Discrimination in Online Personalization: A Multidisciplinary Inquiry

Amit Datta, Anupam Datta, Cynthia Dwork, Deirdre Mulligan, and Michael Carl Tschantz

(This work is closely related to the work submitted by Michael Tschantz as 00067. We could combine presenting this work with the presentation of 00067.)

Regulators, advocates, and scholars have voiced concern with the possibility that algorithms will mask, produce, or reproduce illegal discrimination, and perpetuate stereotypes. In this paper, we recount experiments reported in the computer science literature finding that the ads shown by Google differ based on the gender set in Google Ad Settings. Male identification resulted in more ads for a career coaching service that promoted high paying jobs. While the gender setting appeared causal in the experiments, the complexity of the advertising ecosystem makes it difficult to assess how it interacts with other components in the large third-party ad network that is Google's DoubleClick. Google provides Ad Settings, which displays inferences Google has made about a user's demographics and interests as well as user contributed data. Users can view and edit these settings at google.com/settings/ads. The Ad Settings profile is used to personalize advertisements. However, the Ad Settings profile is but one element of the advertising ecosystem. It interacts with the demographic and key word selections of advertisers—which allow ads explicitly to be targeted to or away from users based on gender, and can differently target men and women due to keywords or other parameters that acts as a proxy for gender—with the Google controlled optimization algorithm that aims to maximize click-through-rates (CTR) and may explicitly learn and tailor base on gender, or on attributes that align with one gender or the other—and with the prices other advertisers are willing to pay per impression, as well as their choices about demographics and keywords. Thus while toggling between genders yields different ads, exactly why it does so remains opaque.

We document the various ways that an advertising system could yield different ads to men and women. We document the various ways in which advertisers can target ads in the Google platform, and run our own experiments demonstrating the ability to have facially discriminatory advertising practices for employment ads through the Google ad platform. We consider how this may interact with the choices of other advertisers, the bidding for impressions, and Google’s effort to maximize return on investment.

We then explore the legal questions and policy concerns raised by these results. We explore how this behavior relates to prohibitions against gender discrimination in employment advertising under state and federal civil rights law, as well as the interaction between these statutes and Section 230 of the Communications Decency Act, which provides broad protection against liability to interactive computer services for other entities speech. We identify several aspects on the ad serving ecosystem that present challenges for existing legal frameworks, specifically: the lack of upfront visibility into the likely output of adword choices; the inability of individual advertisers to know or control how their choices interact with those of other advertisers to influence the display of advertisements across the sexes; and, the difficult of assigning responsibility or blame in such multi-agent, real-time systems.

Given the difficulties facing individual advertisers seeking to target ads without discriminating, we identify a need to address it at the ad serving ecosystem level, and discuss two structural approaches for mitigating discrimination in online behavioral advertising. The first seeks to create an architecture to operationalize decisions about fairness in the ad serving ecosystem. It would prevent such discrimination before it happens by making machine learning algorithms aware of non-discrimination imperatives. This approach, Fairness through Awareness, offers a formal rigorous approach to operationalizing non-discriminatory machine learning—but significantly requires what is fair to be set outside the system. The second builds upon the detection tools that we discuss in this paper, and relies on the legal process for investigation and accountability. It uses detect such discrimination by an organization and to hold it accountable.

This paper intends to stimulate collaboration between the legal and computer science communities. As computers driven by opaque machine learning algorithms and murky data sets increasingly determine people's fates, limiting discriminatory outcomes will depend upon the creation of technical systems that help promote fair outcomes. The interdependence and secrecy of these complex systems make it difficult to assign blame. Building in fairness is a structural response to a structural problem however, it requires upfront decisions about fairness and the price to be paid to support it.