Auditing for Racial Discrimination in the Delivery of Education Ads

Basileal Imana
imana@princeton.edu
Center for Information Technology Policy, Princeton University
Princeton, New Jersey, USA

Aleksandra Korolova
korolova@princeton.edu
Department of Computer Science and School of Public and International Affairs, Princeton University
Princeton, New Jersey, USA

John Heidemann
johnh@isi.edu
Information Sciences Institute, University of Southern California
Los Angeles, California, USA

ABSTRACT

Digital ads on social-media platforms play an important role in shaping access to economic opportunities. Our work proposes and implements a new third-party auditing method that can evaluate racial bias in the *delivery of ads for education opportunities*. Third-party auditing is important because it allows external parties to demonstrate presence or absence of bias in social-media algorithms. Education is a domain with legal protections against discrimination and concerns of racial-targeting, but bias induced by ad delivery algorithms has not been previously explored in this domain. Prior audits demonstrated discrimination in platforms' delivery of ads to users for housing and employment ads. These audit findings supported legal action that prompted Meta to change their ad-delivery algorithms to reduce bias, but only in the domains of housing, employment, and credit. In this work, we propose a new methodology that allows us to measure racial discrimination in a platform's ad delivery algorithms for education ads. We apply our method to Meta using ads for real schools and observe the results of delivery. We find evidence of racial discrimination in Meta's algorithmic delivery of ads for education opportunities, posing legal and ethical concerns. Our results extend evidence of algorithmic discrimination to the education domain, showing that current bias mitigation mechanisms are narrow in scope, and suggesting a broader role for third-party auditing of social media in areas where ensuring non-discrimination is important.

CCS CONCEPTS

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

KEYWORDS

algorithmic auditing, targeted advertising, ad delivery, education ads, racial discrimination

ACM Reference Format:
Basileal Imana, Aleksandra Korolova, and John Heidemann. 2024. Auditing for Racial Discrimination in the Delivery of Education Ads. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3630106.3659041

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil

ACM ISBN 979-8-4007-0450-5/24/06
https://doi.org/10.1145/3630106.3659041

1 INTRODUCTION

Social-media platforms are a key method of advertising, and with the rapid growth of personalized advertising enabled by them, the platforms play a significant role in shaping access to opportunities. This role has been subject to increased scrutiny through academic research [1, 2, 27, 32, 33], civil rights audits [45], and regulation in the U.S. and E.U. [47].

Evidence of discrimination in how social-media platforms shape targeting and delivery of ads to users has been growing in recent years. Initial reports showed that ad *targeting* options, such as demographic attributes, could be used to discriminate [4, 56, 59], leading platforms to limit the targeting options they make available for housing, employment and credit (HEC) ads [53, 57], each an economic opportunity with prior government oversight. While removal of options for demographic targeting prevents directed discrimination, subsequent audits showed discrimination can occur at ad *delivery* stage, the algorithmic process by which a platforms decide when and which ads to display to users, and how much to charge the advertisers for them. Specifically, Meta's ad delivery algorithms were shown to be discriminatory by race and gender in the delivery of housing and employment ads, even when an advertiser targeted a demographically balanced audience [2, 27, 32]. Follow-up studies showed ad delivery algorithms play an active role in shaping access to information on politics [3] and climate change [54] in a way that is not transparent to both advertisers and end-users.

In response to these findings and as part of a legal settlement with the U.S. Department of Justice, in 2023 Meta deployed a Variance Reduction System (VRS) to reduce bias in delivery of housing ads [44, 58], and subsequently for employment and credit ads [6]. VRS's goal is to ensure the fraction of an HEC ad's impressions allocated to users of a particular gender (resp. race) does not deviate too much from the fraction of impressions allocated to users of that gender (resp. race) among all ads shown to them over the last 30 days [58]. Meta achieves this goal by making adjustments to the bidding strategy Meta executes on behalf of the advertiser.

Although the progress towards de-biasing delivery of HEC ads is promising, concerns remain that delivery of other types of opportunities advertised on social-media platforms may be discriminatory. In particular, in the U.S., the potential for ad delivery discrimination is a concern in domains such as insurance, education, healthcare, or other "public accommodations", a legal term that encompasses various types of businesses that provide service to the general public [36, 62]. No methodology is known to audit ad delivery in those domains. Developing new methods can inform applications of existing anti-discrimination law in these domains, similar to how prior audits have influenced changes to the systems and informed guidance on applications of Fair Housing Act to digital platforms [51].

Challenges for Auditing Ad Delivery: Auditing for discrimination in ad delivery is challenging for three reasons. First, the algorithms that select ads to be displayed and choose the prices to charge for them are opaque to advertisers and users. The *ad delivery algorithms* used for these choices are a closely guarded "secret sauce" of each platform, as they are central to platform monetization. These algorithms often use machine-learning-based predictions to quantify how "relevant" an ad is to the user [37, 39, 65] and how likely a user is to be valuable to the advertiser [44]. Furthermore, the algorithms often decide on the advertiser's behalf how much money to spend on trying to get their ad in front of the particular user [34, 44]. The concern is that these algorithms, designed to achieve the platform's long-term business goals, may propagate or exacerbate historical biases, or may learn biases from training data.

A second challenge is the limited access auditors have to data crucial to examine how platforms' algorithms work and what results they produce. The state-of-the-art in auditing ad delivery algorithms takes a black-box approach, utilizing only the limited features available to advertisers [27, 28]. These approaches rely on creating target ad audiences with a specific demographic composition, a challenging process that often requires auxiliary data sources to provide demographic attributes. Additionally, in conventional evaluations for algorithmic fairness, the auditor has access to individual- or group-level data, including scores assigned by the algorithms to those individuals. However, black-box auditors of ad delivery algorithms have access only to aggregated reports from the platform about who is reached. These reports distance the auditor from the evaluation through several levels of indirection, since they do not identify demographics of interest such as race, nor do they provide data on the relevance or value scores assigned by the platform's algorithms they are trying to evaluate.

Thirdly, the use of a black-box evaluation of platform algorithms must control for confounding factors that may affect ad delivery, such as temporal effects and market forces. For example, an observation of an ad's delivery to relatively larger number of White individuals than Black individuals does not necessarily imply discrimination in the ad delivery algorithm against Black users. Instead, it could be attributed to higher number of White individuals using the platform during the observation period. Another reason for differences in delivery could be competition with other advertisers which have larger budgets for some racial groups [13, 34]. The work of Ali and Sapiezynski et al. was the first to control for such factors by running *paired ads* that are equally affected by the confounding factors, and evaluating *relative difference* in the delivery of the ads for different demographic groups [2]. A subsequent study considered differences in job qualification as a relevant confounding factor for evaluating whether bias in delivery of employment ads constitutes discrimination, and proposed a method to control for it using a particular type of paired ads [27].

Why Education? In this work, we explore *education* as a new domain for which to study the potential role of ad delivery for discrimination. Avoiding discrimination in the delivery of educational opportunities matters because education has long-term impact on the career and financial well-being of individuals [16]. One specific concern in U.S. higher education is the potential for for-profit colleges to produce poor financial outcomes for their graduates [38]. Such colleges often devote considerable resources to advertising, targeting racial minorities disproportionately [31]. Prior work has not studied whether ad delivery algorithms propagate historical racial targeting biases in such colleges, a unique aspect of the education domain. In addition, education is an important category because of advertiser large spending in this domain. According to Meta's whitepaper ([64], Figure 4(a)), education is one of the top eight verticals among its advertisers, significantly ahead of verticals such as political advertising which have received much scrutiny.

Why Disparate Impact? We specifically consider *disparate impact* in education ad delivery because it is an outcome-based legal doctrine for assessing discrimination that is agnostic to the reasons for such outcome [7, 11]. Under this legal doctrine, disparate impact occurs if the outcome of ad delivery differs significantly by a demographic attribute such as race. The burden for understanding and justifying the source of the bias as legally permissible shifts to the platform. We expand on the implications of our work for legal liability in §4.2.

Our Contributions: Our first contribution is a new method for testing for the presence of racial discrimination in how platforms deliver *education ads* (§2). Inspired by a methodology developed in prior work for testing for gender discrimination in job ad delivery [27], we extend this approach racial discrimination in education ads by identifying colleges with different historical recruitment by race. By pairing ads for a for-profit and a public college, then looking at relative differences in ad delivery by race, we can isolate the role of platform's algorithm from other confounding factors. The insight in our method is to employ a pair of education ads for seemingly equivalent opportunities to users, but with one of the opportunities, in fact, tied to a historical racial disparity that ad delivery algorithms may propagate. If the ad for the for-profit college is shown to relatively more Black users than the ad for the public college, we can conclude that the algorithmic choices of the platform are racially discriminatory.

Our second contribution is to apply our method to Meta, where *we detect racial discrimination in the delivery of education ads* (§3). We first evaluate the platform's algorithm using neutral ad creatives that control for confounding factors. Our experiments with neutral creatives demonstrate that Meta's ad delivery algorithm shows ads for for-profit schools to relatively more Black users, and the difference is statistically significant for two out of three pairs of schools we study. In additional experiments, we show that when we use realistic ad creatives that the schools use in practice, the racial skew in delivery is *increased* (§3.2). These results provide strong evidence that Meta's algorithms shape the racial demographics of who sees which education opportunities, providing new evidence of Meta's potential long-term impact on careers and the financial well-being of individuals. Our results also open questions of Meta's legal liability under the doctrine of disparate impact of discrimination [7, 11].

Finally, we use our methodology to investigate whether Meta's algorithm steers delivery of ads for schools with historical predatory marketing practices disproportionately towards Black individuals. Over-delivery of ads for such schools could disproportionately harm students who are unaware of their documented risks to attendees [12, 31]. We pick ads for three for-profit schools that were subject to legal action by the U.S. Department of Education due to their predatory practices [50], and evaluate the racial difference in delivery of their ads compared to delivery of ads for public schools. We find that Meta shows the ads for the schools with historical predatory practices to relatively more Black individuals. Our results show that it is not sufficient for schools to target their ads equitably, but that ad platforms also need to ensure their algorithms are not introducing biases along legally protected characteristics such as race.

Our findings of discrimination in delivery of ads for real education opportunities show the issue of algorithmic discrimination is not limited to housing, employment, and credit opportunities and raises a broader question of whether the narrow scope of current solutions proposed by platforms such as Meta's VRS are sufficient. Our work thus underscores the need for platforms, auditors and regulators to expand the set of domains where impacts of ad delivery algorithms are considered.

2 METHODOLOGY: USING PAIR OF SCHOOLS WITH DE-FACTO RACIAL SKEW IN RECRUITMENT

We now describe how we test for discriminatory ad delivery in the education domain by pairing ads for for-profit and public schools while controlling for confounding factors.

The key insight of our design is to identify a historical racial disparity in the higher-education sector that we hypothesize may be propagated by ad delivery algorithms. Specifically, we use known differences in enrollment of Blacks among for-profit and public schools to select content and landing pages for a pair of ads that are designed to probe the platform's algorithms and test whether it perpetuates the existing differences, even when both ads and their descriptions appear as similar education opportunities to users. Our approach is inspired by prior research that used a pair of job opportunities requiring identical qualifications but exhibiting a de-facto gender skew among different employers [27].

2.1 Identifying a De-facto Skew In Education

We identify candidate schools to advertise for based on the de-facto racial skew towards Black students in for-profit colleges and towards White students in public colleges. Per 2022 College Scorecard data from the U.S. Department of Education [60], Black students make up 25% of the student body at for-profit colleges, whereas they account for only 14% of students at public colleges. This difference serves as the basis for designing our experiments to assess whether the ad delivery algorithms lead to discriminatory outcomes. From the list of all public and for-profit colleges in College Scorecard data, we first build two shortlists of four-year colleges that have a de-facto racial skew, one for for-profit schools whose demographics skews towards Black students, and the second for public schools whose demographics skews towards White students. Our hypothesis is that, if a platform's algorithm for education ad delivery is discriminatory in a way consistent with the de-facto skew, a for-profit school ad will be delivered to a disproportionately larger fraction of Black users than the public school ad.

Our hypothesis for the potential for this methodology to showcase discriminatory delivery stems from knowing platforms' ad delivery algorithms factor in historical data. The algorithms are trained on data about relationships between users and entities they interact with collected from myriad of sources on and off the platform [18], and from both online and offline sources [24]. Historically, for-profit colleges have disproportionately targeted racial minorities [31], which also reflects in the current demographics of the students in those schools. We know platforms' ad delivery algorithms consider not only a particular user's prior interactions with colleges, but also interactions of other "similar" users [23]. Therefore, a Black person may receive a higher relevance score for a for-profit college ad than a White person because, historically, other Black people have interacted with the school or been targeted by similar schools. Our method is designed to take advantage of this kind of "learned" bias to interrogate the algorithm for disparate outcome discrimination.

2.2 Identifying a Pair of Education Opportunities to Minimize Confounding Factors

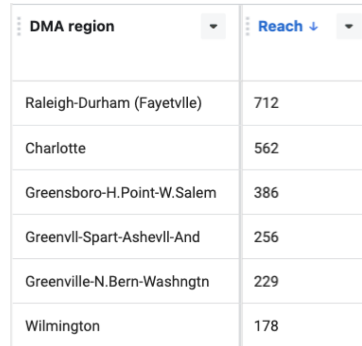
Based on the de-facto skew, we next identify pairs of schools that offer educational opportunities that are equivalent in terms of selectivity and availability of online programs, but differ in their for-profit status. We narrow down our shortlist of for-profit and public schools to those that are not very selective (acceptance rate $\approx> 50\%$) to minimize potential skew in delivery due to differences in educational qualifications among users in our audience. We select schools that offer online, part-time degree programs to minimize effects of school location relative to the location of our target audience.

Additionally, we aim to ensure the platform has sufficient signal and data about the schools we pick. We thus further narrow down the list to schools that have at least 5,000 students, have an active page on the Meta platform, and actively run ad campaigns on the platform.

Using the above criteria, we narrow down our shortlist to 9 public schools and 3 for-profit schools. We sort the for-profit schools based on the descending order of the differences between the percentages of Black and White students enrolled. Simultaneously, we sort the public schools in descending order of the differences between the percentages of White and Black students enrolled. We then group the three for-profit schools with a public school from the corresponding location in the sorted lists. The pairs of schools selected for our experiments and the racial makeup of each school's student body are shown in Table 1. Although the schools in each pair may be differently ranked academically, we do not think this difference affects our evaluation of discrimination. For both types of schools, non-discrimination should result in similar ad delivery to the different racial groups regardless of school ranking.

Pair ID	For-profit School	Public School
epair-1a	Strayer University (B=79%, W=13%, O=8%) (Admit: 100%)	Colorado State University (B=7%, W=64%, O=29%) (Admit: 98%)
epair-2a	American InterContinental University (B=29%, W=26%, O=45%) (Admit: 100%)	Fort Hays State University (B=2%, W=50%, O=48%) (Admit: 91%)
epair-3a	Monroe College (B=42%, W=3%, O=55%) (Admit: 49%)	Arizona State University (B=7%, W=58%, O=35%) (Admit: 73%)

Table 1: List of pairs of schools we use in our experiments on delivery of education ads. For each school, the table shows the racial makeup of the student body ("B" = Black students, "W" = White students, "O" = Other) and the admission rate.



DMA region	Reach ↓
Raleigh-Durham (Fayetvlle)	712
Charlotte	562
Greensboro-H.Point-W.Salem	386
Greenvll-Spart-Ashevll-And	256
Greenville-N.Bern-Washngtn	229
Wilmington	178

Figure 1: A partial screenshot of Meta's Ads Manager that demonstrates aggregate location data that Meta reports for ad recipients.

We pick schools that are correlated with race in terms of school demographics and for-profit status to elicit bias that may exist in the ad delivery algorithm. Picking schools in such a way may introduce confounding factors such as differences in familial or social ties to alumni [26] and familiarity with a school's brand [30]; these may affect how likely a prospective student is to be admitted or how relevant the opportunity is to the student. Consideration of ties with alumni, in particular, is a controversial practice among elite schools that is known to affect the racial composition of admitted students [5], and is being challenged by the U.S. Department of Education [55]. Our approach minimizes the effect of such factors by picking non-elite schools and targeting a large sample of random individuals from U.S. states that are not tied to the specific location of the schools (see §2.3.1 and Appendix A for details). Furthermore, because platforms currently do not provide special access to support auditing, it limits the conceivable confounding factors we can control for.

2.3 Running Ads on Meta and Evaluating their Delivery

Having described the key insight of our methodology based on which we identify pairs of education opportunities to advertiser for, we next describe the steps we take to actually run the ads on Meta's ad platform.

At a high-level, the steps include building an ad audience, selecting ad creatives, budgets and other ad campaign parameters, launching the ad campaigns, collecting data on their delivery, and then using that data to evaluate skew in how the ads are delivered by race. Meta does not allow targeting by race and does not return information about the race of ad recipients, so our ad audience construction and evaluations of the ad's performance are crafted to be able to make such inferences.

2.3.1 Building Audiences so that Location Reports Correspond to Race. Following the approach from prior work [2], we build the target audiences for our ads in a way that will allow us to infer the race of ad recipients from Meta's reports on locations of ad recipients. We use two features of Meta's advertising system to do so – its reports on the number of impressions an ad receives broken down by Designated Market Area (DMA); and the Custom Audience feature, that allows advertisers to create a target audience using a list of personally identifiable information, such as names, email and home addresses [41]. Figure 1 shows a screenshot of Meta's Ads Manager portal, illustrating a report provided by Meta to the advertiser, of the locations of the ad impressions aggregated by DMAs.

We construct ad audiences by (DMA, race) pairs, so that we can infer the breakdown of delivery by race from the breakdown of delivery by DMA provided by the platform. We rely on voter datasets that contain race information of individuals, and build each Custom Audience so that half of it consist of White people from one group of DMAs and another half of it consists of Black people from another group of DMAs (non-overlapping with the first group). For our experiments, we use publicly available voter dataset from North Carolina (NC) [48] (see the summary statistics in Table 2). For example, say we include only Black individuals from Raleigh DMA and only White individuals from Charlotte DMA in our ad audience. Then, whenever our ad is shown in Raleigh, we can infer it was shown to a Black person and, when it is shown in Charlotte – we can infer it was shown to a White person.

If an ad is delivered to a user outside the DMAs listed in Table 2, for example, due to people traveling, we disregard the impression in our evaluation since we cannot infer the race of the user. To ensure location does not skew our results, we replicate all our experiments using "flipped" audiences where we reverse the group of DMAs from which we pick Black and White individuals we include in our audience.

To evaluate reproducibility of our auditing results without introducing test-retest bias, we repeat our experiments on randomly selected audience partitions that are subsets of the voter lists. Each partition contains 15K White and 15K Black individuals, which we find by running test ads is a large enough audience size to get enough samples for our experiments. We conducted our test ads on partitions distinct from those we use in our experiments to evaluate bias in ad delivery. Like prior audits of ad delivery [2, 27], our methodology does not require targeting an equal number of Black

Table 2: List of voter datasets we use to construct ad audiences.

Group ID	Group of DMAs	# of Blacks	# of Whites
Group 1	Raleigh-Durham, Wilmington, Greenville-Spartaburg, Norfolk-Portsmouth	697,492	2,282,243
Group 2	Charlotte, Greensboro, Greenville-New Bern	818,599	2,564,627

and White users for validity, but we still do so for ethical reasons to avoid discriminating as part of conducting our audits. We name each audience partition based on whether the audience is a flipped version or not. For example, a partition named "aud-nc-1" indicates we included Black individuals from DMA group 1 (from Table 2) and White individuals from DMA group 2; and a partition named "aud-nc-1f" is a flipped version of the audience.

2.3.2 Selecting Ad Creatives and Campaign Parameters. For ad creatives, we first use ad text and images that are neutral and consistent across each pair of ads to control for the possible effects due to creative choices [32]. We then run ads for the same schools using realistic creatives taken from each school's Meta Ad Library page to test how the delivery algorithm may amplify implicit cues in ad creatives.

Neutral creatives: To minimize the possible influence of ad creative choice on delivery, we ensure the image and text of our ads are consistent and neutral. For each school, we use a picture of the school's campus or logo, and avoid using images with people's faces, which prior work has shown influences ad delivery [32] or may influence users' engagement [46]. We use a consistent headline text for all ads that prompts recipients to enroll in an online degree program, with the sole difference being the incorporation of the respective school's name. The destination sites, however, are different for each ad – they link to the school's official website dedicated to online programs. We do so to ensure participants who are interested in the ad are provided access to the actual opportunity. Figure 2a and Figure 2b show an example pair of neutral ad creatives we use in our experiments.

Realistic creatives: To test the effects of ad delivery for real-world ads used by schools, we also run experiments using realistic ad creatives. We take a snapshot of a list of ads run by the schools from Meta's public Ad Library [40] and manually annotate the images used in the ads by (perceived) race. The manual effort introduces the potential for annotation bias but extreme precision is not mandatory for our methodology. For each school, we select from the list one representative ad that includes the face of an individual whose race is represented in majority of the ads. Figure 2c and Figure 2d show an example pair of realistic ad creatives we use in our experiments.

Similar to prior work [2, 27], we run each pair of ads in an experiment simultaneously with the same campaign parameters including the budget, the audience, and time duration for the ads. Running a pair of ads in such a way controls for temporal factors and market effects that may otherwise confound our measurement, and thus allows to isolate the role of ad delivery for any differences in outcomes. We run all our ads with a "Traffic" objective that aims to increase traffic to the website that our ads link to [43]. We run them for a full 24 hours with a total budget of $50 per ad. We do not label our education ads as "Special Ad Category", as only housing, employment, credit and social issue ads are required to do so. We limit targeting to the Custom Audiences we built based on voter data and limit delivery to United States. We do not add any additional targeting parameters.

2.3.3 Launching and Monitoring Performance. We launch and monitor our ads using APIs that Meta provides to advertisers. We do not launch our experiments until both ads have been approved by Meta. Once our experiment starts and the pair of ads starts being shown to users, we use a script that fetches ad performance data at least once every hour to track delivery over time. Because we build our audience in such a way that uniquely maps a location of a recipient to their race, we use the DMA attribute that Meta provides to calculate the number of unique impressions by location, and in turn, by the corresponding race.

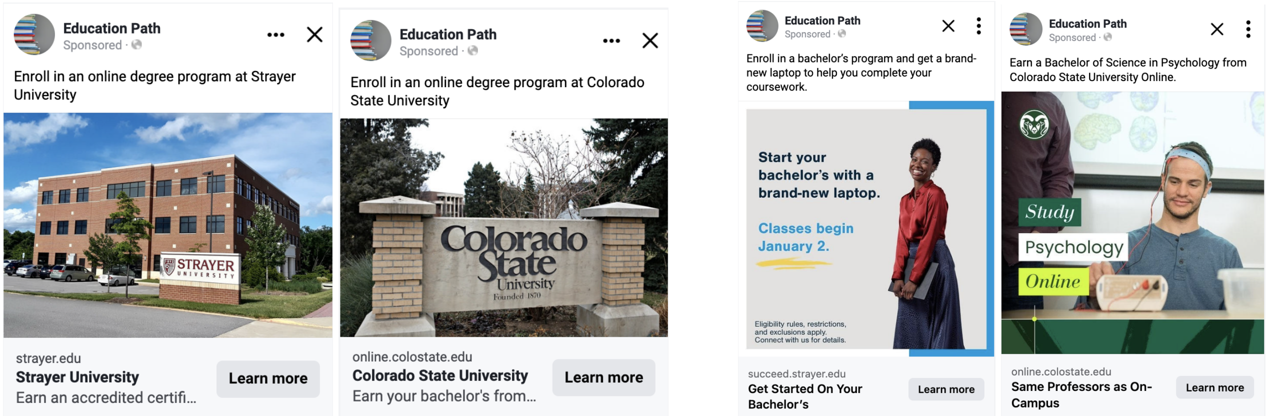
2.3.4 Evaluating Skew. We apply a skew metric to the racial breakdown of unique ad impressions to evaluate statistical significance of any racial skew we observe in the delivery of the pair of education ads. We use a metric that is established in the literature for comparing the delivery of a pair of ads [2, 3, 27]. We next introduce the notations and statistical test for the metric that underlie the empirical findings we present in §3.

Let $n_{f,b}$ and $n_{f,w}$ represent the number of users that saw the for-profit school ad, and are Black and White, respectively. We observe these numbers from Meta's reported "Reach" metric on the ad's performance, which corresponds to unique impressions (number of people that saw the ad). One can define the same terms for the public school ad ($n_{p,b}$ and $n_{p,w}$). We can calculate the fractions of Black users among the for profit school ad's and public school ad's recipients as:

$$s_{f,b} = \frac{n_{f,b}}{n_{f,b} + n_{f,w}} \quad \text{and} \quad s_{p,b} = \frac{n_{p,b}}{n_{p,b} + n_{p,w}}.$$

We apply a statistical test to compare $s_{f,b}$ and $s_{p,b}$ and evaluate whether there is statistically significant racial skew in ad delivery. When the ad delivery algorithm is not skewed, we expect $s_{f,b} = s_{p,b}$ because we ensure other confounding factors affect both ads equally. As long as the two fractions are equal, even if they are not equal to 0.5, there is no delivery bias. For example, both ads may be delivered to 60% Blacks due to fewer White people being online at the time of the experiments or other advertisers bidding higher for Whites than Blacks, and thus this skewed outcome would not be due to ad delivery algorithm's bias. However, if there is a *relative difference* between $s_{f,b}$ and $s_{p,b}$, we can attribute it to choices made by the platform's ad delivery algorithm.

We use D to represent the difference between $s_{f,b}$ and $s_{p,b}$: $D = s_{f,b} - s_{p,b}$. We apply one-sided Z-test for difference in proportions to test whether the difference between the two fractions is statistically significant, where our null hypothesis is $D = 0$ and our alternate hypothesis is $D > 0$. The test statistic is given by:



(a) For-profit (neutral creative) **(b) Public (neutral creative)** **(c) For-profit (realistic)** **(d) Public (realistic)**

Figure 2: Example ad creatives for studying racial skew in delivery of education ads. The two left figures use neutral ad creatives that do not include people. The two right figures use realistic creatives taken from each school's ad library page and include people of a specific perceived race.

$$Z = \frac{D}{SE} \quad \text{where} \quad SE = \sqrt{\hat{s}_b(1-\hat{s}_b)\left(\frac{1}{n_f}+\frac{1}{n_p}\right)} \qquad (1)$$

where $n_f = n_{f,b} + n_{f,w}$ and $n_p = n_{p,b} + n_{p,w}$ and $\hat{s}_b$ is the fraction of Black users in combined set of all people that saw at least one of the two ads: $\hat{s}_b = \frac{n_{f,b}+n_{p,b}}{n_f+n_p}$. Finally, we pick a level of significance α (typically, 0.05), to determine the corresponding critical value of Z_α from the Z-table for standard normal distribution, and conclude that there is a statistically significant racial skew in the ad delivery algorithm if $Z > Z_\alpha$. We use a 95% confidence level ($Z_\alpha = 1.64$) for all of our statistical tests. This hypothesis assumes the samples of individuals that see the ads are independent and that n is large. The delivery audiences may have overlapping samples that are dependent if the same person sees both ads, but we target a large audience to minimize such an outcome. The sample sizes vary by each experiment, but they are at least 1,500 as shown in the "n" column in Figure 3.

To check robustness of our results, we additionally apply a multiple-test correction. We use Holm's method [25], a statistical technique that corrects for the issue of multiple testing by adjusting the threshold for statistical significance depending on the number of tests.

3 EXPERIMENTS AND RESULTS

We apply the methodology we developed to create, run and compare the delivery of ads on Meta for the pairs of for-profit and public schools. We ran all ads between April 2023 and April 2024. Our hypothesis is that, if a platform's algorithm for education ad delivery is discriminatory, the for-profit school ad will be delivered to a disproportionately larger fraction of Black users than the simultaneously run public school ad. Our findings demonstrate evidence of racial discrimination in Meta's algorithms. We make data from our experiments publicly available at [29].

We apply our method to Meta because it is a major ad platform with billions of users, and has known risks of algorithmic bias, as documented in prior academic work [2, 3, 27, 56], civil rights audits [35], and the settlement with the U.S. Department of Justice [58].

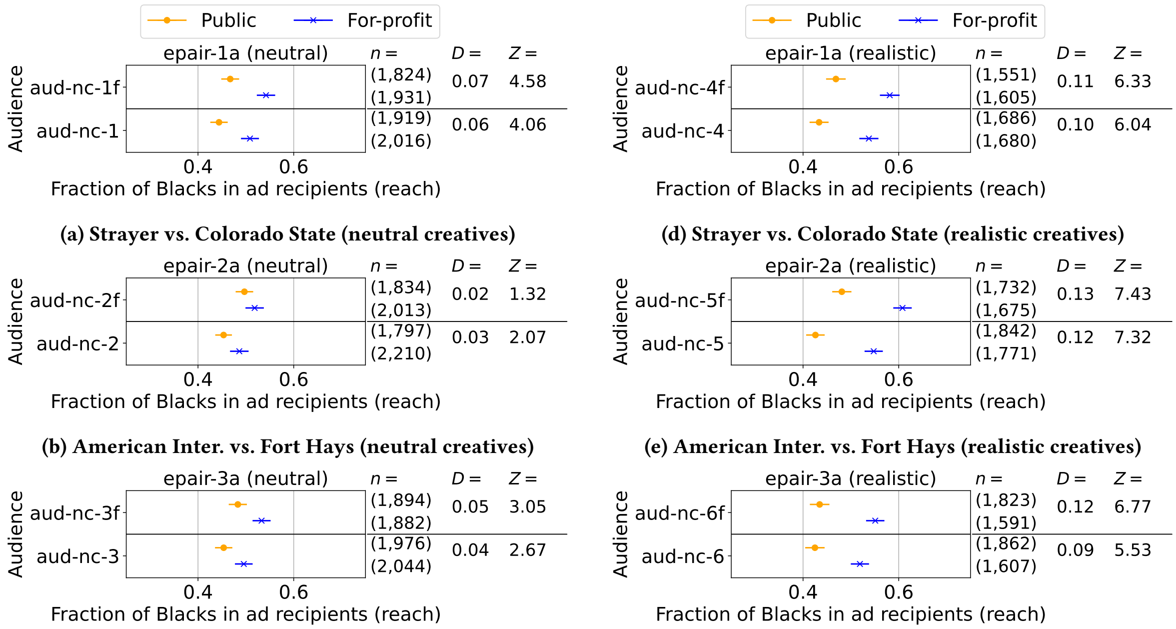
3.1 Demonstrating Discriminatory Ad Delivery using Neutral Ad Creatives

We present results of our experiments using neutral ad creatives (see Figure 2a and Figure 2b for examples of such creatives for one pair of for-profit and public schools).

Recall that Meta reports location of ad recipients, but we are interested in the racial breakdown of the recipients. As discussed in §2.3.1, we repeat each experiment on two audiences by flipping the DMAs from which we pick White and Black individuals. Evaluating both combinations allows us to factor out location as a confounding factor. By applying this procedure to the three pairs of schools we identified in Table 1, we run a total of six experiments. In Appendix A, we present additional experimental results where we use states, instead of DMAs, as a way to construct audiences with a location - race correspondence.

We show our direct observations of ad delivery by race for the experiments in the left column of Figure 3. Each row of that column shows the result of an experiment of running ads for a pair of schools, one public (top, orange) and the second – for-profit (bottom, blue). For each pair of ads, we report D computed by subtracting the fraction of Blacks who saw the public school ad from the fraction of Blacks who saw the for-profit school ad ($D = s_{f,b} - s_{p,b}$). We expect D to be positive if there is discriminatory ad delivery consistent with the de-facto racial skew in the demographics of the schools' student body (Table 1).

We find that the ads for for-profit school in all six experiments with neutral creatives are shown to a higher fraction of Black users (D is positive), as shown in Figure 3a, Figure 3b and Figure 3c. Moreover, the racial skew is statistically significant in five out of



(a) Strayer vs. Colorado State (neutral creatives)

(b) American Inter. vs. Fort Hays (neutral creatives)

(c) Monroe vs. Arizona State (neutral creatives)

(d) Strayer vs. Colorado State (realistic creatives)

(e) American Inter. vs. Fort Hays (realistic creatives)

(f) Monroe vs. Arizona State (realistic creatives)

Figure 3: Results for Meta's delivery of educations ads for neutral creatives (left) and realistic creatives (right). Bars show 95% confidence intervals around each fraction. n is the number of individuals each ad was shown to. D is the difference between fraction of Blacks seeing for-profit and public school ads. Z is the test statistic for significance of this difference. An audience named "aud-nc-*" is built using Black individuals from DMA group 1 (Table 2) and White individuals from group 2; "aud-nc-*f" is a flipped version.

the six experiments. To illustrate, consider Figure 4a which reports the test statistics for each pair of ads. We compute the significance test statistics using the formula in Equation 1, and we compare it to the threshold of statistical significance, shown by the dotted horizontal line. The racial difference in ad delivery for a pair of ads is statistically significant for a pair when a test statistic is above the horizontal line. The test statistic for epair-1a and epair-3a crosses the threshold for both the initial and flipped audiences. For the second pair (epair-2a), the skew in ad delivery we observe is not large enough to be statistically significant for the flipped audience. One possible explanation is that the de-facto racial skew in the demographic of students for the second pair is smaller than the other two pairs (see Table 1), hence resulting in a smaller skew in ad delivery. Our conclusions remain the same after applying Holm's correction for multiple hypothesis testing over the six tests.

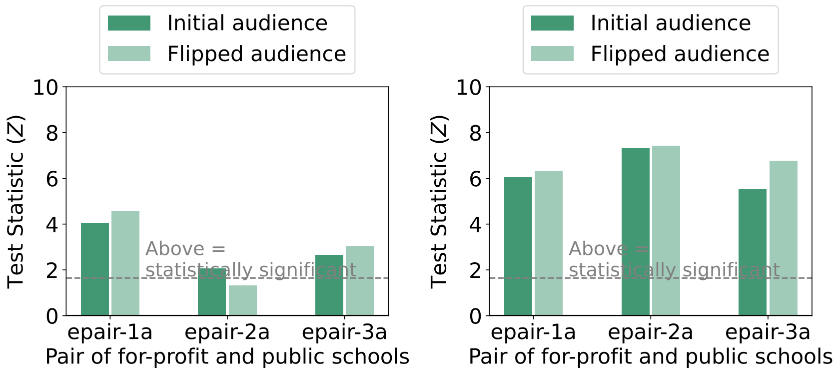
Our finding of algorithm-induced skew in majority of the cases suggests *Meta's ad platform racially discriminates in the delivery of education ads*. Because we measured the relative difference in the delivery of paired ads, the bias we observed cannot be explained by confounding factors such as targeting choices, differences in who is online, competition from concurrently running ads, or market forces. Such factors affect both the for-profit and public college ads in our experiments equally. Therefore, the racial bias we measured in the delivery of the education ads is a product of choices made by Meta's ad delivery algorithms.

3.2 Demonstrating Amplified Skew Using Realistic Ad Creatives

We next re-run the ads for the same schools as in §3.1, but using realistic ad creatives (sampled from those actually used by schools), to measure by how much racial skew in ad delivery increases in the real world. Our expectation is that creatives that include faces of students may propagate assumptions about the racial mix of student body, giving a platform's ad delivery algorithm additional, implicit information to determine to whom the ad may be relevant. Prior work has shown this additional factor may *increase* the skew [32] and studied the potential for discrimination through selective use of images in the employment advertising context [46].

Using the method described in §2.3.2, we observe that the for-profit schools have more ads depicting Black faces than ads depicting White faces, and vice versa for the public schools. We thus pick a representative ad that includes the face of a (perceived) Black person for the for-profit school and a face of a White person for the public school, respectively.

We find that the racially skewed delivery of the ads for the for-profit school is further amplified for the realistic ad creatives (Figure 3; right column). We can see the amplification by comparing the levels of skew (D) in the two columns of the figure. For example, for epair-1a, the skew is $D = 0.07$ when measured using neutral ad creatives ("aud-nc-1f" in Figure 3a), but is larger, $D = 0.11$, when using realistic ad creatives ("aud-nc-4f" Figure 3d). We illustrate



(a) Using neutral ad creatives.

(b) Using realistic ad creatives.

Figure 4: Statistical significance of racial skew in delivery of education ads on Meta. The test statistic is computed based on the racial skew measured in Figure 3. The racial skew in delivery between a pair of ads is statistically significant if the test statistic bar is above the horizontal line (which corresponds to a 95% confidence level: $Z_\alpha = 1.64$). Each bar corresponds to an experiment.

how the amplified skew affects our conclusion in Figure 4. For each pair of schools, the skew we measure using realistic ad creatives (bars in Figure 4b) are larger than the corresponding measurement using neutral ad creatives (bars in Figure 4a). For the realistic ad creatives, we observe a statistically significant skew for all three pairs, including epair-2a for which the skew we measured using neutral ad creatives is not statistically significant. The results for all three pairs remain statistically significant after applying Holm's correction to the family of tests.

This result shows that, in addition to perpetuating historical racial biases associated with the schools, *platforms amplify implicit cues in ad creatives used by schools*. Taken together, the advertiser's choice of ad creatives and algorithmic steering that further amplifies those choices point to a serious impact of ad delivery algorithms in shaping access to education opportunities.

3.3 Experiments Using Schools with Historical Predatory Practices

We next apply our methodology to examine whether Meta's algorithm delivers ads for for-profit schools with historical predatory marketing practices to disproportionate fraction of Black individuals. Disproportionate promotion of opportunities at such institutions poses potential harm to students who enroll in these schools [12, 31, 38].

We find historically predatory schools based on a list published by the U.S. Department of Education of schools that previously have been sued or fined for deceptive marketing practices [50]. From this list, we pick the three universities with the largest number of students: DeVry, Grand Canyon and Keiser. We then evaluate ad delivery for those schools using our methodology. We would like to test whether Meta's algorithms perpetuate the historical skew even if the schools have improved their marketing practices in response to the legal challenges.

We apply the methodology described in §2, but we modify the school-selection criteria to consider historically predatory schools instead of those with skew in racial demographics. We continue to match these for-profit schools with the same three public schools used in our earlier experiments, following the process outlined in §2.2. Table 3 provides an overview of the school pairings we used in this experiment. Similar to §3.2, we use realistic ad creatives that are taken from each school's Meta ad library page (see Appendix B for ad screenshots). As in our previous experiments, we replicate the experiments on a "flipped" audience to ensure location does not skew our results.

We find that, for all three pairs, the historically predatory for-profit school ad is shown to relatively larger fraction of Blacks compared to the public school ad (Figure 5; left column). For example, for the first pair of schools (top row of Figure 5a), the public school ad (orange circle) is delivered to 44% Blacks whereas the for-profit school ad (blue cross mark) is delivered to relatively larger fraction of Blacks (52%). For two out of the three pairs, the racial skew in delivery is statistically significant (see Figure 5d) and remains statistically significant after applying Holm's correction.

In summary, our findings show that *Meta's algorithms disproportionately deliver educational opportunities that could be harmful to individuals based on race*. Our study illustrates how this harm can happen in practice, as the for-profit schools with historical predatory advertising are currently active advertisers on Meta's ad platform.

4 DISCUSSION

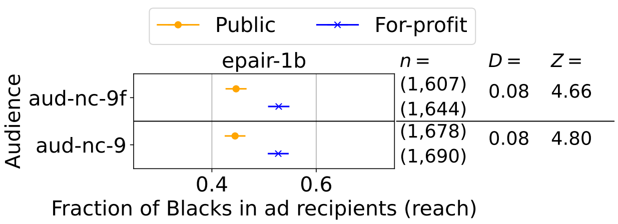
Our work has implications for how platforms shape access to education opportunities, the potential legal liability that they may incur as a result, and the need for platforms, researchers, and regulators to follow a more holistic approach to identifying and addressing the issues of bias and discrimination in ad delivery.

4.1 Platforms Shaping Access to Education Opportunities

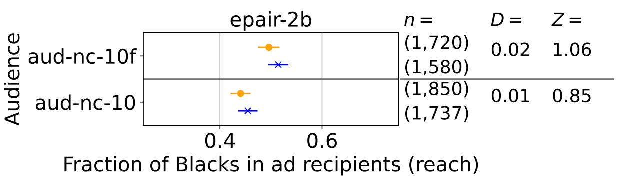
Our findings highlight the negative impact of ad delivery algorithms in shaping access to education opportunities, adding to existing concerns about the disadvantages of for-profit schools and their historical involvement in predatory advertising. Studies have raised concerns that for-profit colleges provide poorer outcomes, with their students showing higher loan default rate, lower earnings and employment than comparable students at other post-secondary institutions [12, 16, 38]. In addition, it is known that for-profit colleges have historically engaged in predatory marketing that targeted racial minorities [31], and enroll a disproportionate number of students with low family income [9].

Table 3: Racial make-up of historically predatory for-profit schools and the public schools we pair them with.

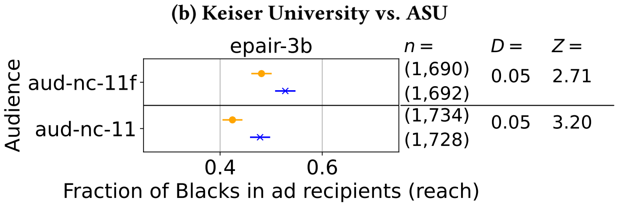
Pair ID	For-profit (historically predatory) school	Public school
epair-1b	DeVry University (B=26% W=44%, O=29%) (Admit: 44%)	Colorado State University (B=7%, W=64%, O=29%) (Admit: 98%)
epair-2b	Grand Canyon University (B=16% W=48%, O=37%) (Admit: 81%)	Fort Hays State University (B=3%, W=50%, O=47%) (Admit: 91%)
epair-3b	Keiser University (B=19% W=30%, O=51%) (Admit: 97%)	Arizona State University (B=7%, W=58%, O=35%) (Admit: 73%)



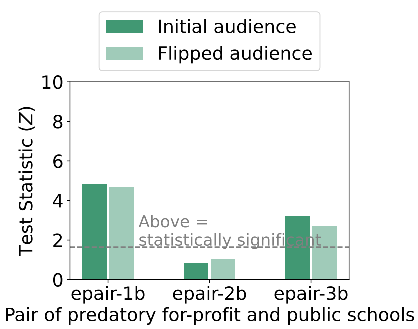
(a) DeVry University vs. Colorado State



(b) Keiser University vs. ASU



(c) Grand Canyon University vs. Fort Hays



(d) Statistical significance of skew

Figure 5: Skew measured by comparing the delivery of ads for a historically predatory for-profit school and a public school.

The platform-induced racial bias we demonstrate illustrates another significant factor that decides exposure to education opportunities: ad delivery algorithms. We show Meta's algorithms deliver ads for the for-profit schools to relatively more Black individuals than ads for public schools. The racial difference we observe in the delivery is not due to the advertiser's targeting choices since we select racially balanced audiences. It is also not due to market effects or difference in platform use by race since our methodology controls for those. Therefore, even if for-profit institutions aim for racially balanced ad targeting, Meta's algorithms would recreate historical racial skew in who the ad are shown to, and would do so unbeknownst to the advertisers. Our findings show it is not enough for schools to target their ads equitably; platforms also need to ensure their ad delivery algorithms are not biased by race.

4.2 Legal Liability for Ad Platforms

In addition to the harm in selectively showing potentially lower-quality education opportunities disproportionally to racial minorities, discriminatory delivery of education ads risks legal liability for Meta.

Educational opportunities have legal protections that prohibit racial discrimination and may apply to ad platforms. The Civil Rights Act of 1974 prohibits discrimination by race, color, or national origin [61] for schools that receive federal funding. D.C.'s Human Rights Act, which provides one of the broadest protections among the states in the U.S. [36], similarly prohibits discrimination by these and other attributes such as sex, age and religion for several domains including education [63].

Our results show Meta should be under legal scrutiny for the role its algorithms play in the delivery of education ads. Because we show discrimination in the *outcome* of ad delivery, the platform may be liable under the *disparate impact* doctrine of discrimination [7, 11]. Under this doctrine, a claim of discrimination can be made if the algorithmic outcome differs significantly by a protected demographic attribute, regardless of the source of bias for such outcome [7]. In this case, the burden for justifying the specific source of bias shifts to the platform. Given our findings of discriminatory ad delivery after controlling for conceivable confounding factors, these results suggest that Meta may need to either justify the concern, or address it with modifications to its algorithms.

5 ETHICAL CONSIDERATIONS

We conduct our audits with consideration of the ethical implications to both individuals engaging with our advertisements and to platforms. First, we build our audiences with voter registration datasets available to the public through election offices of U.S. states. Since this data is already public, it poses minimal new privacy risks. Furthermore, we use this public information only to build our ad audiences; we do not interact directly with the users, nor do we receive or collect identifiers about individuals who see our ads. We observe and report only aggregate statistics about the results of ad performance. Although we consider these risks minimal, this use of voter datasets may require additional considerations in cases where GDPR applies. Second, while we purchase ads, our spending on the purchases is tiny and the benefits of learning about potential discrimination in platform algorithms greatly outweighs potential cost they impose on the ad recipients. While our ads for for-profit schools may substitute for better opportunities, our ad spending is vanishingly small (a few hundred dollars) relative to the advertising budgets of for-profit colleges (where their median marketing cost to recruit a single student exceeds $4k [19]). We minimize any overhead our ads place on their viewers – they link to real education opportunities, and we select audiences of equal sizes by race. Our approaches do not harm the platform, follow the terms of service and use only features and APIs Meta makes available to any advertiser. Our work was reviewed and approved by the Institutional Review Boards at the University of Southern California (review #UP-20-00132) and at Princeton University (record #14833).

6 CONCLUSION

Our study demonstrates yet another domain and demographic group for which platforms shape access to important life opportunities. The racial bias we find in education ad delivery shows discriminatory delivery extends beyond the current scope of solutions, which have been limited to housing, employment and credit domains [6], and raises the broader questions of what other domains with legal concerns of discrimination or ethical concerns of bias need equal level of attention from platforms and regulators. A recent Executive Order towards regulating Artificial Intelligence in the U.S. identifies a number of other domains, such as insurance, healthcare, and childcare, as domains where AI impacts access to opportunities [8, 15]. Prior studies have also demonstrated that ad delivery algorithms shape public discourse on topics such as politics [3] and climate-change [54].

Given the range of domains with concerns of algorithmic bias, we call on platforms to, first, conduct impact assessment of their ad delivery algorithms across all domains relevant to civil rights and societally important topics, and publicize their findings. Second, platforms should do more to allow for independent external scrutiny of the impacts of their algorithms. Current platform transparency efforts are limited to data about content, not algorithms [42, 66] and do not provide an infrastructure for experimentation. Thus the scrutiny of algorithmic impacts so far has been limited to intricate black-box audits custom designed for each new domain where a concern of discrimination arises. In the rare cases when algorithmic scrutiny has been supported by platforms, it was done only through close collaboration between the platforms and select researchers [20–22, 49], without guarantees of full independence. Our work is evidence that the advocacy in the E.U. and the U.S. [10, 14] for a wider range of public-interest researchers to be able to scrutinize the algorithms is justified, and indeed, essential to make progress. A platform-supported approach that gives researchers a standardized interface to not only public data but to the outputs of the algorithms in a privacy-preserving manner is a promising approach towards this goal [28, 52].

ACKNOWLEDGMENTS

This work work was funded in part by a grant from The Rose Foundation and National Science Foundation grants CNS-1956435, CNS-1916153, CNS-2333448, CNS-1943584, CNS-2344925, CNS-2319409, and CNS-1925737. We are grateful to Jonathan Mayer for feedback on the manuscript.

REFERENCES

[1] Muhammad Ali, Angelica Goetzen, Alan Mislove, Elissa M. Redmiles, and Piotr Sapiezynski. 2023. Problematic Advertising and its Disparate Exposure on Facebook. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA, 5665–5682. https://dl.acm.org/doi/abs/10.5555/3620237.3620554

[2] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. 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*, Vol. 3. ACM, 1–30. https://dl.acm.org/doi/10.1145/3359301

[3] Muhammad Ali, Piotr Sapiezynski, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2021. Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (Virtual Event, Israel) *(WSDM '21)*. Association for Computing Machinery, New York, NY, USA, 13–21. https://doi.org/10.1145/3437963.3441801

[4] Julia Angwin and Terry Paris Jr. 2016. Facebook Lets Advertisers Exclude Users by Race – ProPublica. Retrieved February 18, 2021 from https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race

[5] Peter Arcidiacono, Josh Kinsler, and Tyler Ransom. 2022. *What the Students for Fair Admissions Cases Reveal About Racial Preferences.* Working Paper 29964. National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/w29964

[6] Roy L. Austin. 2023. An Update on Our Ads Fairness Efforts. https://about.fb.com/news/2023/01/an-update-on-our-ads-fairness-efforts/.

[7] Solon Barocas and Andrew D Selbst. 2016. Big data's disparate impact. *California law review* (2016), 671–732.

[8] Joseph R Biden. 2023. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Retrieved November 30, 2023 from https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

[9] Raj Chetty, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. *Mobility Report Cards: The Role of Colleges in Intergenerational Mobility.* Working Paper 23618. National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/w23618

[10] Chris Coons, Rob Portman, Amy Klobuchar, and Bill Cassidy. 2022. Platform Accountability and Transparency Act. Retrieved December 21, 2021 from https://www.coons.senate.gov/news/press-releases/senator-coons-colleagues-introduce-legislation-to-provide-public-with-transparency-of-social-media-platforms

[11] Amit Datta, Anupam Datta, Jael Makagon, Deirdre K Mulligan, and Michael Carl Tschantz. 2018. Discrimination in Online Personalization: A Multidisciplinary Inquiry. *Proceedings of the 2018 ACM Conference on Fairness, Accountability, and Transparency* 81 (2018), 20–34.

[12] David Deming, Claudia Goldin, and Lawrence Katz. 2013. For-Profit Colleges. *The Future of Children* 23, 1 (2013), 137–163. http://www.jstor.org/stable/23409492

[13] Cynthia Dwork and Christina Ilvento. 2019. Fairness Under Composition. In *10th Innovations in Theoretical Computer Science Conference (ITCS 2019)*, Vol. 124. 33:1–33:20. https://doi.org/10.4230/LIPIcs.ITCS.2019.33

[14] European Commission. 2022. Digital Services Act: Commission welcomes political agreement on rules ensuring a safe and accountable online environment. Retrieved April 27, 2022 from https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2545

[15] Executive Office of the President. 2023. Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence. Retrieved December 27, 2023 from https://ai.gov/wp-content/uploads/2023/11/AI-in-Government-Memo-Public-Comment.pdf

[16] Federal Reserve. 2023. Report on the Economic Well-Being of U.S. Households in 2022. Retrieved November 1, 2023 from https://www.federalreserve.gov/publications/files/2022-report-economic-well-being-us-households-202305.pdf

[17] Florida Division of Elections. 2022. Voter Extract Disk Request. https://dos.fl.gov/elections/data-statistics/voter-registration-statistics/voter-extract-disk-request/

[18] Carolin Gerlitz and Anne Helmond. 2013. The Like Economy: Social buttons and the data-intensive web. *New Media & Society* 15, 8 (2013), 1348–1365. https:

//doi.org/10.1177/1461444812472322
[19] Howard R. Gold. 2019. Who's at Fault for Student-Loan Defaults? Retrieved January 29, 2023 from https://www.chicagobooth.edu/review/whos-fault-student-loan-defaults
[20] Sandra González-Bailón, David Lazer, Pablo Barberá, Meiqing Zhang, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Deen Freelon, Matthew Gentzkow, Andrew M. Guess, Shanto Iyengar, Young Mie Kim, Neil Malhotra, Devra Moehler, Brendan Nyhan, Jennifer Pan, Carlos Velasco Rivera, Jaime Settle, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker. 2023. Asymmetric ideological segregation in exposure to political news on Facebook. *Science* 381, 6656 (2023), 392–398. https://www.science.org/doi/abs/10.1126/science.ade7138
[21] Andrew M. Guess, Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Edward Kennedy, Young Mie Kim, David Lazer, Devra Moehler, Brendan Nyhan, Carlos Velasco Rivera, Jaime Settle, Daniel Robert Thomas, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker. 2023. How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science* 381, 6656 (2023), 398–404. https://www.science.org/doi/abs/10.1126/science.abp9364
[22] Andrew M. Guess, Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Edward Kennedy, Young Mie Kim, David Lazer, Devra Moehler, Brendan Nyhan, Carlos Velasco Rivera, Jaime Settle, Daniel Robert Thomas, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker. 2023. Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science* 381, 6656 (2023), 404–408. https://www.science.org/doi/pdf/10.1126/science.add8424
[23] John Hegeman and Jared Morgenstern. 2009. User feedback-based selection and prioritizing of online advertisements. https://patents.google.com/patent/US20110106630A1/en. *United States Patent Application* (2009).
[24] Paul Hitlin and Lee Rainie. 2019. Facebook algorithms and personal data. *Pew Research Center* 16 (2019).
[25] Sture Holm. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6, 2 (1979), 65–70. http://www.jstor.org/stable/4615733
[26] Michael Hurwitz. 2011. The impact of legacy status on undergraduate admissions at elite colleges and universities. *Economics of Education Review* 30, 3 (2011), 480–492. https://doi.org/10.1016/j.econedurev.2010.12.002
[27] Basileal Imana, Aleksandra Korolova, and John Heidemann. 2021. Auditing for Discrimination in Algorithms Delivering Job Ads. In *Proceedings of the Web Conference 2021*. Association for Computing Machinery, New York, NY, USA, 3767–3778. https://doi.org/10.1145/3442381.3450077
[28] Basileal Imana, Aleksandra Korolova, and John Heidemann. 2023. Having Your Privacy Cake and Eating It Too: Platform-Supported Auditing of Social Media Algorithms for Public Interest. *Proceedings of the ACM Human Computer Interaction* 7, CSCW1, Article 134 (04 2023), 33 pages. https://doi.org/10.1145/3579610
[29] Basileal Imana, Aleksandra Korolova, and John Heidemann. 2024. Dataset of ad content and delivery statistics used in "Auditing for Racial Discrimination in the Delivery of Education Ads". https://ant.isi.edu/datasets/addelivery-education/.
[30] Mathew Joseph, Eileen Wall Mullen, and Deborah F. Spake. 2012. University branding: Understanding students' choice of an educational institution. *Journal of Brand Management* 20 (2012), 1 – 12. https://api.semanticscholar.org/CorpusID:168107438
[31] Suzanne Kahn, Mark Huelsman, and Jen Mishory. 2019. Bridging Progressive Policy Debates: How Student Debt and the Racial Wealth Gap Reinforce Each Other. Retrieved January 18, 2023 from https://rooseveltinstitute.org/wp-content/uploads/2020/07/RI_Student-Debt-and-RWG-201909.pdf
[32] Levi Kaplan, Nicole Gerzon, Alan Mislove, and Piotr Sapiezynski. 2022. Measurement and Analysis of Implied Identity in Ad Delivery Optimization. In *Proceedings of the 22nd ACM Internet Measurement Conference* (Nice, France) *(IMC '22)*. Association for Computing Machinery, New York, NY, USA, 195–209. https://doi.org/10.1145/3517745.3561450
[33] Michelle S. Lam, Ayush Pandit, Colin H. Kalicki, Rachit Gupta, Poonam Sahoo, and Danaë Metaxa. 2023. Sociotechnical Audits: Broadening the Algorithm Auditing Lens to Investigate Targeted Advertising. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 360 (oct 2023), 37 pages. https://doi.org/10.1145/3610209
[34] Anja Lambrecht and Catherine Tucker. 2019. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management Science* 65, 7 (2019), 2966–2981. http://dx.doi.org/10.2139/ssrn.2852260
[35] Laura Murphy and Associates. 2020. Facebook's Civil Rights Audit – Final Report. Retrieved October 19, 2020 from https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf
[36] Lisa Gabrielle Lerman and Annette K. Sanderson. 1978. Discrimination in Access to Public Places: A Survey of State and Federal Public Accommodations Laws Project. *New York University Review of Law and Social Change* 7 (1978), 215.
[37] LinkedIn. 2021. How the LinkedIn Ads Auction Works — and How You Can Benefit. Retrieved November 9, 2021 from https://www.linkedin.com/business/marketing/blog/linkedin-ads/how-the-linkedin-ads-auction-works-and-how-you-can-benefit
[38] Adam Looney and Constantine Yannelis. 2015. A Crisis in Student Loans? How Changes in the Characteristics of Borrowers and in the Institutions They Attended Contributed to Rising Loan Defaults. *Brookings Papers on Economic Activity* (2015), 1–68. http://www.jstor.org/stable/43752167
[39] Meta. 2024. About Ad Auctions. Retrieved May 2, 2024 from https://www.facebook.com/business/help/430291176997542?id=561906377587030
[40] Meta. 2024. Ad Library. Retrieved May 2, 2024 from https://www.facebook.com/ads/library
[41] Meta. 2024. Create a customer list custom audience. Retrieved May 2, 2024 from https://www.facebook.com/business/help/170456843145568?id=2469097953376494
[42] Meta. 2024. Facebook Open Research & Transparency. Retrieved May 2, 2024 from https://fort.fb.com/
[43] Meta. 2024. How to choose the right ad objective in Meta Ads Manager. Retrieved May 2, 2024 from https://www.facebook.com/business/help/1438417719786914
[44] Meta Technical Report. 2023. Towards Fairness in Personalized Ads. Retrieved November 24, 2023 from https://about.fb.com/wp-content/uploads/2023/01/Toward_fairness_in_personalized_ads.pdf
[45] Meta's Civil Rights Team. 2021. Meta's Progress on Civil Rights Audit Commitments. Retrieved November 22, 2021 from https://about.fb.com/wp-content/uploads/2021/11/Metas-Progress-on-Civil-Rights-Audit-Commitments.pdf
[46] Varun Nagaraj Rao and Aleksandra Korolova. 2023. Discrimination through Image Selection by Job Advertisers on Facebook. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, 1772–1788. https://doi.org/10.1145/3593013.3594115
[47] Brandie Nonnecke and Camille Carlton. 2022. EU and US legislation seek to open up digital platform data. *Science* 375, 6581 (2022), 610–612. https://doi.org/10.1126/science.abl8537 arXiv:https://www.science.org/doi/pdf/10.1126/science.abl8537
[48] North Carolina State Board of Elections. 2020. Voter History Data. https://dl.ncsbe.gov/index.html.
[49] Brendan Nyhan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, et al. 2023. Like-minded sources on Facebook are prevalent but not polarizing. *Nature* 620, 7972 (2023), 137–144. https://www.nature.com/articles/s41586-023-06297-w
[50] U.S. Department of Education. 2018. Law Enforcement Actions Against Predatory Colleges. Retrieved December 25, 2023 from https://sites.ed.gov/naciqi/files/2018/05/NACIQI-Enclosure-1-law-enforcement.pdf
[51] U.S. Department of Housing and Urban Development. 2024. Guidance on Application of the Fair Housing Act to the Advertising of Housing, Credit, and Other Real Estate-Related Transactions through Digital Platforms. Retrieved May 2nd, 2024 from https://www.hud.gov/sites/dfiles/FHEO/documents/FHEO_Guidance_on_Advertising_through_Digital_Platforms.pdf
[52] Victor Ojewale, Ryan Steed, Briana Vecchione, Abeba Birhane, and Inioluwa Deborah Raji. 2024. Towards AI Accountability Infrastructure: Gaps and Opportunities in AI Audit Tooling. https://doi.org/10.48550/arXiv.2402.17861
[53] Sheryl Sandberg. 2019. Doing More to Protect Against Discrimination in Housing, Employment and Credit Advertising. Retrieved October 19, 2020 from https://about.fb.com/news/2019/03/protecting-against-discrimination-in-ads/
[54] Aruna Sankaranarayanan, Erik Hemberg, and Una-May O'Reilly. 2023. The Facebook Algorithm's Active Role in Climate Advertisement Delivery. https://arxiv.org/abs/2308.03191
[55] Michael D. Shear and Anemona Hartocollis. 2023. Education Dept. Opens Civil Rights Inquiry Into Harvard's Legacy Admissions. Retrieved January 9, 2024 from https://www.nytimes.com/2023/07/25/us/politics/harvard-admissions-civil-rights-inquiry.html
[56] Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Potential for Discrimination in Online Targeted Advertising. In *Proceedings of Machine Learning Research*, Vol. 81. PMLR, New York, NY, USA, 5–19. http://proceedings.mlr.press/v81/speicher18a.html
[57] Scott Spencer. 2020. Upcoming Update To Housing, Employment, And Credit Advertising Policies. Retrieved October 19, 2020 from https://www.blog.google/technology/ads/upcoming-update-housing-employment-and-credit-advertising-policies/
[58] The US Department of Justice. 2022. Justice Department Secures Groundbreaking Settlement Agreement with Meta Platforms, Formerly Known as Facebook, to Resolve Allegations of Discriminatory Advertising. Retrieved June 29, 2022 from https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known

[59] Ariana Tobin and Jeremy B. Merrill. 2018. Facebook Is Letting Job Advertisers Target Only Men – ProPublica. Retrieved February 18, 2021 from https://www.propublica.org/article/facebook-is-letting-job-advertisers-target-only-men
[60] U.S. Department of Education. 2022. College Scorecard. https://collegescorecard.ed.gov/.
[61] USC. 1964. 42 USC Section 2000d. https://www.law.cornell.edu/uscode/text/42/2000d.
[62] USC. 1974. 20 USC Chapter 39 — Equal educational opportunities and transportation of students. https://www.law.cornell.edu/uscode/text/20/chapter-39.
[63] USC. 1977. DC Human Rights Act – D.C. Code § 2-14. http://dccode.elaws.us/code?no=2-14.
[64] Nils Wernerfelt, Anna Tuchman, Bradley Shapiro, and Robert Moakler. 2022. Estimating the Value of Offsite Data to Advertisers on Meta. *University of Chicago, Becker Friedman Institute for Economics Working Paper* 114 (2022).
[65] X. 2024. Bidding and auctions FAQs. Retrieved May 2, 2024 from https://business.twitter.com/en/help/troubleshooting/bidding-and-auctions-faqs.html
[66] YouTube. 2022. YouTube Researcher Program. Retrieved July 14, 2022 from https://research.youtube/

A EXPERIMENTS USING STATE RATHER THAN DMA WHEN CONSTRUCTING AUDIENCES

We present additional experimental results that use states instead of DMAs when constructing custom audiences with location - race correspondence. The results are consistent with those in §3.2 and §3.1, with slightly more variation, particularly in replications on initial and flipped audiences. We hypothesize these differences may be due to to state-dependent variation in platform usage, but only Meta has the data to confirm.

For these experiments, we use publicly available voter datasets from two states: Florida (FL) and North Carolina (NC) [17, 48] (see the summary statistics in Table 4). We run each of the ads on two audiences – with Black individuals and White individuals from FL and NC, respectively, and then "flip" it with White individuals from FL and Black individuals from NC. A partition named "aud1" indicates we included Black individuals from Florida, White individuals from North Carolina, and a partition named "aud1f" is a flipped version of the audience.

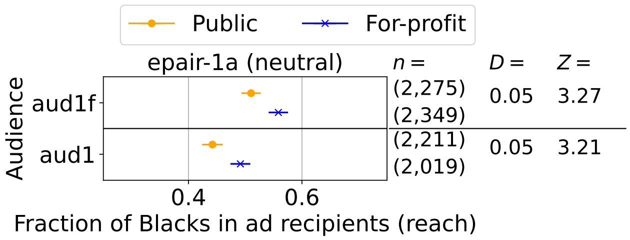
Table 4: List of voter datasets we use to construct ad audiences using state as a proxy for race.

State	# of Blacks	# of Whites
Florida (FL)	2,090,303	9,438,537
North Carolina (NC)	1,546,944	4,842,453

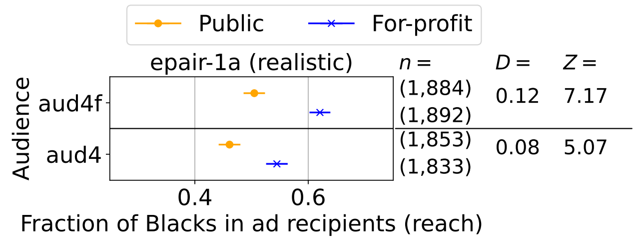
Similar to our findings in §3.2 and §3.1, these experiments show evidence of racial bias in the delivery of education ads (left column of Figure 6), and show the degree of bias in delivery increases when we use realistic ad creatives (right column of Figure 6). However, we see some variation in the experiments we replicate on flipped audiences for which we expected the results to be similar. For example, in Figure 6e, approximately 52% of recipients of the for-profit school ad are Black for aud5, whereas 70% or recipients are Black when the audience is flipped (aud5f). Because we do not look at absolute numbers but rather relative differences in the delivery of the for-profit and public schools, these variations do not affects our evaluation of skew in the algorithm. However, the results show using states when constructing audiences can introduce additional variations one must keep into account.

B REALISTIC AD CREATIVES USED FOR HISTORICALLY PREDATORY FOR-PROFIT SCHOOLS

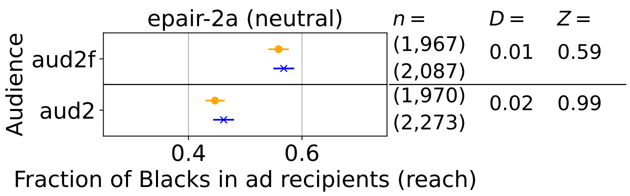
In Figure 7, we show the realistic ad creatives used for experiments involving historically predatory for-profit schools in §3.3. These ad creatives are taken from each school's page on Meta's public ad library (see §2.3.2 for details on creative choice).



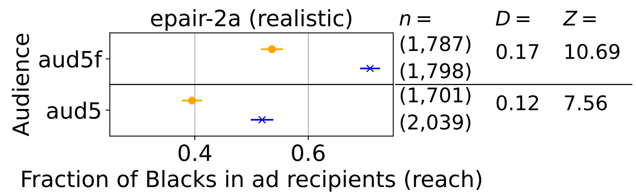
(a) Strayer vs. Colorado State (neutral creatives)



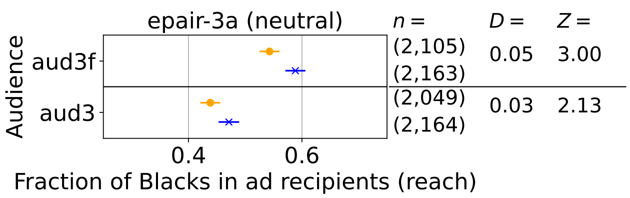
(d) Strayer vs. Colorado State (realistic creatives)



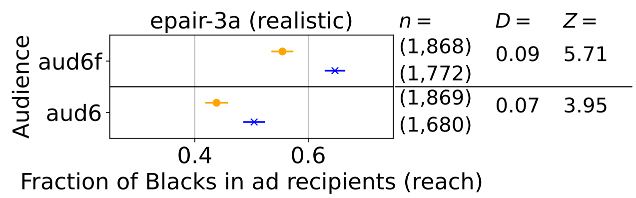
(b) American Inter. vs. Fort Hays (neutral creatives)



(e) American Inter. vs. Fort Hays (realistic creatives)

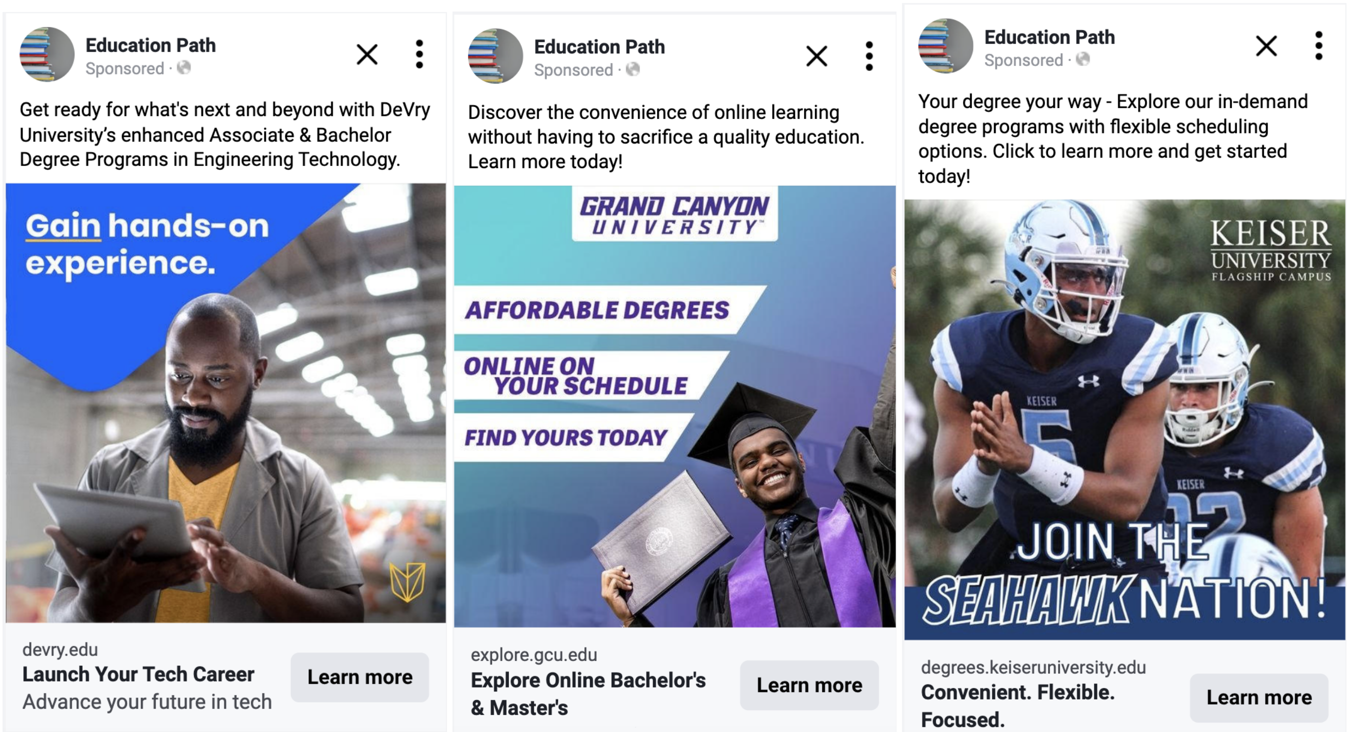


(c) Monroe vs. Arizona State (neutral creatives)



(f) Monroe vs. Arizona State (realistic creatives)

Figure 6: Additional results for racial skew in the delivery of education ads on Meta. These experiments were conducted using states (NC and FL) as a proxy, hence showing some variation between the flipped and non-flipped results.



(a) DeVry University. (b) Grand Canyon University. (c) Keiser University.

Figure 7: Realistic ad creatives used for for-profit schools with historical practices of predatory marketing. The ad texts and images are taken from each school's page on Meta's ad library.