add to Cart](#)

HBR'S 10 MUST READS

BON
“Now W
Managin
By Joan
Suzanne

The definitive
management ideas
of the year from
Harvard Business Review.

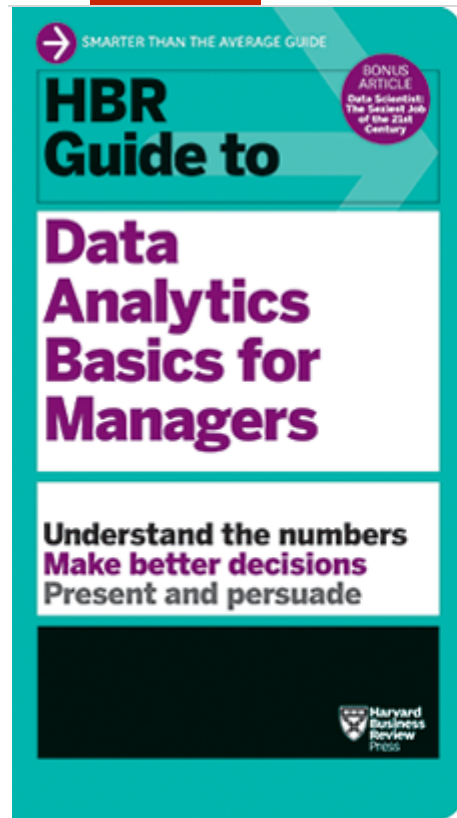
201

PAPERBACK + E

• [HBR's 10 Must Reads 2019: The Definitive...](#)

BUNDLES

34.95 [Add to Cart](#)



DOC

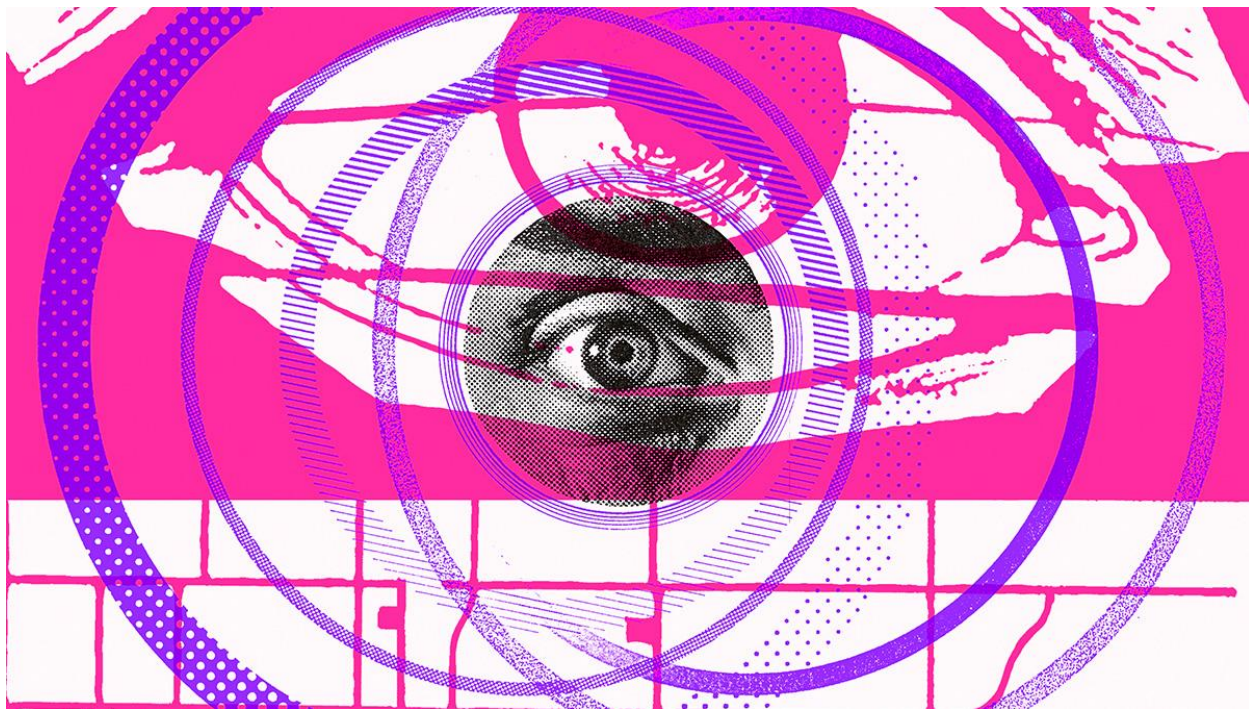
EBOOK + TOOLS

•

[HBR Guide to Data Analytics Basics for...](#)

BOOK

49.95 [Add to Cart](#)



CSA IMAGES/PRINTSTOCK COLLECTION/GETTY IMAGES

Eminent industry leaders [worry](#) that the biggest risk tied to artificial intelligence is the militaristic downfall of humanity. But there's a smaller community of people committed to addressing two more tangible risks: AI created with harmful biases built into its core, and AI that does not reflect the diversity of the users it serves. I am proud to be part of the second group of concerned practitioners. And I would argue that not addressing the issues of bias and diversity could lead to a different kind of weaponized AI.

The good news is that AI is an opportunity to build technology with less human bias and built-in inequality than has been the case in previous innovations. But that will only happen if we expand AI talent pools and explicitly test AI-driven technologies for bias.

Eliminating Biases in AI: The People

Technology inevitably reflects its creators in a myriad of ways, conscious and unconscious. The tech industry remains very male and fairly culturally homogeneous. This lack of diversity is reflected in the products it produces. For instance, AI assistants like Apple's Siri or Amazon's Alexa, which have default female names, voices, and personas, are largely seen as helpful or passive supporters of a user's lifestyle. Meanwhile, their male-branded counterparts like IBM's Watson or Salesforce's Einstein are perceived as complex problem-solvers tackling global issues. The quickest way to flip this public perception on its head is to render AI genderless, something I advocate for tirelessly and practice with Sage's personal finance assistant, [Pegg](#). The more long-term approach requires expanding the talent pool of people working on the next generation of AI technologies.

Diversifying the AI talent pool isn't just about gender. Currently, AI development is a PhD's game. The community of credentialed people creating scalable AI for businesses is relatively small. While the focus on quality and utility needs to remain intact, expanding the diversity of people working on AI to include people with nontechnical professional backgrounds and less advanced degrees is vital to AI's sustainability. As a start, companies developing AI should consider hiring creatives, writers, linguists, sociologists, and passionate people from nontraditional professions. Over time, they should commit to supporting training programs that can broaden the talent pool beyond those who've graduated from elite universities. Recruiting diverse sets of people will also help to improve and reinvent AI user experiences.

Eliminating Biases in AI: The Technology

Software and hardware engineers regularly test new technology products to ensure they are not harmful to people or businesses that will use them in the real world. Engineers conduct testing in labs and research facilities before a product launches. Ideally, any harmful attributes of the product are uncovered and removed during the testing phase. While that is not always the case, a fundamental and virtually universal commitment to testing does significantly decrease potential risks for everyone producing the products.

The same testing approach is used in the development of new AI technologies. People building AI test their systems for utility, safety, and scalability, with a focus on eliminating product flaws and security vulnerabilities. However, AI is unique in that it typically keeps learning and therefore changing after it leaves the lab.

Currently, though, there is a lack of testing AI products throughout their development cycle to detect potential harms they may do to humans socially, ethically, or emotionally once they hit the market. One way to remedy this is by adding bias testing to a new product's development cycle. Adding such a test in the R&D phase would help companies remove harmful biases from algorithms that run their AI applications and datasets that they pull from to interact with people.

Bias testing would also help account for the blind spots created by the lack of diversity at present in terms of who builds AI. It could help highlight issues which might not be immediately or traditionally obvious to the engineer and, most importantly, to the end user, [like how an automated assistant should respond when harassed](#). Some degree of testing will need to continue after a product is released, since AI algorithms can evolve as they encounter new data.

AI-driven technologies will continue to integrate into the everyday lives of people around the world in meaningful ways. It will become commonplace at the office and at home – and not only in the form of voice assistants like Amazon's Alexa or Google Home. AI-driven enterprise technologies will improve commercial productivity, close workforce skill gaps and bolster customer experience across industries. That's why now is the right time to implement methods that eliminate harmful biases and take gender out of the equation,

expand the population of people working on technologies, and address trust issues with AI. The technology community must do all of these things to make AI's journey into the mainstream one that improves people's lives rather than tearing them apart.

[Kriti Sharma](#) is the vice president of AI at [Sage](#), a market leader in business software that helps manage everything from money to people. Kriti was listed in Forbes 30 under 30 in 2017 and was number 44 in the Recode 100 list of influencers in Technology, Business and Media.