



DS-GA 3001.009: Responsible Data Science

Transparency: Online Ad Delivery

Prof. Julia Stoyanovich
Center for Data Science
Computer Science and Engineering at Tandon

@stoyanoj

<http://stoyanovich.org/>
<https://dataresponsibly.github.io/>

Transparency themes

- **Explaining black-box models**
 - **LIME**: local interpretable explanations [Ribeiro et al., KDD 2016]
 - **QII**: causal influence of features on outcomes [Datta et al., SSP 2016]
 - **SHAP**: Shapley additive explanations [Lundberg and Lee, NeurIPS 2017]
- **Online ad targeting**
 - Racially identifying names [Sweeney, CACM 2013]
 - Ad Fisher [Datta et al., PETS 2015]
 - Discrimination through optimization [Ali et al., CSCW 2019]
- **Interpretability**

Racially identifying names

[Latanya Sweeney; CACM 2013]



Ads by Google

[Latanya Sweeney, Arrested?](#)
1) Enter Name and State. 2) Access F
Checks Instantly.
www.instantcheckmate.com/

[Latanya Sweeney](#)
Public Records Found For: Latanya S
www.publicrecords.com/

[La Tanya](#)

LATANYA SWEENEY
1420 Centre Ave
Pittsburgh, PA 15219
DOB: Oct 27, 1959 (53 years old)

CERTIFIED

Personal
Name, aliases, birthdate, phone numbers, etc.

Location
Detailed address history and related data, maps, etc.

Criminal History
Rate This Content: ★★★★★
This section contains possible citation, arrest, and criminal records for the subject of this report. While our database does contain hundreds of millions of arrest records, different counties have different rules regarding what information they will and will not release.
We share with you as much information as we possibly can, but a clean slate here should not be interpreted as a guarantee that Latanya Sweeney has never been arrested; it simply means that we were not able to locate any matching arrest records in the data that is available to us.

Possible Matching Arrest Records

| Name | County and State | Offenses | View Details |
|--|------------------|----------|--------------|
| No matching arrest records were found. | | | |

Racism is Poisoning Online Ad Delivery, Says Harvard Professor

Google searches involving black-sounding names are more likely to serve up ads suggestive of a criminal record than white-sounding names, says computer scientist

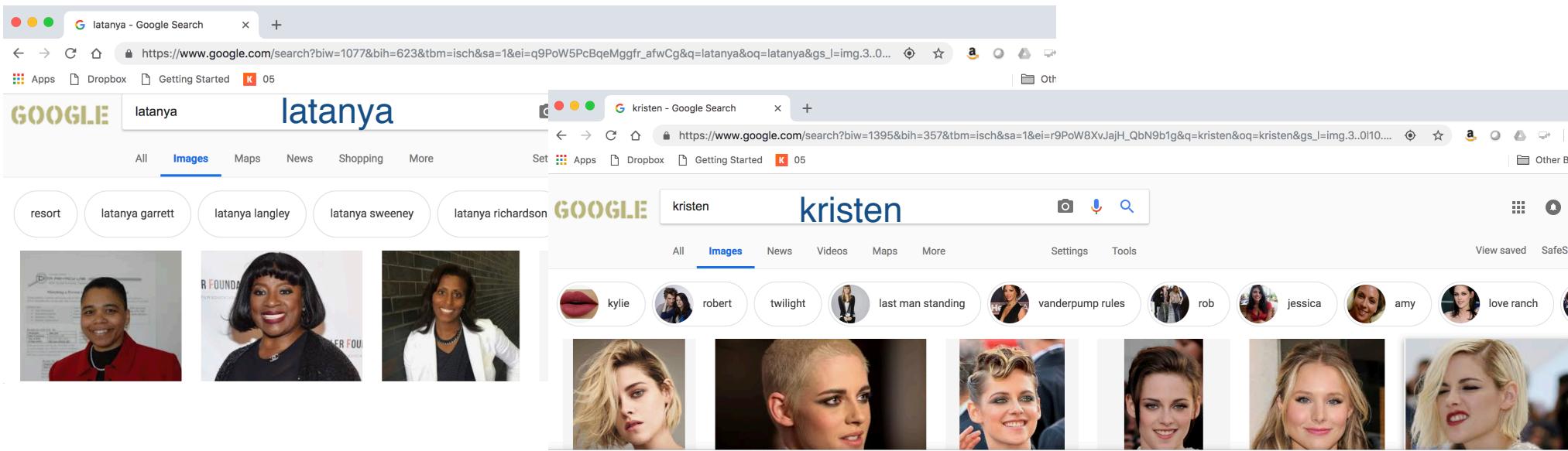
racially identifying names trigger ads suggestive of a criminal record

<https://www.technologyreview.com/s/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/>

Observations

[Latanya Sweeney; CACM 2013]

- Ads suggestive of a criminal record, linking to Instant Checkmate, appear on [google.com](#) and [reuters.com](#) in response to searches for “Latanya Sweeney”, “Latanya Farrell” and “Latanya Lockett”*
- No Instant Checkmate ads when searching for “Kristen Haring”, “Kristen Sparrow”* and “Kristen Lindquist”*
- * next to a name associated with an actual arrest record



Racially identifying names: details

[Latanya Sweeney; CACM 2013]

"A greater percentage of Instant Checkmate ads having the word arrest in ad text appeared for black-identifying first names than for white-identifying first names within professional and netizen subsets, too. On Reuters.com, which hosts Google AdSense ads, **a black-identifying name was 25% more likely to generate an ad suggestive of an arrest record.**"

More than 1,100 Instant Checkmate ads appeared on Reuters.com, with 488 having black-identifying first names; of these, 60% used arrest in the ad text. Of the 638 ads displayed with white-identifying names, 48% used arrest. This difference is statistically significant, with less than a 0.1% probability that the data can be explained by chance (chi-square test: $\chi^2 (1)=14.32$, $p < 0.001$).

The EEOC's and U.S. Department of Labor's adverse impact test for measuring discrimination is 77 in this case, so if this were an employment situation, a charge of discrimination might result. (The adverse impact test uses the ratio of neutral ads, or 100 minus the percentages given, to compute disparity: $100-60=40$ and $100-48=52$; dividing 40 by 52 equals 77.)

Why is this happening?

[Latanya Sweeney; CACM 2013]

Possible explanations (from Latanya Sweeney):

- Does Instant Checkmate serve ads specifically for black-identifying names?
- Is Google's AdSense explicitly biased in this way?
- Does Google's AdSense learn racial bias based on click-through rates?

How do we know which explanation is right?

We need transparency!

Response

<https://www.technologyreview.com/s/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/>

In response to this blog post, a **Google** spokesperson sends the following statement:

“AdWords does not conduct any racial profiling. We also have a strict violence policy which states that we will not allow ads that discriminate against an organisation, person or group of people. It is up to individual advertisers to choose which keywords they want to choose to trigger their ads.”



Instantcheckmate.com sends the following statement:

“As a point of fact, Instant Checkmate would like to state we have never engaged in racial profiling in Google AdWords. **We have the technology in place to even connect a name with a race** but we have no desire to do so. We have no desire to do any attempt to do so. The very idea is contrary to our company principles and values.”



Who is responsible?

- Who benefits?
- Who is harmed?
- What does the law say?
- Who is in a position to mitigate?

transparency responsibility trust

Gender discrimination in online job ads



Automated Experiments on Ad Privacy Settings (AdFisher)

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

Online job ads



Samuel Gibbs

Wednesday 8 July 2015 11.29 BST

Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs



One experiment showed that Google displayed adverts for a career coaching service for executive jobs 1,852 times to the male group and only 318 times to the female group. Photograph: Alamy

The AdFisher tool simulated job seekers that did not differ in browsing behavior, preferences or demographic characteristics, except in gender.

One experiment showed that Google displayed ads for a career coaching service for “\$200k+” executive jobs **1,852 times to the male group and only 318 times to the female group**. Another experiment, in July 2014, showed a similar trend but was not statistically significant.

<https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study>

Ad targeting online

- **Users** browse the Web, consume content, consume ads (see / click / purchase)
- **Content providers** (or **publishers**) host online content that often includes ads. They outsource ad placement to third-party ad networks
- **Advertisers** seek to place their ads on publishers' websites
- **Ad networks** track users across sites, to get a global view of users' behaviors. They connect advertisers and publishers

Google ad settings

Google ad settings aims to provide **transparency** / give **control to users** over the ads that they see

Your Google profile

The screenshot shows the Google Ad Settings interface. At the top, there's a "Your Google profile" section with a blue circular icon containing a white silhouette of a person, labeled "Gender". Next to it is another blue circular icon with the text "35-44", labeled "Age". Below this, there's a section titled "Ads based on your interests" with a green toggle switch set to "ON". Underneath, a sub-section says "Improve your ad experience when you are signed in to Google sites". The main content area is divided into two columns: "With Ads based on your interests ON" (green background) and "With Ads based on your interests OFF" (grey background). The "ON" column lists benefits like ads being delivered based on search queries and YouTube watch history. The "OFF" column lists what happens when ads are turned off, such as seeing ads based on general location.

| With Ads based on your interests ON | With Ads based on your interests OFF |
|---|---|
| <ul style="list-style-type: none">The ads you see will be delivered based on your prior search queries, the videos you've watched on YouTube, as well as other information associated with your account, such as your age range or genderOn some Google sites like YouTube, you will see ads related to your interests, which you can edit at any time by visiting this pageYou can block some ads that you don't want to see | <ul style="list-style-type: none">You will still see ads and they may be based on your general location (such as city or state)Ads will not be based on data Google has associated with your Google Account, and so may be less relevantYou will no longer be able to edit your interestsAll the advertising interests associated with your Google Account will be deleted |

<http://www.google.com/settings/ads>

Google ad settings

Do users truly have transparency / choice or is this a placebo button?

The screenshot shows the 'Control your Google ads' page. At the top, it says 'Control your Google ads' and 'You can control the ads that are delivered to you based on your Google Account, across devices, by editing these settings. These ads are more likely to be useful and relevant to you.' Below this, under 'Your interests', there are two columns of checkboxes:

| Interest Category | Interests |
|--------------------------|--|
| Action & Adventure Films | Cats |
| Cooking & Recipes | Fitness |
| History | Hybrid & Alternative Vehicles |
| Hygiene & Toiletries | Make-Up & Cosmetics |
| Mobile Phones | Parenting |
| Phone Service Providers | Recording Industry |
| Reggaeton | Search Engine Optimization & Marketing |
| Vehicle Brands | |

At the bottom left, there is a blue button labeled '+ ADD NEW INTEREST'. To the right of the interests, a link says 'WHERE DID THESE COME FROM?'. A dark gray callout box contains the text: 'These interests are derived from your activity on Google sites, such as the videos you've watched on YouTube. This does not include Gmail interests, which are used only for ads within Gmail. [Learn more](#)'.

<http://www.google.com/settings/ads>

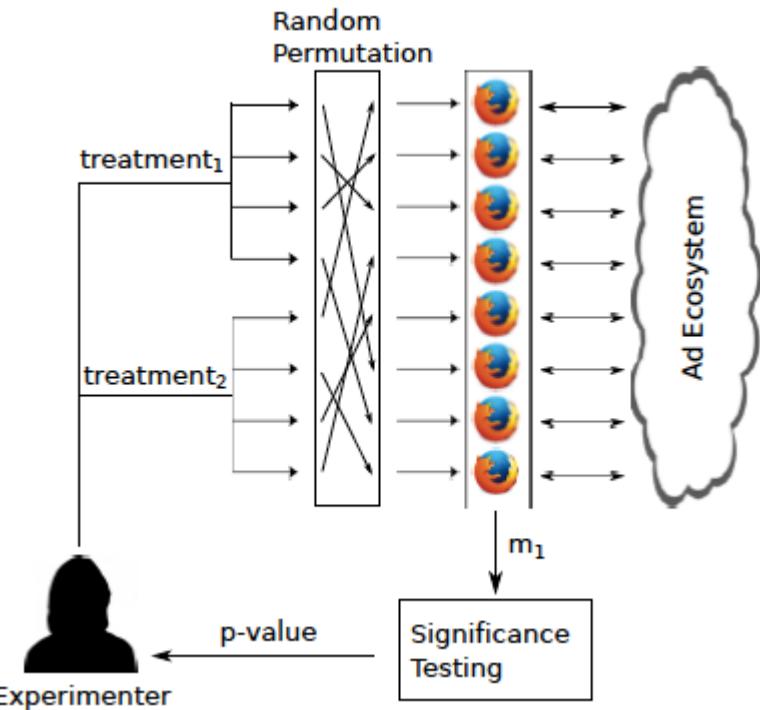
AdFisher

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

From anecdotal evidence to statistical insight: How do user behaviors, ads and ad settings interact?

Automated randomized controlled experiments for studying online tracking

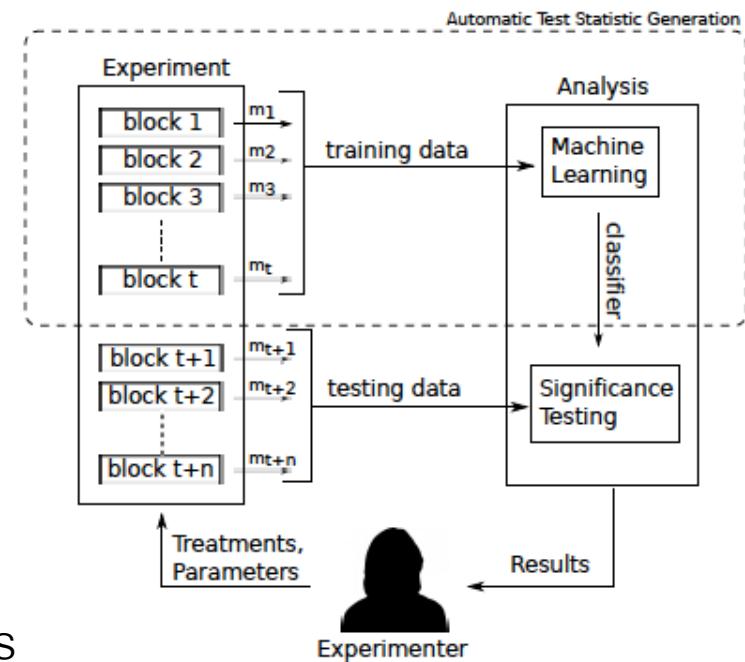
Individual data use transparency: ad network must share the information it uses about the user to select which ads to serve to him



AdFisher: methodology

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

- Browser-based experiments, simulated users
 - **input:** (1) visits to content providing websites; (2) interactions with Google Ad Settings
 - **output:** (1) ads shown to users by Google; (2) change in Google Ad Settings
- Fisher randomized hypothesis testing
 - **null hypothesis** inputs do not affect outputs
 - control and experimental treatments
 - AdFisher can help select a test statistic



AdFisher: gender and jobs

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

Non-discrimination: Users differing only in protected attributes are treated similarly

Causal test: Find that a protected attribute changes ads

Experiment: **gender and jobs**

Specify gender (male/female) in Ad Settings, simulate interest in jobs by visiting employment sites, collect ads from Times of India or the Guardian

Result: males were shown ads for higher-paying jobs significantly more often than females (1852 vs. 318)

violation

AdFisher: substance abuse

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

Transparency: User can view data about him used for ad selection

Causal test: Find attribute that changes ads but not settings

Experiment 2: **substance abuse**

Simulate interest in substance abuse in the experimental group but not in the control group, check for differences in Ad Settings, collect ads from Times of India

Result: no difference in Ad Settings between the groups, yet significant differences in what ads are served: rehab vs. stocks + driving jobs

violation

AdFisher: online dating

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

Ad choice: Removing an interest decreases the number of ads related to that interest.

Causal test: Find that removing an interest causes a decrease in related ads

Experiment 3: **online dating**

Simulate interest in online dating in both groups, remove “Dating & Personals” from the interests on Ad Settings for experimental group, collect ads

Result: members of experimental group do not get ads related to dating, while members of the control group do

compliance

Recall the set-up

[A. Datta, A. Datta, J. Makagon, D. Mulligan, M. Tschantz; *FAT* 2018*]

- **Users** browse the Web, consume content, consume ads (see / click / purchase)
- **Content providers** (or **publishers**) host online content that often includes ads. They outsource ad placement to third-party ad networks
- **Advertisers** seek to place their ads on publishers' websites
- **Ad networks** track users across sites, to get a global view of users' behaviors. They connect advertisers and publishers

Why are males seeing ads for high-paying jobs more often?

What is causing gender-based discrimination?

(1) who is responsible and (2) how is discrimination enacted?

Who is responsible?

[A. Datta, A. Datta, J. Makagon, D. Mulligan, M. Tschantz; *FAT* 2018*]

- **Google alone:** explicitly programming the system to show the ad less often to females, e.g., based on independent evaluation of demographic appeal of product (**explicit and intentional discrimination**)
- **The advertiser:** targeting of the ad through explicit use of demographic categories (**explicit and intentional**), selection of proxies (**hidden and intentional**), or through those choices without intent (**unconscious selection bias**), and **Google** respecting these targeting criteria
- **Other advertisers:** others outbid our advertiser when targeting to females
- **Other users:** Male and female users behaving differently to ads, and Google learning to predict this behavior

How is targeting done?

[A. Datta, A. Datta, J. Makagon, D. Mulligan, M. Tschantz; *FAT* 2018*]

- on gender directly
- on a proxy of gender, i.e., on a known correlate of gender because it is a correlate
- on a known correlate of gender, but not because it is a correlate
- on an unknown correlate of gender

Secretary Jobs
possibility.cylab.cmu.edu/jobs
Full time jobs in Florida
Excellent pay and relocation

(a)

Truck Driving Jobs
possibility.cylab.cmu.edu/jobs
Full time jobs in Florida
Excellent pay and relocation

(b)

Figure 1: Ads approved by Google in 2015. The ad in the left (right) column was targeted to women (men).

experiments show that is possible to use Google AdWords to target on gender

“This finding demonstrates that an advertiser with discriminatory intentions can use the AdWords platform to serve employment related ads disparately on gender.”

What are the legal ramifications?

[A. Datta, A. Datta, J. Makagon, D. Mulligan, M. Tschantz; *FAT* 2018*]

- Each actor in the advertising ecosystem may have contributed inputs that produced the effect
- **It is impossible to know, without additional information, what the different actors - other than the consumers of the ads - did or did not do**
- In particular, impossible to asses intent, which *may* be necessary to asses the extent of legal liability. Or it may not!
 - **Title VII of the 1964 Civil Rights Act** makes it unlawful to discriminate based on sex in several stages of employment. It includes an **advertising prohibition** (think sex-specific *help wanted* columns in a newspaper), which does not turn on intent
 - **Title VII does not directly apply here** because it is limited in scope to employers, labor organizations, employment agencies, joint labor-management committees
 - **Fair Housing Act (FHA)** is perhaps a better guide than Title VII, limiting both content and activities that target advertisement based on protected attributes

In the news: Facebook ads

POLICY \ US & WORLD \ TECH \

THE VERGE

83

Facebook has been charged with housing discrimination by the US government

'Facebook is discriminating against people based upon who they are and where they live,' says HUD secretary

By Russell Brandom | Mar 28, 2019, 7:51am EDT

The Department of Housing and Urban Development has [filed charges](#) against Facebook for housing discrimination, escalating the company's ongoing fight over discrimination in its ad targeting system. The charges build on [a complaint filed in August](#), finding that there is reasonable cause to believe Facebook has served ads that violate the Fair Housing Act.

ProPublica first raised concerns over housing discrimination on Facebook in 2016, when reporters found that [the “ethnic affinities” tool](#) could be used to exclude black or Hispanic users from seeing specific ads. If those ads were for housing or employment opportunities, the targeting could easily violate federal law. At the time, Facebook had no internal safeguards in place to prevent such targeting.

<https://www.theverge.com/2019/3/28/18285178/facebook-hud-lawsuit-fair-housing-discrimination>

In the news: Facebook ads

POLICY \ US & WORLD \ TECH \

THE VERGE

83

Facebook has been charged with housing discrimination by the US government

'Facebook is discriminating against people based upon who they are and where they live,' says HUD secretary

By Russell Brandom | Mar 28, 2019, 7:51am EDT

Facebook has struggled to effectively address the possibility of discriminatory ad targeting. The company pledged to step up anti-discrimination enforcement in the wake of *ProPublica's* reporting, but a follow-up report in 2017 found the same problems persisted nearly a year later.

According to the HUD complaint, many of the options for targeting or excluding audiences are shockingly direct, including a map tool that explicitly echoes redlining practices. “[Facebook] has provided a toggle button that enables advertisers to exclude men or women from seeing an ad, a search-box to exclude people who do not speak a specific language from seeing an ad, and a map tool to exclude people who live in a specified area from seeing an ad by drawing a red line around that area,” the complaint reads.

<https://www.theverge.com/2019/3/28/18285178/facebook-hud-lawsuit-fair-housing-discrimination>

In the news: Google and Twitter ads

POLICY \ US & WORLD \ TECH

THE VERGE

83

Facebook has been charged with housing discrimination by the US government

'Facebook is discriminating against people based upon who they are and where they live,' says HUD secretary

By Russell Brandom | Mar 28, 2019, 7:51am EDT

This is the first federal discrimination lawsuit to deal with racial bias in targeted advertising, a milestone that lawyers at HUD said was overdue. "Even as we confront new technologies, the fair housing laws enacted over half a century ago remain clear—discrimination in housing-related advertising is against the law," said HUD General

POLICY \ US & WORLD \ TECH

HUD reportedly also investigating Google and Twitter in housing discrimination probe

By Adi Robertson | @thedextriarchy | Mar 28, 2019, 3:52pm EDT

<https://www.theverge.com/2019/3/28/18285899/housing-urban-development-hud-facebook-lawsuit-google-twitter>

Discrimination in Facebook's ad delivery



Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; *arXiv 2019*]

Discrimination in Facebook's ad delivery

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; *arXiv 2019*]

- Follow-up work on AdFisher (Google ads, gender-based discrimination for the purposes of employment) ascertained that it was possible to target on gender for job ads
- Platforms have since taken steps to address such blatant violations

*“... Facebook currently has several policies in place to avoid discrimination for certain types of ads. Facebook also recently **built tools to automatically detect ads offering housing, employment, and credit**, and pledged to prevent the use of certain targeting categories with those ads. Additionally, Facebook relies on advertisers to self-certify that they are not in violation of Facebook’s advertising policy prohibitions against discriminatory practices. More recently, in order to settle multiple lawsuits stemming from these reports, **Facebook stated that they will soon no longer allow age, gender, or ZIP code-based targeting for housing, employment or credit ads**, and that they would also block other detailed targeting attributes that are “describing or appearing to relate to protected classes”.*

- Yet, the question still remains: **Does the ad delivery platform itself embed discriminatory outcomes?**

Facebook ad delivery

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; *arXiv 2019*]

Part 1: ad creation

- ad contents
- audience selection
- bidding strategy

Part 2: ad delivery

For every opportunity to show a user an ad (e.g., **an ad slot** is available as the user is browsing the service), the ad platform will run an **ad auction** to determine, from among all of the ads that include the current user in the audience, which ad should be shown.

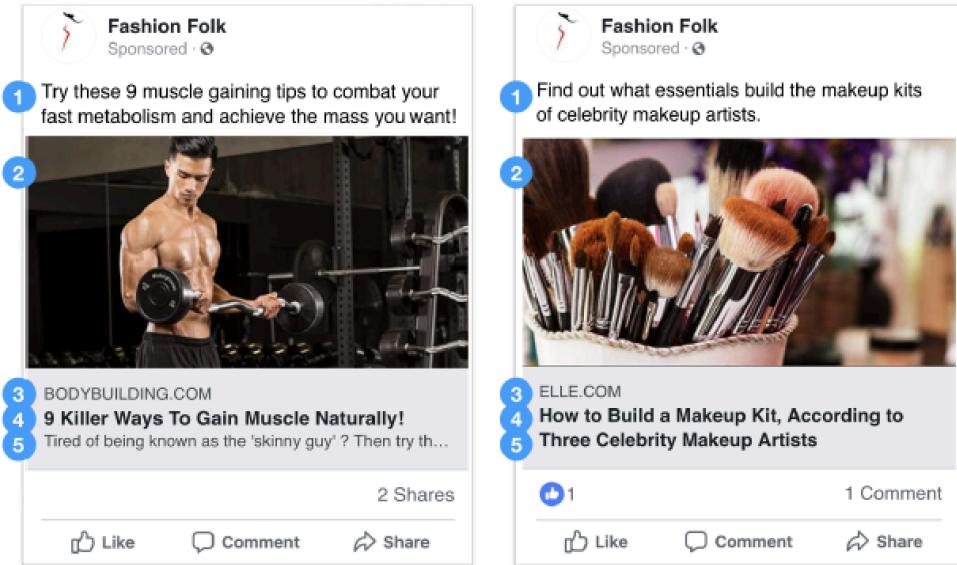


Figure 1: Each ad has five elements that the advertiser can control: (1) the ad headline and text, entered manually by the advertiser, (2) the images and/or videos, (3) the domain, pulled automatically from the HTML meta property `og:site_name` of the destination URL, (4) the title, pulled automatically from the HTML meta property `og:title` of the destination URL, and (5) the description from meta property `og:description` of the destination URL.

Facebook ad delivery

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; *arXiv 2019*]

Part 1: ad creation

- ad contents
- audience selection
- bidding strategy

Part 2: ad delivery

For every opportunity to show a user an ad (e.g., **an ad slot** is available as the user is browsing the service), the ad platform will run an **ad auction** to determine, from among all of the ads that include the current user in the audience, which ad should be shown.

When Facebook has ad slots available, it runs an ad auction among the active advertisements bidding for that user. However, **the auction does not just use the bids placed by the advertisers**; Facebook says:

*“The ad that wins an auction and gets shown is the one with the highest **total value**. Total value isn’t how much an advertiser is willing to pay us to show their ad. It’s combination of 3 major factors: (1) Bid, (2) Estimated action rates, and (3) Ad quality and relevance.”*

*“During ad set creation, you chose a target audience ... and an optimization event ... **We show your ad to people in that target audience who are likely to get you that optimization event.**”*

Facebook ad delivery: insights

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; *arXiv 2019*]

Facebook ad delivery results can be skewed **in ways that advertisers do not intend**

- Skew can arise due to:
 - financial optimization effects
 - the ad delivery platform's predictions about the relevance of its ads to different user categories
- What contributes to the skew?
 - ad content - both text and images, which are likely automatically analyzed by Facebook
 - advertiser budget

Skew was observed along gender and racial lines, in ads for employment and housing opportunities

Budget impacts demographics

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; *arXiv 2019*]

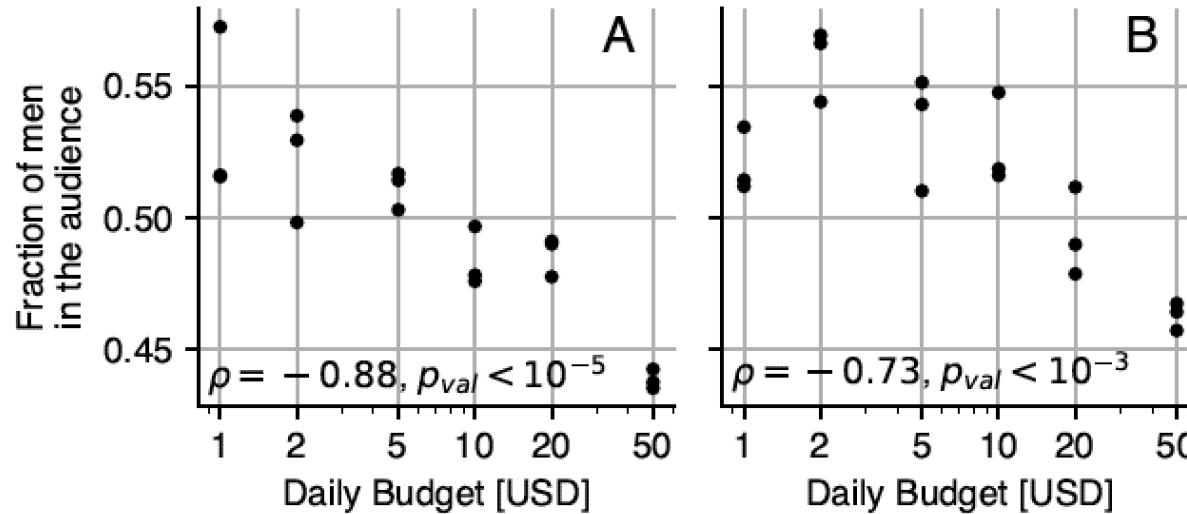


Figure 2: Gender distributions of the audience depend on the daily budget of an ad, with higher budgets leading to a higher fraction of women. The left graph shows an experiment where we target all users located in the U.S.; the right graph shows an experiment where we target our random phone number custom audiences.

Ad creative impacts demographics

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; *arXiv 2019*]

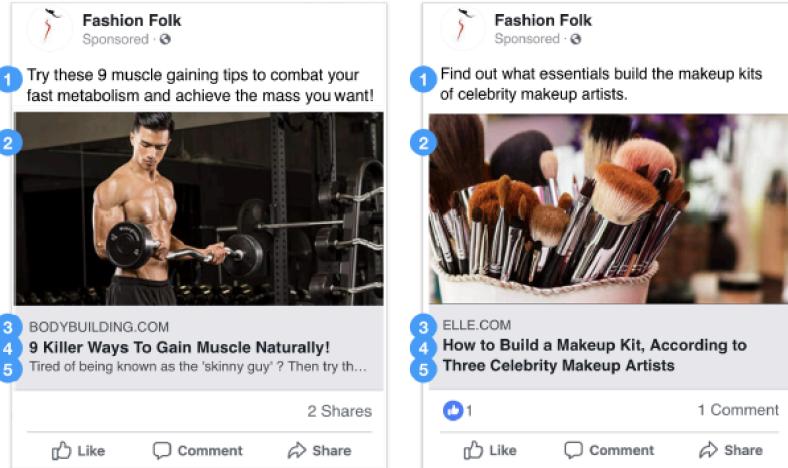


Figure 1: Each ad has five elements that the advertiser can control: (1) the ad headline and text, entered manually by the advertiser, (2) the images and/or videos, (3) the domain, pulled automatically from the HTML meta property `og:site_name` of the destination URL, (4) the title, pulled automatically from the HTML meta property `og:title` of the destination URL, and (5) the description from meta property `og:description` of the destination URL.

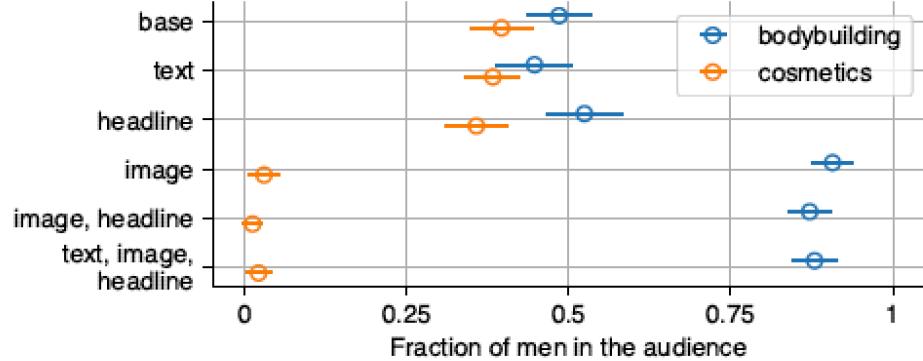


Figure 3: “Base” ad contains a link to a page about either bodybuilding or cosmetics, a blank image, no text, or headline. There is a small difference in the fraction of male users for the base ads, and adding setting the “text” only decreases it. Setting the “headline” sets the two ads apart but the audience of each is still not significantly different than that of the base version. Finally, setting the ad “image” causes drastic changes: the bodybuilding ad is shown to a 91% male audience, the cosmetics ad is shown to very few men, despite the same target audience.

Transparent images are still targeted!

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; *arXiv 2019*]

| No. | Male | | Female | |
|-----|---|-----------|---|-----------|
| | Visible | Invisible | Visible | Invisible |
| 1 |  | |  | |
| 2 |  | |  | |
| 3 |  | |  | |
| 4 |  | |  | |
| 5 |  | |  | |

Table 2: Diagram of the images used in the transparency experiments. Shown are the five stereotypical male and female images, along with the same images with a 98% alpha channel, denoted as invisible. The images with the alpha channel are almost invisible to humans, but are still delivered in a skewed manner.

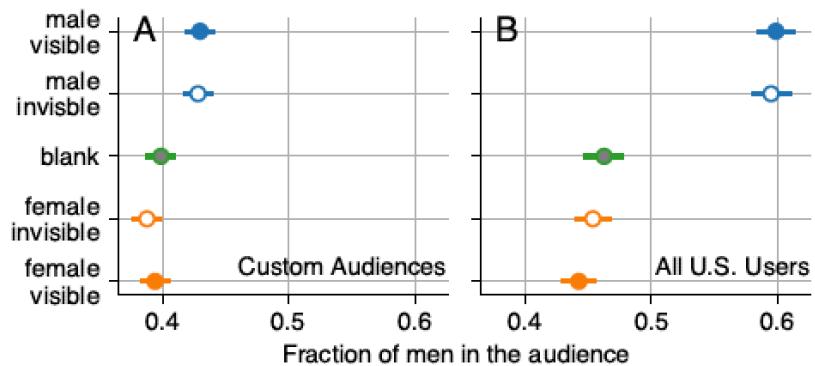


Figure 6: Ad delivery to ads with the images from Table 2, targeting general US audience as well as the random phone number custom audience. The solid markers are visible images, and the hollow markers are the same images with 98% opacity. Also shown is the delivery to truly white images ("blank"). We can observe that a difference in ad delivery exists, and that that difference is statistically significant between the male and female invisible images. This suggests that automated image classification is taking place.

Racial discrimination in housing ads

[M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke; *arXiv 2019*]

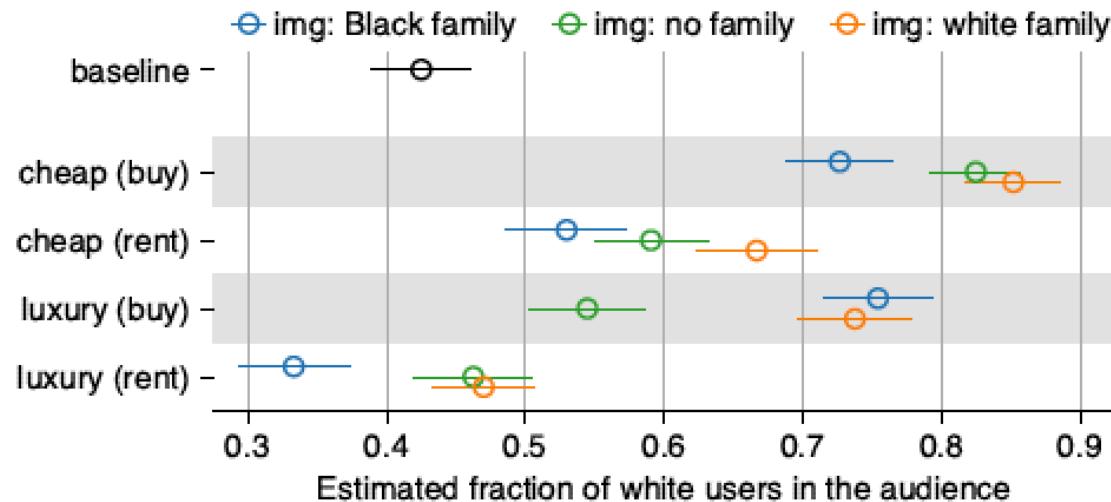


Figure 9: Results for housing ads, showing a breakdown in the ad delivery audience by race. Despite being targeted in the same manner, using the same bidding strategy, and being run at the same time, we observe significant skew in the makeup of the audience to whom the ad is delivered (ranging from over 85% white users to over 65% Black users).