Big Data and Gender-Biased Algorithms

ERKAN SAKA
İstanbul Bilgi University, Turkey

Algorithms and Big Data have recently become household terms and, almost immediately, issues of bias began to surface in the critical literature. Race bias was the first to be named but gender bias in algorithms quickly became visible as well. For example, Google searches demonstrated that selecting professions which are typically associated with men will generate images which are predominantly of men, even if women are an equal or majority share of that profession's workforce. Amazon's experimental artificial intelligence (AI) recruiting engine had to be taken out of action when the company realized that the algorithm was beginning to exclude women. The original aim of the search was to find successful but emerging professionals to recruit for the ever-growing company, but the AI started with data from already successful professionals who are mostly men. The AI "decided" to exclude women based on a historically biased data relating to the characteristics of the relevant workforce. In the meantime, an MIT study about facial recognition software projects demonstrated that both race and gender bias had an impact in all significant projects generated by Amazon, Microsoft, and IBM, albeit to differing degrees: the group most affected were Black women. In one project, the software mis-recognized darker-skinned women as men almost a third of the time.

Much of the recent research focusing on gender bias suggests that the primary reason it occurs is because of the predominance of a particular gender type behind the making of algorithms. For example, Gehl, Moyer-Horner, and Yeo (2017) reviewed 102 peer-reviewed articles on CVPF (computer vision-based pornography filtering) software and found that pictures of lone, thin, naked female bodies constituted the ideal type of pornography: most software ignored not only the penis but also a varied range and size of human bodies, male, female, and trans. Researchers argue that computer scientists, who are overwhelmingly straight men, project and then tool their own preferences for pornography into the design of the software, thus normalizing and idealizing as objective what is actually an entirely subjectively produced process. An investigation of artificial intelligence studies (West, Whittaker, & Crawford, 2019) also emphasizes the lack of participation of individuals with diverse gender identities and a narrow focus on "women in tech" which may privilege White women and exclude everyone else. In any case, software systems, like any other human-generated process, take shape according to the social organization of production (Raymond & Steele, 1996; Seaver, 2018).

Another primary reason for gender bias relates to the Big Data that feed the algorithms. Datasets may be unrepresentative in the sense that minority gender perspectives do not pull through into the sampling. However, even if they are involved in the sample in adequate numbers, such inclusion may nonetheless ignore the complexity and context of social systems. If the data thus ignore social change, then both past and

The International Encyclopedia of Gender, Media, and Communication. Karen Ross (Editor-in-Chief),
Ingrid Bachmann, Valentina Cardo, Sujata Moorti, and Marco Scarcelli (Associate Editors).

DOI: 10.1002/9781119429128.iegmc267

present biases may simply be forwarded into the future. The collection, handling, and purpose of large datasets need to be further explored and exposed to understand better how these processes can perpetuate gender and racial bias and discrimination. In what they call "The Bride Problem," Zou and Schiebinger (2018) found that using the term "bride" in ImageNet would very likely yield a White woman who wears a white wedding dress. In contrast, a North Indian woman who wears a wedding sari or *lehenga* will be labeled under "performance art," although she is also a bride. As nearly half of bride images come from the United States, the dataset here ignores both geodiversity and representational proportions. Relatedly, humanoid robotics and especially those linking the language of assistance with a feminine voice, recirculate existing gender stereotypes of female domesticity. Linguistic biases may also reinforce gender biases and Gendered Innovations, a Stanford-based initiative, provides examples from Google Translate. This tool uses a "faulty" algorithm which selects the "most-used" pronoun by default so that when it translates languages without gendered pronouns into English, it defaults to "he said" because the phrase appears more on the internet than "she said."

Few Big Data projects fully integrate information ethics in their research efforts and some critics call for an auditing initiative (Mittelstadt, 2016) and more transparency and supervision for algorithmic processes (Diakopoulos, 2014).The Big Data that feed the algorithms seem to be the root cause of subjective decision making in processes of classification, prioritization, association, and filtering (Diakopoulos, 2014). There are, however, limitations to relying solely on increased transparency as a countermeasure since contemporary networked discrimination is more subtle than the traditional biases related to race, class, and gender (boyd, Levy, & Marwick, 2014). Still, design choices in platform selection may also reinforce gender stereotypes (Adams & Ní Loideáin, 2019). Platforms have recently become significant venues for gender politics since platform centrality has become evident in digital communications. Both the #Gamergate scandal and the "Fappening" leaks became very visible examples of gender bias on platforms such as Reddit, 4chan, and Twitter. Design decisions and assumptions about users, made these platforms "nests" of misogynistic activism (Massanari, 2017) and hostile places for anyone who was not a heterosexual male. While less critiqued, popular platforms such as Google and Facebook also contribute to gender biases through the design of their interfaces, protocols, and databases which are created by biased algorithms.

Algorithms' impact goes beyond their intended functions. They deliver a deep mediatization of reality (Couldry & Hepp, 2018) and in a way, they are replacing credentialized experts, the scientific method, common sense, or even the word of God as voices to rely on (Gillespie, 2014). With this power, algorithms can introduce new forms of gender-based discrimination and/or reproduce existing ones. Alternatively, seamless depersonalized domination by algorithms seems unlikely, thanks to the very complexity and arbitrary nature of human intervention (Seaver, 2018) and the unstable and malleable nature of algorithms themselves (Seaver, 2013). However, their contribution to gender bias in everyday life is undeniable and diversification of datasets that are used to build algorithms (Nafus, 2018), more diversity among algorithm designers and builders, and more transparency in algorithmic processes are all mechanisms which can mitigate their biasing impact.

SEE ALSO: Gender and Media; Gender and Technology; Gendered Hate Online; Women, Technology, and the Gender Gap

References

Adams, R., & Ní Loideáin, N. (2019). *Addressing indirect discrimination and gender stereotypes in AI virtual personal assistants: The role of international human rights law*. Paper presented at Annual Cambridge International Law Conference 2019, New Technologies: New Challenges for Democracy and International Law. University of Cambridge, UK.

boyd, d., Levy, K., & Marwick, A. (2014). The networked nature of algorithmic discrimination. In S. Peña Gangadharan (with V. Eubanks & S. Barocas) (Eds.), *Data and Discrimination: Collected Essays* (pp. 53–57). New America's Open Technology Institute. Retrieved from https://www.ftc.gov/system/files/documents/public_comments/2014/10/00078-92938.pdf

Couldry, N., & Hepp, A. (2018). *The mediated construction of reality*. Oxford, UK: Polity Press.

Diakopoulos, N. (2014). *Algorithmic accountability reporting: On the investigation of black boxes*. Tow Center for Digital Journalism. Retrieved from https://academiccommons.columbia.edu/doi/10.7916/D8ZK5TW2

Gehl, R. W., Moyer-Horner, L., & Yeo, S. K. (2017). Training computers to see internet pornography: Gender and sexual discrimination in computer vision science. *Television & New Media, 18*(6), 529–547.

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–194). Cambridge, MA: MIT Press.

Massanari, A. (2017). #Gamergate and the fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society, 19*(3), 329–346.

Mittelstadt, B. (2016). Auditing for transparency in content personalization systems. *International Journal of Communication, 10,* 4991–5002. Retrieved from https://ijoc.org/index.php/ijoc/article/view/6267/1808

Nafus, D. (2018). Exploration or algorithm? The undone science before the algorithms. *Cultural Anthropology, 33*(3), 368–374.

Raymond, E. S., & Steele, G. L. (1996). *The new hacker's dictionary*. Cambridge, MA: MIT Press.

Seaver, N. (2013). *Knowing algorithms*. Presented at the Media in Transition 8, Cambridge, MA. Retrieved from http://nickseaver.net/papers/seaverMiT8.pdf

Seaver, N. (2018). What should an anthropology of algorithms do? *Cultural Anthropology, 33*(3), 375–385.

West, S. M., Whittaker, M., & Crawford, K. (2019). *Discriminating systems: Gender, race, and power in AI*. AI Now Institute. Retrieved from https://ainowinstitute.org/discriminatingsystems.html

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist—it's time to make it fair. *Nature, 559,* 324–326. Retrieved August 20, 2019, from https://www.nature.com/articles/d41586-018-05707-8

Further Reading

Collett, C., & Dillon, S. (2019). *AI and gender: Four proposals for future research*. Cambridge, UK: The Leverhulme Centre for the Future of Intelligence.

Social Media Collective. (2011–2018). Critical algorithm studies: A reading list. Retrieved from https://socialmediacollective.org/reading-lists/critical-algorithm-studies

Erkan Saka (PhD Anthropology, Rice University) is an associate professor of media and journalism studies at İstanbul Bilgi University. He has been a visiting scholar in anthropology at University of California, Irvine (Fall 2018) and Science and Technology Studies at Massachusetts Institute of Technology (Spring 2019). His recent and current research topics include online information fact-checking and verification issues, political trolling, digital web archiving and internet histories, and ethnography of cryptocurrency circles in Turkey.