Auditing for Discrimination in Algorithms Delivering Job Ads

Basileal Imana
University of Southern California
Los Angeles, CA, USA

Aleksandra Korolova
University of Southern California
Los Angeles, CA, USA

John Heidemann
USC/Information Science Institute
Los Angeles, CA, USA

ABSTRACT

Ad platforms such as Facebook, Google and LinkedIn promise value for advertisers through their targeted advertising. However, multiple studies have shown that ad delivery on such platforms can be skewed by gender or race due to hidden algorithmic optimization by the platforms, even when not requested by the advertisers. Building on prior work measuring skew in ad delivery, we develop a new methodology for black-box auditing of algorithms for *discrimination* in the delivery of *job advertisements*. Our first contribution is to identify the distinction between skew in ad delivery due to protected categories such as gender or race, from skew due to differences in qualification among people in the targeted audience. This distinction is important in U.S. law, where ads may be targeted based on qualifications, but not on protected categories. Second, we develop an auditing methodology that distinguishes between skew explainable by differences in qualifications from other factors, such as the ad platform's optimization for engagement or training its algorithms on biased data. Our method controls for job qualification by comparing ad delivery of two concurrent ads for similar jobs, but for a pair of companies with different de facto gender distributions of employees. We describe the careful statistical tests that establish evidence of non-qualification skew in the results. Third, we apply our proposed methodology to two prominent targeted advertising platforms for job ads: Facebook and LinkedIn. We confirm skew by gender in ad delivery on Facebook, and show that it cannot be justified by differences in qualifications. We fail to find skew in ad delivery on LinkedIn. Finally, we suggest improvements to ad platform practices that could make external auditing of their algorithms in the public interest more feasible and accurate.

CCS CONCEPTS

• **Social and professional topics → Technology audits**; **Employment issues**; **Socio-technical systems**; **Systems analysis and design**.

ACM Reference Format:
Basileal Imana, Aleksandra Korolova, and John Heidemann. 2021. Auditing for Discrimination in Algorithms Delivering Job Ads. In *Proceedings of the Web Conference 2021 (WWW '21), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3442381.3450077

1 INTRODUCTION

Digital platforms and social networks have become popular means for advertising to users. These platforms provide many mechanisms that enable advertisers to target a specific audience, i.e. specify the criteria that the member to whom an ad is shown should satisfy. Based on the advertiser's chosen parameters, the platforms employ optimization algorithms to decide who sees which ad and the advertiser's payments.

Ad platforms such as Facebook and LinkedIn use an automated algorithm to deliver ads to a subset of the targeted audience. Every time a member visits their site or app, the platforms run an ad auction among advertisers who are targeting that member. In addition to the advertiser's chosen parameters, such as a bid or budget, the auction takes into account an ad *relevance score*, which is based on the ad's predicted engagement level and value to the user. For example, from LinkedIn's documentation [37]: "scores are calculated … based on your predicted campaign performance and the predicted performance of top campaigns competing for the same audience." Relevance scores are computed by ad platforms using algorithms; both the algorithms and the inputs they consider are proprietary. We refer to the algorithmic process run by platforms to determine who sees which ad as *ad delivery optimization.*

Prior work has hypothesized that ad delivery optimization plays a role in skewing recipient distribution by gender or race even when the advertiser targets their ad inclusively [15, 31, 54, 56]. This hypothesis was confirmed, at least for Facebook, in a recent study [2], which showed that for jobs such as lumberjack and taxi driver, Facebook delivered ads to audiences skewed along gender and racial lines, even when the advertiser was targeting a gender- and race-balanced audience. The Facebook study [2] established that the skew is not due to advertiser targeting or competition from other advertisers, and hypothesized that it could stem from the proprietary ad delivery algorithms trained on biased data optimizing for the platform's objectives (§2.1).

Our work focuses on developing an auditing methodology for measuring skew in the delivery of *job ads*, an area where U.S. law prohibits discrimination based on certain attributes [59, 61]. We focus on expanding the prior auditing methodology of [2] to bridge the gap between audit studies that demonstrate that a platform's ad delivery algorithm results in skewed delivery and studies that provide evidence that the skewed delivery is discriminatory, thus bringing the set of audit studies one step closer to potential use by regulators to enforce the law in practice [14]. We identify one such gap in the context of job advertisements: controlling for bona fide occupational qualifications [61] and develop a methodology

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-8312-7/21/04.
https://doi.org/10.1145/3442381.3450077

to address it. We focus on designing a methodology that assumes no special access beyond what a regular advertiser sees, because we believe that auditing of ad platforms in the public interest needs to be possible by third-parties — and society should not depend solely on the limited capabilities of federal commissions or self-policing by the platforms.

Our first contribution is to examine how the occupational qualification of an ad's audience affects the legal liability an ad platform might incur with respect to discriminatory advertising (§2). Building upon legal analysis in prior work [14], we make an additional distinction between skew that is due to a difference in occupational qualifications among the members of the targeted ad audience, and skew that is due to (implicit or explicit use of) protected categories such as gender or race by the platform's algorithms. This distinction is relevant because U.S. law allows differential delivery that is justified by differences in qualifications [61], an argument that platforms are likely to use to defend themselves against legal liability when presented with evidence from audit studies such as [2, 15, 31, 54, 56].

Our second contribution is to propose a novel auditing methodology (§4) that distinguishes between a delivery skew that could be a result of the ad delivery algorithm merely incorporating job qualifications of the members of the targeted ad audience from skew due to other algorithmic choices that correlate with gender- or racial- factors, but are not related to qualifications. Like the prior study of Facebook [2], to isolate the role of the platform's algorithms we control for factors extraneous to the platform's ad delivery choices, such as the demographics of people on-line during an ad campaign's run, advertisers' targeting, and competition from other advertisers. Unlike prior work, our methodology relies on simultaneously running *paired* ads for several jobs that have *similar qualification requirements* but have *skewed de facto (gender) distribution.* By "skewed de facto distribution", we refer to existing societal circumstances that are reflected in the skewed (gender) distribution of employees. An example of such a pair of ads is a delivery driver job at Domino's (a pizza chain) and at Instacart (a grocery delivery service). Both jobs have similar qualification requirements but one is de facto skewed male (pizza delivery) and the other – female (grocery delivery) [17, 52]. Comparing the delivery of ads for such pairs of jobs ensures skew we may observe can not be attributed to differences in qualification among the underlying audience.

Our third contribution is to show that our proposed methodology distinguishes between the behavior of ad delivery algorithms of different real-world ad platforms, and identify those whose delivery skew may be going beyond what is justifiable on the basis of qualifications, and thus may be discriminatory (§5). We demonstrate this by registering as advertisers and running job ads for real employment opportunities on two platforms, Facebook and LinkedIn. We apply the same auditing methodology to both platforms and observe contrasting results that show statistically significant gender-skew in the case of Facebook, but not LinkedIn.

We conclude by providing recommendations for changes that could make auditing of ad platforms more accessible, efficient and accurate for public interest researchers (§6.2).

2 PROBLEM STATEMENT

Our goal is to develop a novel methodology that measures skew in ad delivery that is not justifiable on the basis of differences in job qualification requirements in the targeted audience. Before we focus on qualification, we first enumerate the different potential sources of skew that need to be taken into consideration when measuring the role of the ad delivery algorithms. We then discuss how U.S. law may treat qualification as a legitimate cause for skewed ad delivery.

We refer to algorithmic decisions by ad platforms that result in members of one group being over- or under-represented among the ad recipients as "skew in ad delivery". We consider groups that have been identified as legally protected (such as gender, age, race). We set the baseline population for measuring skew as the qualified and available ad platform members targeted by the campaign (see §4.4 for a quantitative definition).

2.1 Potential Sources of Skew

Our main challenge is to isolate the role of the platform's algorithms in creating skew from other factors that affect ad delivery and may be used to explain away any observed skew. This is a challenge for a third-party auditor because they investigate the platform's algorithms as a black-box, without access to the code or inputs of the algorithm, or access to the data or behavior of platform members or advertisers. We assume that the auditor has access only to ad statistics provided by the platform.

Targeted advertising consists of two high-level steps. The advertiser *creates* an ad, specifies its target audience, campaign budget, and the advertiser's objective. The platform then *delivers* the ad to its users after running an auction among advertisers targeting those users. We identify four categories of factors that may introduce skew into this process:

First, an advertiser can select **targeting parameters and an audience** that induce skew. Prior work [5, 6, 54, 57, 64] has shown that platforms expose targeting options that advertisers can use to create discriminatory ad targeting. Recent changes in platforms have tried to disable such options [22, 49, 55].

Second, an ad platform can make **choices in its ad delivery optimization algorithm** to maximize ad relevance, engagement, advertiser satisfaction, revenue, or other business objectives, which can implicitly or explicitly result in a skew. As one example, if an image used in an ad receives better engagement from a certain demographic, the platform's algorithm may learn this association and preferentially show the ad with that image to the subset of the targeted audience belonging to that demographic [2]. As another example, for a job ad, the algorithm may aim to show the ad to users whose professional backgrounds better match the job ad's qualification requirements. If the targeted population of qualified individuals is skewed along demographic characteristics, the platform's algorithm may propagate this skew in its delivery.

Third, an advertiser's **choice of objective** can cause a skew. Ad platforms such as LinkedIn and Facebook support advertiser objectives such as reach and conversion. *Reach* indicates the advertiser wants their ad to be shown to as many people as possible in their target audience, while for *conversion* the advertiser wants as many ad recipients as possible to take some action, such as clicking through to their site [20, 39]. Different demographic groups may have different propensities to take specific actions, so a *conversion* objective can implicitly cause skewed delivery. When the platform's implementation of the advertiser's objective results in a discriminatory skew, the responsibility for it can be a matter of dispute (see §2.2).

Finally, there may be **other confounding factors** that are not under direct control of a particular advertiser or the platform leading to skew, such as differing sign-on rates across demographics, time-of-day effects, and differing rates of advertiser competition for users from different demographics. For example, delivery of an ad may be skewed towards men because more men were online during the run of the ad campaign, or because competing advertisers were bidding higher to reach the women in the audience than to reach the men [2, 18, 31].

In our work, we focus on isolating skew that results from an ad delivery algorithm's optimization (the second factor). Since we are studying job ads, we are interested in further distinguishing skew due to an algorithm that incorporates qualification in its optimization from skew that is due to an algorithm that perpetuates societal biases without a justification grounded in qualifications. We are also interested in how job ad delivery is affected by the objective chosen by the advertiser (the third factor). We discuss our methodology for achieving these goals in §4.

2.2 Discriminatory Job Ads and Liability

Building on a legal analysis in prior work [14], we next discuss how U.S. anti-discrimination law may treat job qualification requirements, optimization objectives, and other factors that can cause skew, and discuss how the applicability of the law informs the design of our methodology[1].

Our work is unique in underscoring the implications of qualification when evaluating potential legal liability ad platforms may incur due to skewed job ad delivery. We also draw attention to the nuances in analyzing the implications of the optimization objective an advertiser chooses. We focus on Title VII, a U.S. law which prohibits preferential or discriminatory employment advertising practices using attributes such as gender or race [61]. We interpret this law to apply not just to actions of advertisers but also to *outcomes* of ad delivery.

Title VII allows entities who advertise job opportunities to legally show preference based on *bona fide occupational qualifications* [61], which are requirements necessary to carry out a job function. While it is unclear whether the scope of Title VII applies to ad platforms (as discussed by Datta *et al.* [14]), to the extent that it may apply, it is conceivable that a platform such as Facebook can use qualification as an exception to argue that the skew arising from its ad delivery optimization does not violate the law. They may argue that skew (shown in prior work [2]) simply reflects job qualifications. Therefore, our goal is to design an auditing methodology that can distinguish between skew due to ad platform's use of qualifications from skew due to other algorithmic choices by the platform. The methodology to make such a distinction is one of our main contributions relative to prior work. It also brings findings from audit studies such as ours a step closer to having the potential to be used by regulators to enforce the law in practice.

As discussed in §2.1, the objective an advertiser chooses can also be a source of skew. If different demographic groups tend to engage with ads differently, using engagement as an objective may result in outcomes that reflect these differences. When an objective that is chosen by the advertiser but is implemented by the platform results in discriminatory delivery, who bears the legal responsibility may be unclear. On one hand, the advertiser (perhaps, unknowingly or implicitly) requested the outcome, and if that choice created a discriminatory outcome, a prior legal analysis [14] suggests the platform may be protected under Section 230 of the Communications Decency Act, a U.S. law that provides ad platforms with immunity from content published by advertisers [62]. On the other hand, to the extent that a platform might be liable under Title VII, one may argue Section 230 does not provide immunity from such liability. We suggest that platforms should be aware that ad objectives that optimize for engagement may cause delivery algorithms to skew who receives a job ad; if it does, the platform may have the responsibility to prevent such skew or disable advertiser's choice of such objectives for employment ads in order to prevent discrimination. Our work does not advocate a position on the legal question, but provides data (§5.2) about outcomes that shows implications of choices of objectives.

In addition to the optimization objective, other confounding sources of skew (§2.1) may have implications for legal liability. The prior legal analysis of the Google's ad platform evaluated the applicability of Section 230 to different sources of skew, and argued Google may not be protected by Section 230 if a skew is fully a product of Google's algorithms [14]. Similarly, our goal is to design a methodology that controls for confounding factors and isolates skew that is enabled solely due to choices made by the platform's ad delivery algorithms.

3 BACKGROUND

We next highlight relevant details about the ad platforms to which we apply our methodology and discuss related work.

3.1 LinkedIn and Facebook Ad Platforms

We give details about LinkedIn's and Facebook's advertising platforms that are relevant to our methodology.

Ad objective: LinkedIn and Facebook advertisers purchase ads to meet different marketing *objectives.* As of February 2021, both LinkedIn and Facebook have three types of objectives: awareness, consideration and conversion, and each type has multiple additional options [20, 39]. For both platforms, the chosen objective constrains the ad format, bidding strategy and payment options available to the advertiser.

Ad audience: On both platforms, advertisers can target an audience using targeting attributes such as geographic location, age and gender. But if the advertiser discloses they are running a job ad, the platforms disable or limit targeting by age

[1] We have updated §2.2 after the original submission to WWW '21 to reflect post-camera-ready improvements to our understanding of the legal issues.

and gender [49]. LinkedIn, being a professional network, also provides targeting by job title, education, and job experience.

In addition, advertisers on both platforms can upload a list of known contacts to create a custom audience (called "Matched Audience" on LinkedIn and "Custom Audience" on Facebook). On LinkedIn, contacts can be specified by first and last name or e-mail address. Facebook allows specification by many more fields, such as zip code and phone number. The ad platforms then match the uploaded list with profile information from LinkedIn or Facebook accounts.

Ad performance report: Both LinkedIn and Facebook provide *ad performance reports* through their website interface and via their marketing APIs [21, 38]. These reports reflect near real-time campaign performance results such as the number of clicks and impressions the ad received, broken down along different axes. The categories of information along which aggregate breakdowns are available differ among platforms. Facebook reports breaks down performance data by location, age, and gender, while LinkedIn gives breakdowns by location, job title, industry and company, but not by age or gender.

3.2 Related Work

Targeted advertising has become ubiquitous, playing a significant role in shaping information and access to opportunities for hundreds of millions of users. Because the domains of employment, housing, and credit have legal anti-discrimination protections in the U.S. [11, 12, 60], the study of ad platform's role in shaping access and exposure to those opportunities has been of particular interest in civil rights discourse [32, 33] and research. We discuss such work next.

Discriminatory ad targeting: Several recent studies consider discrimination in ad targeting: journalists at ProPublica were among the first to show that Facebook's targeting options enabled job and housing advertisers to discriminate by age [6], race [5] and gender [57]. In response to these findings and as part of a settlement agreement to a legal challenge [1], Facebook has made changes to restrict the targeting capabilities offered to advertisers for ads in legally protected domains [22, 49]. Other ad platforms, e.g. Google, have announced similar restrictions [55]. The question of whether these restrictions are sufficient to stop an ill-intentioned advertiser from discrimination remains open, as studies have shown that advanced features of ad platforms, such as custom and lookalike audiences, can be used to run discriminatory ads [23, 51, 54, 64]. Our work assumes a well-intentioned advertiser and performs an audit study using gender-balanced targeting.

Discriminatory ad delivery: In addition to the targeting choices by advertisers, researchers have hypothesized that discriminatory outcomes can be a result of platform-driven choices. In 2013, Sweeney's empirical study found a statistically significant difference between the likelihood of seeing an ad suggestive of an arrest record on Google when searching for people's names assigned primarily to black babies compared to white babies [56]. Datta et al. [15] found that the gender of a Google account influences the number of ads one sees related to high-paying jobs, with female accounts seeing fewer such ads. Both studies could not examine the causes of such outcomes, as their methodology did not have an ability to isolate the role of the platform's algorithm from other possibly contributing factors, such as competition from advertisers and user activity. Gelauff et al. [24] provide an empirical study of the challenges of advertising to a demographically balanced ad audience without using micro-targeting and in the presence of ad delivery optimization. Lambrecht et al. [31] perform a field test promoting job opportunities in STEM using targeting that was intended to be gender-neutral, find that their ads were shown to more men than women, and explore potential explanations for this outcome. Finally, recent work by Ali and Sapiezynski et al. [2] has demonstrated that their job and housing ads placed on Facebook are delivered skewed by gender and race, even when the advertiser targets a gender- and race-balanced audience, and that this skew results from choices of the Facebook's ad delivery algorithm, and is not due to market or user interaction effects. AlgorithmWatch [28] replicate these findings with European user audiences, and add an investigation of Google's ad delivery for jobs. Our work is motivated by these studies, confirming results on Facebook and performing the first study we are aware of for LinkedIn. Going a step further to distinguish between skewed and discriminatory delivery, we propose a new methodology to control for user qualifications, a factor not accounted for in prior work, but that is critical for evaluating whether skewed delivery is, in fact, discriminatory, for job ads. We build on prior work exploring ways in which discrimination may arise in job-related advertising and assessing the legal liability of ad platforms [14], to establish that the job ad delivery algorithms of Facebook may be violating U.S. anti-discrimination law.

Auditing algorithms: The proprietary nature of ad platforms, algorithms, and their underlying data makes it difficult to definitively establish the role platforms and their algorithms play for creation of discriminatory outcomes [4, 8–10, 48]. For advertising, in addition to the previously described studies, recent efforts have explored the possibility of auditing with data provided by Facebook through its public Ad Library [53] (created in response to a legal settlement [1]). Other works have focused on approaches that rely on sock-puppet account creation [7, 34]. Our work uses only ad delivery statistics that platforms provide to regular advertisers. This approach makes us less reliant on the platform's willingness to be audited. We do not rely on transparency-data from platforms, since it is often limited and insufficient for answering questions about the platform's role in discrimination [41]. We also do not rely on an ability to create user accounts on the platform, since experimental accounts are labor-intensive to create and disallowed by most platform's policies. We build on prior work of external auditing [2, 3, 14, 15, 50, 66]. We show that auditing for discrimination in ad delivery of job ads is possible, even when limited to capabilities available to a regular advertiser, and that one can carefully control for confounding factors.

Auditing LinkedIn: To our knowledge, the only work that has studied LinkedIn's ad system's potential for discrimination is that of Venkatadri and Mislove [64]. Their work demonstrates that compositions of multiple targeting options together can result in targeting that is skewed by age and gender, without explicitly targeting using those attributes. They suggest mitigations should be based not on disallowing individual targeting parameters, but on the overall outcome of the targeting. We agree with this goal, and go beyond this prior work by basing our evaluation on the outcome of ad delivery, measuring delivery of real-world ads, and contrasting outcomes on LinkedIn with Facebook's.

LinkedIn has made efforts to integrate fairness metrics into some of its recommendation systems [25, 42]. Our work looks at a different product, its ad platform, for which, to our knowledge, LinkedIn has not made public claims about fairness-aware algorithms.

4 AUDITING METHODOLOGY

We next describe the methodology we propose to audit ad delivery algorithms for potential discrimination.

Our approach consists of three steps. First, we use the advertising platform's custom audience feature (§4.1) to build an audience that allows us to infer gender of the ad recipients for platforms that do not provide ad delivery statistics along gender lines. Second, we develop a novel methodology that controls for job qualifications by carefully selecting job categories (§4.2) for which everyone in the audience is equally qualified (or not qualified) for, yet for which there are distinctions in the real-world gender distributions of employees in the companies. We then run paired ads concurrently for each job category and use statistical tests to evaluate whether the ad delivery results are skewed (§4.3).

Our lack of access to users' profile data, interest or browsing activity prevents us from directly testing whether ad delivery satisfies metrics of fairness commonly used in the literature, such as *equality of opportunity* [26], or recently proposed for ad allocation tasks where users have diverse preferences over outcomes, such as preference-informed individual fairness [29]. In our context of job ads, equality of opportunity means that an individual in a demographic group that is qualified for a job should get a positive outcome (in our case: see an ad) at equal rates compared to an equally qualified individual in another demographic group. While our methodology does not test for this metric, we indirectly account for qualification in the way we select which job categories we run ads for.

We only describe a methodology for studying discrimination in ad delivery along gender lines, but we believe our methodology can be generalized to audit along other attributes such as race and age by an auditor with access to auxiliary data that is needed for picking appropriate job categories.

4.1 Targeted Audience Creation

Unlike Facebook, LinkedIn does not give a gender breakdown of ad impressions, but reports their location at the county level. As a workaround, we rely on an approach introduced in prior work [2, 3] that uses ad recipients' location to infer gender.

To construct our ad audience, we use North Carolina's voter record dataset [46], which among other fields includes each voter's name, zip code, county, gender, race and age. We divide all the counties in North Carolina into two halves. We construct our audience by including only male voters from counties in the first half, and only female voters from counties in the second half (this data is limited to a gender binary, so our research follows). If a person from the first half of the counties is reported as having seen an ad, we can infer that the person is a male, and vice versa. Furthermore, we include a roughly equal number of people from each gender in the targeting because we are interested in measuring skew that results from the delivery algorithm, not the advertiser's targeting choices.

Table 1: Audiences used in our study.

ID	Size	Males	Females	Match Rate
Aud #0	954,714	477,129	477,585	11.83%
Aud #1	900,000	450,000	450,000	11.6%
Aud #2	950,000	450,000	500,000	11.8%
Aud #0f	850,000	450,000	400,000	11.88%
Aud #1f	800,000	400,000	400,000	12.51%
Aud #2f	790,768	390,768	400,000	12.39%

To evaluate experimental reproducibility without introducing test-retest bias, we repeat our experiments across different, but equivalent audience partitions. Table 1 gives a summary of the partitions we used. Aud#0, Aud#1 and Aud#2 are partitions whose size is approximately a quarter of the full audience, while Aud#0f, Aud#1f and Aud#2f are constructed by swapping the choice of gender by county. Swapping genders this way doubles the number of partitions we can use.

On both LinkedIn and Facebook, the information we upload is used to find *exact matches* with information on user profiles. We upload our audience partitions to LinkedIn in the form of first and last names. For Facebook, we also include zip codes, because their tool for uploading audiences notified us that the match rate would be too low when building audiences only on the basis of first and last names. The final targeted ad audience is a subset of the audience we upload, because not all the names will be matched, i.e. will correspond to an actual user of a platform. As shown in Table 1, for each audience partition, close to 12% of the uploaded names were matched with accounts on LinkedIn. Facebook does not report the final match rates for our audiences in order to protect user privacy.

To avoid self-interference between our ads over the same audience we run paired ads concurrently, but ads for different job categories or for different objectives sequentially. In addition, to avoid test-retest bias, where a platform learns from prior experiments who is likely to respond and applies that to subsequent experiments, we generally use different (but equivalent) target audiences.

4.2 Controlling for Qualification

The main goal of our methodology is to distinguish skew resulting from algorithmic choices that are not related to qualifications, from skew that can be justified by differences in user qualifications for the jobs advertised. A novel aspect of our methodology is to *control* for qualifications by running paired ads for jobs with similar qualification requirements, but skewed de facto gender distributions. We measure skew by comparing the *relative difference* between the delivery of a *pair of ads* that run concurrently, targeting the same audience. Each test uses paired jobs that meet two criteria: First, they must have *similar qualification requirements*, thus ensuring that the people that we target our ads with are equally qualified (or not qualified) for both job ads. Second, the jobs must exhibit a *skewed, de facto gender distribution* in the real-world, as shown through auxiliary data. Since both jobs require similar qualifications, our assumption is that on a platform whose ad delivery algorithms are non-discriminatory, the distribution of genders among the recipients of the two ads will be roughly equal. On the other hand, in order to optimize for engagement or business objectives, platforms may incorporate other factors into ad delivery optimization, such as training or historical data. This data may reflect the de facto skew and thus influence machine-learning-based algorithmic predictions of engagement. Since such factors do not reflect differences in job qualifications, they may be disallowed (§2.2) and therefore represent platform-induced discrimination (even if they benefit engagement or the platform's business interests). We will look for evidence of such factors in a *difference* in gender distribution between the paired ads (see §4.4 for how we quantify the difference).

In §5.1, we use the above criteria to select three job categories – delivery driver, sales associate and software engineer – and run a pair of ads for each category and compare the gender make-up of the people to whom LinkedIn and Facebook show our ads. An example of such a pair of ads is a delivery driver job at Domino's (a pizza chain) and at Instacart (a grocery delivery service). The de facto gender distribution among drivers of these services is skewed male for Domino's and skewed female for Instacart [17, 52]. If a platform shows the Instacart ad to relatively more women than a Domino's ad, we conclude that the platform's algorithm is discriminatory, since both jobs have similar qualification requirements and thus a gender skew cannot be attributed to differences in qualifications across genders represented in the audience.

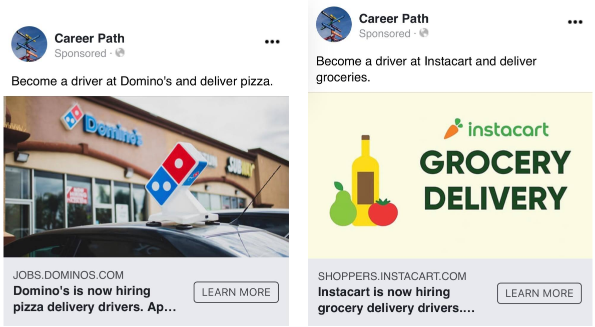
Using paired, concurrent ads that target the same audience also ensures other confounding factors such as timing or competition from other advertisers affect both ads equally [2].

To avoid bias due to the audience's willingness to move for a job, we select jobs in the same physical location. When possible (for delivery driver and sales job categories, but not software engineering), we select jobs in the location of our target audience.

4.3 Placing Ads and Collecting Results

We next describe the mechanics of placing ads on Facebook and LinkedIn, and collecting the ad delivery statistics which we use to calculate the gender breakdown of the audiences our ads were shown to. We also discuss the content and parameters we use for running our ads.

4.3.1 Ad Content. In creating our ads, we aim to use gender-neutral text and image so as to minimize any possible skew due to the input of an advertiser (us). The ad headline and description for each pair of ads is customized to each job category as described in §5.1. Each ad we run links to a real-world job opportunity that is listed on a job search site, pointing to a job posting on a company's careers page (for delivery driver) or to a job posting on LinkedIn.com (in other cases). Figure 1 shows screenshots of two ads from our experiments.



Become a driver at Domino's and deliver pizza.

JOBS.DOMINOS.COM
Domino's is now hiring pizza delivery drivers. Ap...

Become a driver at Instacart and deliver groceries.

SHOPPERS.INSTACART.COM
Instacart is now hiring grocery delivery drivers....

Figure 1: Example delivery driver job ads for Domino's and Instacart.

4.3.2 Ad Optimization Objective. We begin by using the *conversion* objective because searching for people who are likely to take an action on the job ad is a likely choice for advertisers seeking users who will apply for their job (§5.1). For LinkedIn ads, we use "Job Applicants" option, a conversion objective with the goal: "Your ads will be shown to those most likely to view or click on your job ads, getting more applicants." [39]. For Facebook ads, we use "Conversions" option with with the following optimization goal: "Encourage people to take a specific action on your business's site" [20], such as register on the site or submit a job application.

In §5.2, we run some of our Facebook ads using the *awareness* objective. By comparing the outcomes across the two objectives we can evaluate whether an advertiser's objective choice plays a role in the skew (§2.2). We use the "Reach" option that Facebook provides within the awareness objective with the stated goal of: "Show your ad to as many people as possible in your target audience" [20].

4.3.3 Other Campaign Parameters. We next list other parameters we use for running ads and our reasons for picking them.

From the ad format options available for the objectives we selected, we choose *single image ads*, which show up in a prominent part of LinkedIn and Facebook users' newsfeeds.

We run all Facebook and LinkedIn ads with a total budget of $50 per ad campaign and schedule them to run for a full day or until the full budget is exhausted. This price point ensures a reasonable sample size for statistical evaluations, with all of our ads receiving at least 340 impressions.

For both platforms, we request automated bidding to maximize the number of clicks (for the conversion objective) and impressions (for the awareness objective) our ads can get within the budget. We configure our campaigns on both platforms to pay per impression shown. On LinkedIn, this is the only available option for our chosen parameters. We use the same option on Facebook for consistency. On both platforms we disable audience expansion and off-site delivery options. While these options might show our ad to more users, they are not relevant or may interfere with our methodology.

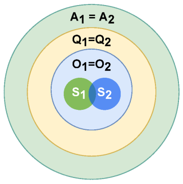


Figure 2: Relation between subsets of audiences involved in running two ads targeting the same audience. The subscripts indicate sets for the first and second ad.

Since our methodology for LinkedIn relies on using North Carolina county names as proxies for gender, we add "North Carolina" as the location for our target audience. We do the same for Facebook for consistency across experiments but we do not need to use location as a proxy to infer gender in Facebook's case.

4.3.4 Launching Ads and Collecting Delivery Statistics. For LinkedIn, we use its *Marketing Developer Platform* API to create the ads, and once the ads run, to get the final count of impressions per county which we use to infer gender. For Facebook, we create ads via its Ad Manager portal. The portal gives a breakdown of ad impressions by gender, so we do not rely on using county names as a proxy. We export the final gender breakdown after the ad completes running.

4.4 Skew Metric

We now describe the metric we apply to the *outcome* of advertising, i.e. the demographic make-up of the audience that saw our ads, to establish whether platform's ad delivery algorithm leads to discriminatory outcomes.

4.4.1 Metric: As discussed in the beginning of this section, out methodology works by running two ads simultaneously and looking at the relative difference in how they are delivered. In order to be able to effectively compare delivery of the two ads, we need to ensure the baseline audience that we use to measure skew is the same for both ads. The baseline we use is people who are qualified for the job we are advertising and are browsing the platform during the ad campaigns. However, we must consider several audience subsets shown in Figure 2: A, the the audience targeted by us, the advertiser (us); Q, the subset of A that the ad platform's algorithm considers qualified for the job being advertised, and O, the subset of Q that are online when the ads are run.

Our experiment design should ensure that these sets are the same for both ads, so that a possible skewed delivery cannot be merely explained by a difference the underlying factors these sets represent. We ensure A, Q, and O match for our jobs by targeting the same audience (same A), ensuring both jobs have similar qualification requirements (same Q) as discussed in §4.2, and by running the two ads at the same time (same O).

To *measure gender skew*, we compare what fraction of people in O that saw our two ads are a member of a specific gender. Possible unequal distribution of gender in the audience does not affect our comparison because it affects both ads equally (because O is the same for both ads). Let S_1 and S_2 denote subsets of people in O who saw the first and second ad, respectively. S_1 and S_2 are not necessarily disjoint sets. To measure gender skew, we compare the fraction of females in S_1 that saw the first ad ($s_{1,f}$) and fraction of females in S_2 that saw the second ad ($s_{2,f}$) with the fraction of females in O that were online during the ad campaign (o_f).

In the absence of discriminatory delivery, we expect, for both ads, the gender make-up of the audience the ad is shown to be representative of the gender make-up of people that were online and participated in ad auctions. Mathematically, we expect $s_{1,f} = o_f$ and $s_{2,f} = o_f$. As an external auditor that does not have access to users' browsing activities, we do not have a handle on o_f but we can directly compare $s_{1,f}$ and $s_{2,f}$. Because we ensure other factors that may affect ad delivery are either controlled or affect both ads equally, we can attribute any difference we might observe between $s_{1,f}$ and $s_{2,f}$ to choices made by the platform's ad delivery algorithm based on factors unrelated to qualification of users, such as revenue or engagement goals of the platform.

4.4.2 Statistical Significance: We use the Z-Test to measure the statistical significance of a difference in proportions we observe between $s_{1,f}$ and $s_{2,f}$. Our null hypothesis is that there is no gender-wise difference between the audiences that saw the two ads, i.e., $s_{1,f} = s_{2,f}$, evaluated as:

$$Z = \frac{s_{1,f} - s_{2,f}}{\sqrt{\hat{s}_f(1 - \hat{s}_f)(\frac{1}{n_1} + \frac{1}{n_2})}}$$

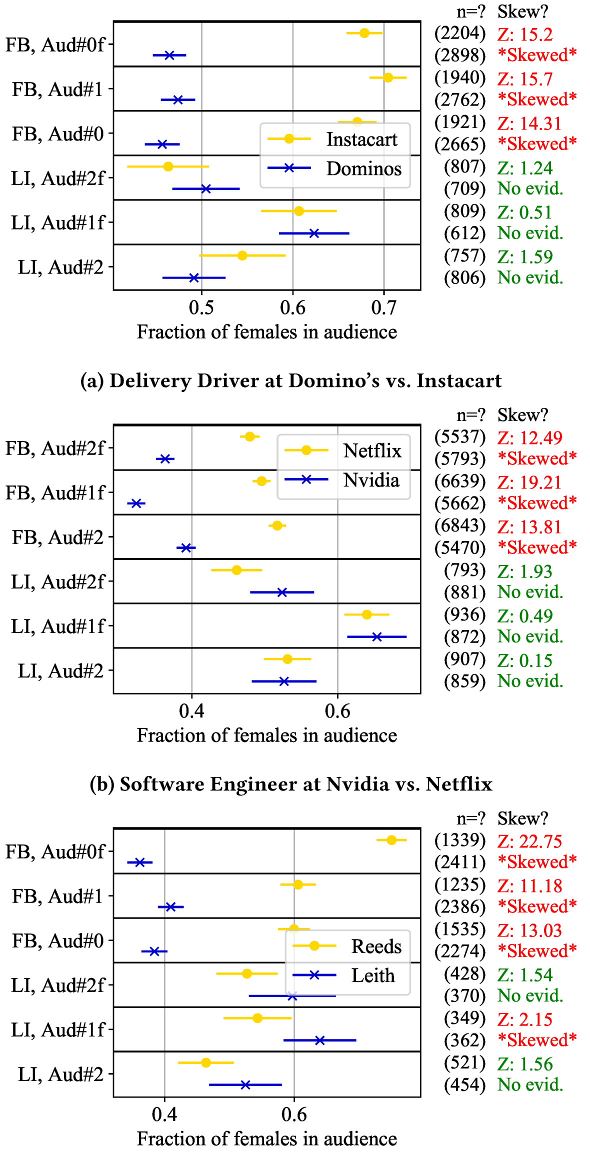
where $\hat{s}_f$ is fraction of females in S_1 and S_2 combined ($S_1 \cup S_2$), and n_1 and n_2 are the sizes of S_1 and S_2, respectively. At α significance level, if $Z > Z_\alpha$, we reject the null hypothesis and conclude that there is a statistically significant gender skew in the ad delivery. We use a 95% confidence level ($Z_\alpha = 1.96$) for all of our statistical tests. This test assumes the samples are independent and n is large. Only the platform knows whom it delivers the ad to, so only it can verify independence. Sample sizes vary by experiment, as shown in figures, but they always exceed 340 and often are several thousands.

4.5 Ethics

Our experiments are designed to consider ethical implications, minimizing harm both to the platforms and the individuals that interact with our ads. We minimize harm to the platforms by registering as an advertiser and interacting with the platform just like any other regular advertiser would. We follow their terms of service, use standard APIs available to any advertiser and do not collect any user data. We minimize harm to individuals using the platform and seeing our ads by having all our ads link to a real job opportunity as described. Finally, our ad audiences aim to include an approximately equal number of males and females and so aim not to discriminate. Our study was classified as exempt by our Institutional Review Board.

5 EXPERIMENTS

We next present the results from applying our methodology to real-world ads on Facebook and LinkedIn. We find contrasting results that show statistically significant evidence of skew that is not justifiable on the basis of qualification in the case of Facebook, but not in the case of LinkedIn. We make data for the ads we used in our experiments and their delivery statistics publicly available at [27]. We ran all ads in February, 2021.



(a) Delivery Driver at Domino's vs. Instacart

(b) Software Engineer at Nvidia vs. Netflix

(c) Sales Associate at Reeds Jewelry vs. Leith Automotive

Figure 3: Skew in delivery of real-world ads on Facebook (FB) and LinkedIn (LI), using "Conversion" objective. n gives total number of impressions. We use our metric (§4.4) to test for skew at 95% confidence level ($Z > 1.96$).

5.1 Measuring Skew in Real-world Ads

We follow the criteria discussed in §4.2 to pick and compare jobs which have similar qualification requirements but for which there is data that shows the de facto gender distribution is skewed. We study whether ad delivery optimization algorithms reproduce these de facto skews, even though they are not justifiable on the basis of differences in qualification.

We pick three job categories: a low-skilled job (delivery driver), a high-skilled job (software engineer), and a low-skilled but popular job among our ad audience (sales associate). Since our methodology compares two ads for each category, we select two job openings at companies for which we have evidence of de facto gender distribution differences, and use our metric §4.4 to measure whether there is a statistically significant gender skew in ad delivery. In each job category, we select pairs of jobs in the same state to avoid skew (§4.2).

For each experiment, we run the same pair of ads on both Facebook and LinkedIn and compare their delivery. For both platforms, we repeat the experiments on three different audience partitions for reproducibility. We run the ads for each job category at different times to avoid self-competition (§4.1). We run these first set of ads using the conversion objective (§4.3.2).

As discussed in §4.3.1, we build our ad creatives (text and image) using gender-neutral content to minimize any skew due to an advertiser's (our) input. For delivery driver and sales associate categories, Facebook ad text uses modified snippets of the real job descriptions they link to (for example, "Become a driver at Domino's and deliver pizza"). Images use a company's logo or a picture of its office. To ensure any potential skew is not due to keywords in the job descriptions that could appeal differently to different audiences, we ran the software engineering Facebook ads using generic headlines with a format similar to the ones shown in Figure 1, and found similar results to the runs that used modified snippets. All LinkedIn ads were ran using generic ad headlines similar to those in Figure 1.

5.1.1 *Delivery Drivers.*

We choose *delivery driver* as a job category to study because we were able to identify two companies – Domino's and Instacart – with significantly different de facto gender distributions among drivers, even though their job requirements are similar. 98% of delivery drivers for Domino's are male [17], whereas more than 50% of Instacart drivers are female [52]. We run ads for driver positions in North Carolina for both companies, and expect a platform whose ad delivery optimization goes beyond what is justifiable by qualification and reproduces de facto skews to show the Domino's ad to relatively more males than the Instacart ad.

Figure 3a shows gender skews in the results of ad runs for delivery drivers, giving the gender ratios of ad impressions with 95% confidence intervals. These results show evidence of a *statistically significant gender skew on Facebook*, and show *no gender skew on LinkedIn.* The skew we observe on Facebook is in the same direction as the de facto skew, with the Domino's ad delivered to a higher fraction of men than the Instacart ad. We confirm the results across three separate runs for both platforms, each time targeting a different audience partition.

5.1.2 *Software Engineers.*

We next consider the *software engineer* (SWE) job category, a high-skilled job which may be a better match for LinkedIn users than delivery driver jobs.

We pick two companies based employee demographics stated in their diversity report . Because we are running software engineering ads, we specifically look at the percentage of female employees who work in a tech-related position. We pick Netflix and Nvidia for our paired ad experiments. At Netflix, 35% of employees in tech-related positions are female [44] according to its 2021 report. At Nvidia, 19% of all employees are female according to [47], and third-party data as of 2020 suggests that the percentage of female employees in tech-related positions is as low as 14% [16]. For both companies, we find job openings in the San Francisco Area and run ads for those positions. We expect a platform whose algorithm learns and perpetuates the existing difference in employee demographics will show the Netflix ad to more women than the Nvidia ad.

Figure 3b shows the results. The *Facebook results show skew by gender in all three trials*, with a statistically different gender distribution between the delivery of the two ads. The skew is in the direction that confirms our hypothesis, a higher fraction of women seeing the Netflix ads than the Nvidia ads. *LinkedIn results are not skewed in all three trials.* These results confirm the presence of delivery skew not justified by qualifications on Facebook for a second, higher-skilled job category.

5.1.3 Sales Associates. We consider *sales associate* as a third job category. Using LinkedIn's audience estimation feature, we found that many LinkedIn users in the audience we use identified as having sales experience, so we believe people with experience in sales are well-represented in the audience. The Bureau of Labor Statistics (BLS) data shows that sales jobs skew by gender in different industries, with women filling 62% of sales associates in jewelry stores and only 17.9% in auto dealerships [58]. We pick Reeds Jewelers (a retail jeweler) and Leith Automotive (an auto dealership) to represent these two industries with open sales positions in North Carolina. If LinkedIn's or Facebook's delivery mimics skew in the de facto gender distribution, we expect them to deliver the Reeds ad to relatively more women than the Leith ad.

Figure 3c presents the results. All three trials on both platforms confirm our prior results using other job categories, with *statistically significant delivery skew between all jobs on Facebook but not for two of the three cases on LinkedIn.* One of the three trials on LinkedIn (Aud#1f) shows skew just above the threshold for a statistical significance, and surprisingly it shows bias in the opposite direction from expected (more women for the Reeds ad). We observe that these cases show the smallest response rates (349 to 521) and their Z-scores (1.54 to 2.15) are close to the threshold ($Z = 1.96$), while Facebook shows consistently large skew (11 or more).

5.1.4 Summary: These experiments confirm that our methodology proposed in §4.2 is feasible to implement in practice. Moreover, the observed outcomes are different among the two platforms. Facebook's job ad delivery is skewed by gender, even when the advertiser is targeting a gender-balanced audience, consistent with prior results of [2]. However, because our methodology controls for qualifications, our results imply that the skew cannot be explained by the ad delivery algorithm merely reflecting differences in qualifications. Thus, based on the discussion of legal liability in §2.2, our findings suggests that Facebook's algorithms may be responsible for unlawful discriminatory outcomes.

Our work provides the first analysis of LinkedIn's ad delivery algorithm. With the exception of one experiment, we did not find evidence of skew by gender introduced by LinkedIn's ad delivery, a negative result for our investigation, but perhaps a positive result for society.

5.2 "Reach" vs. "Conversion" Objectives

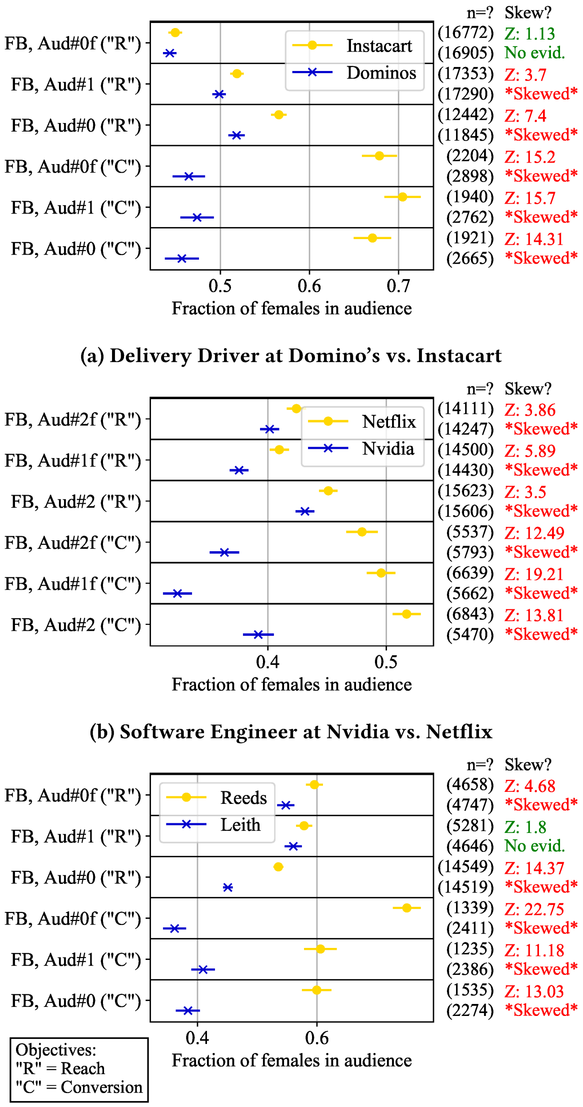
In §5.1, we used the *conversion* objective, assuming that this objective would be chosen by most employers running ads and aiming to maximize the number of job applicants. However, both LinkedIn and Facebook also offer advertisers the choice of the *reach* objective, aiming to increase the number of people reached with (or shown) the ad, rather than the number of people who apply for the job. We next examine how the use of the *reach* objective affects skew in ad delivery on Facebook, compared to the use of the *conversion* objective. We focus on Facebook because we observed evidence of skew that cannot be explained by differences in qualifications in their case, and we are interested in exploring whether that skew remains even with a more "neutral" objective. While there may be a debate about allocating responsibility for discrimination between advertiser and platform when using a conversion objective (see §2.2), we believe that the responsibility for any discrimination observed when the advertiser-selected objective is *reach* rests on the platform.

We follow our prior approach (§5.1) with one change: we use *reach* as an objective and compare with the prior results that used the *conversion* objective. The job categories and other parameters remain the same and we repeat the experiments on different audience partitions for reproducibility.

Figure 4a, Figure 4b and Figure 4c show the delivery of *reach* ads for the delivery driver, software engineer and sales associate jobs, respectively. For comparison, the figures include the prior Facebook experiments ran using *conversion* objective (from Figure 3). For all three job categories, the results show a statistically significant skew in at least two out of the three experiments using the reach objective. This result confirms our result in §5.1 that showed Facebook's ad delivery algorithm introduces gender skew even when advertiser targets a gender-balanced audience. Since skewed delivery occurs even when the advertiser chooses the *reach* objective, the skew is attributable to the platform's algorithmic choices and not to the advertiser's choice.

On the other hand, we notice two main differences in the delivery of the ads run with the *reach* objective. For all three job categories (Figure 4a, Figure 4b and Figure 4c) the gap between gender delivery for each pair of ads is reduced for the *reach* ads compared to the *conversion* ads. And, for two of the job categories (delivery driver and sales associate), one of the three cases does not show a statistically significant evidence for skew, while all three showed such evidence in the *conversion* ads. These observations indicate that the degree of skew may be reduced when using the *reach* objective, and, therefore, an advertiser's request for the *conversion* objective may increase the amount of skew because, according to Facebook's algorithmic predictions, conversions may correlate with particular gender choices for certain jobs.

Revisiting our discussion of the legal responsibility for discrimination (§2.2) in light of these results, the fact that both



(a) Delivery Driver at Domino's vs. Instacart

(b) Software Engineer at Nvidia vs. Netflix

(c) Sales Associate at Reeds Jewelry vs. Leith Automotive

Figure 4: Comparison of ad delivery with "reach" and "conversion" objectives on Facebook.

the advertiser and the platform make choices about the ad recipients can blur who is legally responsible. If the discriminatory outcome occurs regardless of the advertiser-chosen objective, as our results with the *reach* objective underscore, then it's clear it is the responsibility of the platform. On the other hand, if we saw skew with advertiser-specified objectives that optimize for engagement and not others (which was not the case in our experiments), the platform may claim it is just doing what the advertiser requested, and may even state that the blame (or legal culpability) for any skew therefore rests on the advertiser. However, even in this case, one could argue that it is the ad platform that has full control over determining how the optimization algorithm actually works and what its inputs are. Therefore, if an advertiser discloses that they are running job ads (which we did in our experiments), the ad platform may still have the ethical and legal responsibility to ensure its algorithm does not produce a discriminatory outcome regardless of the advertiser objective it is optimizing for.

Table 2: Breakdown of impressions for LinkedIn ads run using Aud#2. "Unreported" shows percentages in unreported counties (whose genders we cannot infer).

Company	Total	Males	Females	Unreported (%)
Domino's	806	241	233	41.19
Instacart	757	194	232	43.73
Nvidia	859	232	258	42.96
Netflix	907	240	272	43.55
Leith	454	145	160	32.82
Reeds	521	192	166	31.29

6 FUTURE WORK

We next discuss the limitations of our study, give some directions for future study and, motivated by the challenges we faced in our work, provide recommendations as to what ad platforms can do to make auditing more feasible and accurate.

6.1 Limitations and Further Directions

Our experiments focus on skew from gender, but we believe our methodology can be used to study other attributes such as age or race. It requires the auditor having access to data about age and gender distributions among employees of different companies in the same category, so that the auditor can pick job ads that fit the criteria of our methodology. It also requires the ability to create audiences whose age and race distributions are known. The voter dataset we use includes age and race, so can be adapted to test for discrimination along those attributes.

Like prior studies, we use physical location as a proxy to infer the gender of the ad recipient, an approach which has some limitations. LinkedIn hides location when there are two or fewer ad recipients, so our estimates may be off in those areas. These cases account for 31-43% of our ad recipients, as shown in Table 2. Assuming gender distribution is uniform by county in North Carolina's population, we reason that these unreported cases do not significantly distort our conclusions.

We tested three job categories, with three experiment repetitions each. Additional categories and repetitions would improve confidence in our results. Although we found it difficult to select job categories with documented gender bias that we could target, such data is available in private datasets. Another question worth investigating with regards to picking job categories is whether delivery optimization algorithms are the same for all job categories, i.e., whether relatively more optimization happens for high-paying or scarce jobs.

Some advertisers will wish to target their ads by profession or background. We did not evaluate such targeting because our population data is not rich and large enough to support such comparisons with statistical rigor. Evaluation of this question would be future work, especially if the auditor has access to richer population data.

6.2 Recommendations

Prior work has shown that platforms are not consistent when self-policing their algorithms for undesired societal consequences, perhaps because the platforms' business objectives are at stake. Therefore, we believe independent (third party) auditing fills an important role. We suggest recommendations to make such external auditing of ad delivery algorithms more accessible, accurate and efficient, especially for public interest researchers and journalists.

Providing more targeting and delivery statistics: First, echoing sentiments from prior academic and activism work [2, 41], we note the value of surfacing additional ad targeting and delivery data in a privacy-preserving way. Public interest auditors often rely on features that the ad platforms make available for any regular advertiser to conduct their studies, which can make performing certain types of audits challenging. For example, in the case of LinkedIn, the ad performance report does not contain a breakdown of ad impressions by gender or age. To overcome such challenges, prior audit studies and our work rely on finding workarounds such as proxies to measure ad delivery along sensitive demographic features. On one hand, providing additional ad delivery statistics could help expend the scope of auditors' investigations. On the other hand, there may be an inherent trade-off between providing additional statistics about ad targeting and delivery and the privacy of users (see e.g. [23, 30]) or business interests of advertisers. We believe that privacy-preserving techniques, such as differentially private data publishing [19] may be able to strike a balance between auditability and privacy, and could be a fruitful direction for future work and practical implementation in the ad delivery context.

It is also worth asking what additional functionalities or insights about its data or ad delivery optimization algorithms the platforms can or should provide which would allow for more accessible auditing without sacrificing independence of the audits. Recent work has explored finding a balance between independence and addressing the challenges of external auditing by suggesting a *cooperative audit* framework [65], where the target platform is aware of the audit and gives the auditor special access but there are certain protocols in place to ensure the auditor's independence. In the context of ad platforms, we recognize that providing a special access option for auditors may open a path for abuse where advertisers may pretend to be an auditor for their economic or competitive benefit.

Replacing ad-hoc privacy techniques: Our other recommendation is for ad platforms to replace ad-hoc techniques they use as a privacy enhancement with more rigorous approaches. For example, LinkedIn gives only a rough estimate of audience sizes, and does not give the sizes if less than 300 [36] It also does not give the number of impressions by location if the count per county is less than three [35].

Such ad-hoc approaches have two main problems. First, it is not clear based on prior work on the ad platforms how effective they are in terms of protecting privacy of users [63, 64]. We were also able to circumvent the 300-minimum limit for audience size estimates on LinkedIn with repeated queries by composing one targeting parameter with another, then repeating a decomposed query and calculating the difference. More generally, numerous studies show ad-hoc approaches often fail to provide the privacy that they promise [13, 43]. Second, ad-hoc approaches can distorts statistical tests that auditors perform [45]. Therefore, we recommend ad platforms use approaches with rigorous privacy guarantees, and whose impact on statistical validity can be precisely analyzed, such as differentially private algorithms [19], where possible.

Reducing cost of auditing: Auditing ad platforms via black-box techniques incurs a substantial cost of money, effort, and time. Our work alone required several months of research on data collection and methodology design, and cost close to $5K to perform the experiments by running ads. A prior study of the impact of Facebook's ad delivery algorithms on political discourse cost up to $13K [3]. These costs quickly accumulate if one is to repeat experiments to study trends, increase statistical confidence, or reproduce results. One possible solution is to provide a discount for auditors. They would have similar access to the platform features like any other advertiser but would pay less to run ads. However, as with other designed-auditor techniques, this approach risks abuse.

Overall, making auditing ad delivery systems more feasible to a broader range of interested parties can help ensure that the systems that shape job opportunities people see operate in a fair manner that does not violate anti-discrimination laws. The platforms may not currently have the incentives to make the changes proposed and, in some cases, may actively block transparency efforts initiated by researchers and journalists [40]; thus, they may need to be mandated by law.

7 CONCLUSION

We study gender bias in the delivery of job ads due to platform's optimization choices, extending existing methodology to account for the role of qualifications in addition to the other confounding factors studied in prior work. We are the first to methodologically address the challenge of controlling for qualification, and also draw attention to how qualification may be used as a legal defense against liability under applicable laws. We apply our methodology to both Facebook and LinkedIn and show that our proposed methodology is applicable to multiple platforms and can identify distinctions between their ad delivery practices. We also provide the first analysis of LinkedIn for potential skew in ad delivery. We confirm that Facebook's ad delivery can result in skew of job ad delivery by gender beyond what can be legally justified by possible differences in qualifications, thus strengthening the previously raised arguments that Facebook's ad delivery algorithms may be in violation of anti-discrimination laws [2, 14]. We do not find such skew on LinkedIn. Our approach provides a novel example of feasibility of auditing algorithmic systems in a black-box manner, using only the capabilities available to all users of the system. At the same time, the challenges we encounter lead us to suggest changes that ad platforms could make (or that should be mandated of them) to make external auditing of their performance in societally impactful areas easier.

ACKNOWLEDGEMENTS

This work was funded in part by NSF grants CNS-1755992, CNS-1916153, CNS-1943584, CNS-1956435, and CNS-1925737. We are grateful to Aaron Rieke for feedback on the manuscript.

REFERENCES

[1] ACLU. Facebook EEOC complaints. https://www.aclu.org/cases/facebook-eeoc-complaints?redirect=node/70165.

[2] Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., and Rieke, A. Discrimination through optimization: How facebook's ad delivery can lead to biased outcomes. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing* (2019).

[3] Ali, M., Sapiezynski, P., Korolova, A., Mislove, A., and Rieke, A. Ad delivery algorithms: The hidden arbiters of political messaging. In *14th ACM International Conference on Web Search and Data Mining* (2021).

[4] Andrus, M., Spitzer, E., Brown, J., and Xiang, A. "What We Can't Measure, We Can't Understand": Challenges to demographic data procurement in the pursuit of fairness. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT)* (2021).

[5] Angwin, J., and Paris Jr., T. Facebook lets advertisers exclude users by race – ProPublica. https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race, October 26, 2016.

[6] Angwin, J., Scheiber, N., and Tobin, A. Dozens of companies are using Facebook to exclude older workers from job ads – ProPublica. https://www.propublica.org/article/facebook-ads-age-discrimination-targeting, December 20, 2017.

[7] Asplund, J., Eslami, M., Sundaram, H., Sandvig, C., and Karahalios, K. Auditing race and gender discrimination in online housing markets. In *Proceedings of the International AAAI Conf. on Web and Social Media* (2020).

[8] Barocas, S., and Selbst, A. D. Big data's disparate impact. *California Law Review 104*, 3 (2016), 671–732.

[9] Bogen, M., and Rieke, A. Help wanted: an examination of hiring algorithms, equity, and bias. *Technical report, Upturn* (2018).

[10] Bogen, M., Rieke, A., and Ahmed, S. Awareness in practice: tensions in access to sensitive attribute data for antidiscrimination. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (2020).

[11] CFR. 12 CFR section 202.4 (b)—discouragement. https://www.law.cornell.edu/cfr/text/12/202.4.

[12] CFR. 24 CFR section 100.75—discriminatory advertisements, statements and notices. https://www.law.cornell.edu/cfr/text/24/100.75.

[13] Cohen, A., and Nissim, K. Linear program reconstruction in practice. *Journal of Privacy and Confidentiality 10*, 1 (2020).

[14] Datta, A., Datta, A., Makagon, J., Mulligan, D. K., and Tschantz, M. C. Discrimination in online personalization: A multidisciplinary inquiry. *FAT* (2018).

[15] Datta, A., Tschantz, M. C., and Datta, A. Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 1 (2015).

[16] DiversityReports.org. Diversity reports - Nvidia. https://www.diversityreports.org/company-information/nvidia, 2020. Last accessed on Feb 28, 2021.

[17] Dominos. Gender pay gap report 2018. https://investors.dominos.co.uk/sites/default/files/attachments/dominos-corporate-stores-sheermans-limited-gender-pay-gap-2018-report.pdf, 2018. Last accessed on October 6, 2020.

[18] Dwork, C., and Ilvento, C. Fairness Under Composition. In *10th Innovations in Theoretical Computer Science Conference (ITCS)* (2019).

[19] Dwork, C., and Roth, A. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science* (2014).

[20] Facebook. Choose the right objective. https://www.facebook.com/business/help/1438417719786914.

[21] Facebook. Marketing API—Facebook for developers. https://developers.facebook.com/docs/marketing-apis/.

[22] Facebook. Simplifying targeting categories. https://www.facebook.com/business/news/update-to-facebook-ads-targeting-categories/, 2020.

[23] Faizullabhoy, I., and Korolova, A. Facebook's advertising platform: New attack vectors and the need for interventions. In *IEEE Workshop on Technology and Consumer Protection (ConPro)* (2018).

[24] Gelauff, L., Goel, A., Munagala, K., and Yandamuri, S. Advertising for demographically fair outcomes. *arXiv preprint arXiv:2006.03983* (2020).

[25] Geyik, S. C., Ambler, S., and Kenthapadi, K. Fairness-aware ranking in search and recommendation systems with application to linkedin talent search. In *Proc. 25th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining* (2019).

[26] Hardt, M., Price, E., and Srebro, N. Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems* (2016).

[27] Imana, B., Korolova, A., and Heidemann, J. Dataset of content and delivery statistics of ads used in "Auditing for discrimination in algorithms delivering job ads". https://ant.isi.edu/datasets/addelivery/.

[28] Kayser-Bril, N. Automated discrimination: Facebook uses gross stereotypes to optimize ad delivery. https://algorithmwatch.org/en/story/automated-discrimination-facebook-google/, October 18, 2020.

[29] Kim, M. P., Korolova, A., Rothblum, G. N., and Yona, G. Preference-informed fairness. In *Innovations in Theoretical Computer Science* (2020).

[30] Korolova, A. Privacy violations using microtargeted ads: A case study. *Journal of Privacy and Confidentiality 3*, 1 (2011), 27–49.

[31] Lambrecht, A., and Tucker, C. Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management Science 65*, 7 (2019), 2966–2981.

[32] Laura Murphy and Associates. Facebook's civil rights audit – progress report. https://about.fb.com/wp-content/uploads/2019/06/civilrightaudit_final.pdf, June 30, 2019.

[33] Laura Murphy and Associates. Facebook's civil rights audit – Final report. https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf, July 8 2020.

[34] Lecuyer, M., Spahn, R., Spiliopolous, Y., Chaintreau, A., Geambasu, R., and Hsu, D. Sunlight: Fine-grained targeting detection at scale with statistical confidence. In *CCS* (2015).

[35] LinkedIn. Ads Reporting. https://docs.microsoft.com/en-us/linkedin/marketing/integrations/ads-reporting/ads-reporting.

[36] LinkedIn. Audience Counts. https://docs.microsoft.com/en-us/linkedin/marketing/integrations/ads/advertising-targeting/audience-counts.

[37] LinkedIn. Campaign quality scores for sponsored content. https://www.linkedin.com/help/lms/answer/85406.

[38] LinkedIn. LinkedIn marketing developer platform. https://docs.microsoft.com/en-us/linkedin/marketing/.

[39] LinkedIn. Select a marketing objective for your ad campaign. https://www.linkedin.com/help/lms/answer/94698/select-a-marketing-objective-for-your-ad-campaign.

[40] Merrill, J. B., and Tobin, A. Facebook moves to block ad transparency tools – including ours. https://www.propublica.org/article/facebook-blocks-ad-transparency-tools, January 28, 2019.

[41] Mozilla. Facebook's ad archive API is inadequate. https://blog.mozilla.org/blog/2019/04/29/facebooks-ad-archive-api-is-inadequate/, 2019.

[42] Nandy, P., Diciccio, C., Venugopalan, D., Logan, H., Basu, K., and Karoui, N. E. Achieving fairness via post-processing in web-scale recommender systems. *arXiv preprint arXiv:2006.11350v2* (2021).

[43] Narayanan, A., and Shmatikov, V. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy* (2008).

[44] Netflix. Inclusion takes root at Netflix: Our first report. https://about.netflix.com/en/news/netflix-inclusion-report-2021, 2021.

[45] Nissim, K., Steinke, T., Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., O'Brien, D. R., and Vadhan, S. Differential privacy: A primer for a non-technical audience. *Vand. J. Ent. & Tech. L. 21* (2018).

[46] North Carolina State Board of Elections. Voter history data. https://dl.ncsbe.gov/index.html. Downloaded on April 23, 2020.

[47] Nvidia. Global diversity and inclusion report. https://www.nvidia.com/en-us/about-nvidia/careers/diversity-and-inclusion/, 2021. Last accessed on Feb 28, 2021.

[48] Reisman, D., Schultz, J., Crawford, K., and Whittaker, M. Algorithmic impact assessments: A practical framework for public agency accountability. *AI Now* (2018).

[49] Sandberg, S. Doing more to protect against discrimination in housing, employment and credit advertising. https://about.fb.com/news/2019/03/protecting-against-discrimination-in-ads/, March 19, 2019.

[50] Sandvig, C., Hamilton, K., Karahalios, K., and Langbort, C. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry 22* (2014), 4349–4357.

[51] Sapiezynski, P., Ghosh, A., Kaplan, L., Mislove, A., and Rieke, A. Algorithms that "don't see color": Comparing biases in lookalike and special ad audiences. *arXiv preprint arXiv:1912.07579* (2019).

[52] Selyukh, A. Why suburban moms are delivering your groceries. NPR https://www.npr.org/2019/05/25/722811953/why-suburban-moms-are-delivering-your-groceries, May 25, 2019.

[53] Shukla, S. A better way to learn about ads on facebook. https://about.fb.com/news/2019/03/a-better-way-to-learn-about-ads/, March 28 2019.

[54] Speicher, T., Ali, M., Venkatadri, G., Ribeiro, F. N., Arvanitakis, G., Benevenuto, F., Gummadi, K. P., Loiseau, P., and Mislove, A. Potential for discrimination in online targeted advertising. In *Proceedings of Machine Learning Research* (2018), S. A. Friedler and C. Wilson, Eds.

[55] Spencer, S. Upcoming update to housing, employment, and credit advertising policies. https://www.blog.google/technology/ads/upcoming-update-housing-employment-and-credit-advertising-policies/, 2020.

[56] Sweeney, L. Discrimination in online ad delivery: Google ads, black names and white names, racial discrimination, and click advertising. *Queue* (2013).

[57] Tobin, A., and Merrill, J. B. Facebook is letting job advertisers target only men – ProPublica. https://www.propublica.org/article/facebook-is-letting-job-advertisers-target-only-men, September 18, 2018.

[58] U.S. Bureau of Labor Statistics. Employed persons by detailed industry, sex, race, and Hispanic or Latino ethnicity. https://www.bls.gov/cps/cpsaat18.pdf, 2018.

[59] U.S. Equal Employment Opportunity Commission. Prohibited employment policies/practices. https://www.eeoc.gov/prohibited-employment-policiespractices.

[60] USC. 29 USC section 623—prohibition of age discrimination. https://www.law.cornell.edu/uscode/text/29/623.

[61] USC. 42 USC section 2000e-3—other unlawful employment practices. https://www.law.cornell.edu/uscode/text/42/2000e-3.

[62] USC. 47 USC section 230—protection for private blocking and screening of offensive material. https://www.law.cornell.edu/uscode/text/47/230.

[63] Venkatadri, G., Andreou, A., Liu, Y., Mislove, A., Gummadi, K. P., Loiseau, P., and Goga, O. Privacy risks with Facebook's PII-based targeting: Auditing a data broker's advertising interface. In *IEEE Symposium on Security and Privacy (SP)* (2018).

[64] Venkatadri, G., and Mislove, A. On the Potential for Discrimination via Composition. In *Internet Measurement Conference (IMC'20)* (2020).

[65] Wilson, C., Ghosh, A., Jiang, S., Mislove, A., Baker, L., Szary, J., Trindel, K., and Polli, F. Building and auditing fair algorithms: A case study in candidate screening. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT)* (2021).

[66] Zhang, J., and Bareinboim, E. Fairness in decision-making - the causal explanation formula. In *Association for the Advancement of Artificial Intelligence* (2018).