Tech Ethics Animated – AI & Discrimination

Transcript

An internet search engine sold ads to advertisers that would be shown to users. As someone was looking up where to live, for example, the search engine would show targeted ads on the side based on who the user was. The search engine, however, allowed advertisers to exclude people of "unknown gender" from seeing their ads. While in default mode, gender was not considered when determining who sees the ads. However, companies running ads on the platform could choose not to show their ads to people who didn't identify themselves as "male" or "female." So, if someone did not identify as male or female using this platform, they would not be shown the ads of employers that decided not to allow them to see the ads.

In the U.S., two concepts dominate discrimination law: disparate treatment and disparate impact. Both are related to protected classes.

Disparate treatment occurs when someone is denied equal treatment in their access to products, services, or opportunities because of discriminatory practices or policies. In these cases, members of protected classes such as race, sex, religion, national origin, or ethnic group are evaluated in selection procedures, such as employment and housing, with different standards based on their class. An example is the search engine allowing advertisers to exclude non-binary people from seeing their ads—the search engine provided tools so that the advertisers could intentionally exclude people from seeing ads for jobs, housing, or financial products based on their gender.

Disparate impact has a similar effect—someone is excluded or discriminated against based on protected status—but does not require intent. In this case, policies or practices appear to be neutral but have a disproportionately adverse effect on members of protected groups more than others. For example, a social networking service excluded women from seeing certain job ads based on gender. However, in this case, there was no option to specify the audience based on demographics, but their algorithm discriminated against women. Apparently, their algorithm automatically targeted based on the current demographic distribution of jobs, which is supposedly related to historical reasons. While there was no intent, women were still impacted.

Today, there is a reoccurring problem of data and algorithms reproducing biases against historically marginalized groups. Efforts of trying to resolve this problem have matched anti-discrimination discourses in the law. And discrimination analysis is helpful to understand what is wrong in a given application of AI. However, Professor Anna Lauren

Hoffmann highlights the limitations of *only* using a discrimination lens when analyzing the ethical implications of AI.

Discrimination is important but should be the bare minimum we expect from our technologies. Professor Hoffmann notes that discrimination tends to focus on specific bad actors—like the search engine that supported discrimination in advertising and hiring on their platform—when in reality, there may be systems of power that are impacting the moral harms. We also focus on only particular classes of individuals such as race or gender with discrimination analysis and mistakenly think of them as easily classified rather than possibly intersectional—for instance, if the user on the platform identified as non-binary and is Hispanic, as well. Finally, only some types of decisions are governed by these laws, and those are usually around getting a job, a loan, a house, etc. However, there could be more advertising discrimination that went unnoticed due to the lack of governance for other decisions.

That being said, there are many instances where the decisions being augmented by AI—such as finding a job, being approved for a loan, getting into a school—are governed by discrimination law. And ensuring AI decisions are not discriminating against users is a first step to the ethical analysis of AI.

CREDITS

Animation Team

        Josiah Broughton (Lead)

        Michael Simon

Script

        Brooke Anquillare

Faculty Advisor

        Carolina Villegas-Galaviz, Ph.D.

Associated Readings

Martin, K. (2022). Discrimination and Data Analytics. In *Ethics of Data and Analytics: Concepts and Cases*. Taylor & Francis: USA, 291-295.

Barocas, S., and Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, 104(3).

Hoffmann, A. L. (2019). Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society*, 22(7).

Merrill, J. B. (2021). Google Has Been Allowing Advertisers to Exclude Nonbinary People from Seeing Job Ads. *The Markup*.

Hao, K. (2021). Facebook's ad algorithms are still excluding women from seeing jobs. *MIT Technology Review*.

TEC Technology Ethics Center

UNIVERSITY OF NOTRE DAME