Problematic Advertising and its Disparate Exposure on Facebook

Muhammad Ali
w. Angelica Goetzen, Alan Mislove, Elissa M. Redmiles, Piotr Sapiezynski
USENIX Security ‘23
Anaheim, CA
August 11, 2023
Ad delivery can have discriminatory outcomes

Discrimination through optimization: How Facebook’s ad delivery can lead to biased outcomes

MUHAMMAD ALI*, Northeastern University
PIOTR SAPIEZYNSKI*, Northeastern University
MIRANDA BOGEN, Upturn
ALEKSANDRA KOROLOVA, University of Southern California
ALAN MISLOVE, Northeastern University
AARON RIEKE, Upturn

Facebook’s ad system seems to discriminate by race and gender
New research shows that Facebook's ad-distribution software is disturbingly biased

Facebook’s ad-serving algorithm discriminates by gender and race
Even if an advertiser is well-intentioned, the algorithm still prefers certain groups of people over others.
By Karen Hao
April 5, 2019

Facebook’s ad delivery could be inherently discriminatory, researchers say
Meta also will develop a new system to address racial and other disparities caused by its use of personalization algorithms in its ad delivery system for housing ads.

We are building into our ads system a method — referred to in the settlement as the “variance reduction system” — designed to make sure the audience that ends up seeing a housing ad more closely reflects the eligible targeted audience for that ad.
Are disparate outcomes of advertising solved?

• Maybe for housing ads alone…
• What about domains not protected by law? e.g. scams, clickbait, vulnerabilities?
• What about variances in individual experiences?
User-informed “bad” ads exist in the marketplace

In this talk: Problematic advertising and its disparate exposure on Facebook

Research Questions

1. What types of ads do users consider problematic?  
   RQ1

2. Are there skews in the distribution of such ads?  
   RQ2

3. Who is responsible for skews?  
The advertisers or ad delivery/personalization?  
   RQ3
Methodology

Panel of Facebook Users (n = 132)

Nov. 2021 — Sep. 2022 (11 months); rolling recruitment; each participant stays 3 months

NEU Ad Observer (Chrome + Firefox)

Ad “diets” and targeting data

Annotation Codebook (8 Ad Types)

8,701 ads

Panel of Facebook Users

NEU Ad Observer

Ad “diets” and targeting data

Annotation Codebook

Monthly Surveys

88,509

32,587 (~37%)

subsample 200 ads/participant/mo.
Categorizing ads with a codebook

Pilot Data Collection

Mixture of inductive qualitative coding from collected ads + deductive analysis of prior work and Facebook policies

- Healthcare
- Opportunity
- Neutral
- Sensitive
  - Financial
  - Gambling
  - Alcohol
  - Weight loss
  - Online pharmacies
  - Prescription + over-the-counter drugs
- Deceptive
- Potentially Prohibited
- Clickbait

[Zeng et al., CHI ’21]


<table>
<thead>
<tr>
<th>Code</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>20,596</td>
<td>68.52</td>
</tr>
<tr>
<td>Healthcare</td>
<td>3564</td>
<td>11.86</td>
</tr>
<tr>
<td>Opportunity</td>
<td>2267</td>
<td>7.54</td>
</tr>
<tr>
<td>Sensitive: Financial</td>
<td>1429</td>
<td>4.75</td>
</tr>
<tr>
<td>Sensitive: Other</td>
<td>631</td>
<td>2.10</td>
</tr>
<tr>
<td>Clickbait</td>
<td>1182</td>
<td>3.93</td>
</tr>
<tr>
<td>Deceptive</td>
<td>542</td>
<td>1.80</td>
</tr>
<tr>
<td>Potentially Prohibited</td>
<td>253</td>
<td>0.84</td>
</tr>
</tbody>
</table>
RQ1: Which categories of ads do participants perceive as problematic?

Monthly Surveys
RQ1: What categories of ads do participants perceive as problematic?

Monthly Surveys

Reasons of dislike? (compared to Neutral)

- They have higher odds of being considered irrelevant, clickbait, scam.
- Sensitive ads have higher odds of being disliked due to the advertiser of the product.
RQ2: Are there skews in the distribution of problematic ads?
RQ2: Are there skews in the distribution of problematic ads?

Participants’ Ad “Diet”

Fraction of Participant’s Ad Diet

Ad Category
Problematic
Neutral
RQ2: Are there skews in the distribution of problematic ads?

Yes. Do they relate to participant demographics?

Linear Regression

\[ \text{e.g. fraction_clickbait} \sim \text{woman} + \text{hispanic} + \text{older} + \ldots \]
How are problematic ad skews related to demographics?

Ad Diets, Linear Regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate (β) [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Problematic</td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
</tr>
<tr>
<td>Gender: Woman</td>
<td></td>
</tr>
<tr>
<td>Race: Black</td>
<td></td>
</tr>
<tr>
<td>Race: Asian</td>
<td></td>
</tr>
<tr>
<td>Ethnicity: Hispanic</td>
<td></td>
</tr>
<tr>
<td>Education: college and above</td>
<td></td>
</tr>
<tr>
<td>Age: Gen-X and older</td>
<td></td>
</tr>
</tbody>
</table>
How are problematic ad skews related to demographics?

Older participants see 5.1 pp more Problematic ads. Participants identifying as women see 6.4 pp fewer Problematic ads.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate (β) [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.12*** [0.09, 0.15]</td>
</tr>
<tr>
<td>Gender: Woman</td>
<td>-0.064*** [-0.09, -0.04]</td>
</tr>
<tr>
<td>Race: Black</td>
<td>0.025 [-0.01, 0.06]</td>
</tr>
<tr>
<td>Race: Asian</td>
<td>-0.002 [-0.04, 0.04]</td>
</tr>
<tr>
<td>Ethnicity: Hispanic</td>
<td>0.023 [-0.03, 0.08]</td>
</tr>
<tr>
<td>Education: college and above</td>
<td>0.01 [0.001, 0.01]</td>
</tr>
<tr>
<td>Age: Gen-X and older</td>
<td>0.051*** [0.02, 0.08]</td>
</tr>
</tbody>
</table>
How are problematic ad skews related to demographics?

Older participants see 5.1 pp more Problematic ads—including Deceptive and Clickbait content. Black participants see 1.3 pp more Clickbait than other races. Participants identifying as women see 6.4 pp fewer Problematic ads.
How are problematic ad skews related to demographics?

**Older participants** see **5.1 pp more** Problematic ads—including Deceptive and Clickbait content.

**Black participants** see **1.3 pp more** Clickbait than other races.

Participants identifying as **women** see **6.4 pp fewer** Problematic ads—largely due to lower exposure to Financial ads.
RQ3: Who is responsible for skews? Advertisers or algorithms?
Ad targeting data, “Why am I seeing this?”

Ad Targeting
- target audience
- budget
- campaign objective

Ad Delivery
- personalization / relevance judgments
- user auction
- final audience $\subseteq$ target audience

Older participants see 5.1 pp more Problematic ads—including Deceptive and Clickbait content.
RQ3: Who is responsible for skews? Advertisers or algorithms?

Deep-dive into age: are advertisers targeting older participants?

Age targeting has high usage in our data, 49.7% ads. Compared to only 12.1% ads using gender targeting.
RQ3: Who is responsible for skews? Advertisers or algorithms?

Deep-dive into age: are advertisers targeting older participants?

Age targeting has high usage in our data, 49.7% ads. Compared to only 12.1% ads using gender targeting.
RQ3: Who is responsible for skews? Advertisers or algorithms?

Deep-dive into age: are advertisers targeting older participants?

Age targeting has high usage in our data, 49.7% ads. Compared to only 12.1% ads using gender targeting.

Advertisers’ targeting aligns with observed skews: **Clickbait** and **Deceptive** is actively targeted to older users. **Pot. Prohibited** is targeted less to older users.

So advertisers are clearly responsible, what about algorithms?
Isolating algorithm’s influence: ads with “default” targeting

Advertiser has no preference whatsoever

21.2% ads target to all adults in the US, i.e. 267 million users

Linear Regression (as before) on subset of default targeting ads

e.g. \textit{fraction\_clickbait} \sim \text{woman} + \text{hispanic} + \text{older} + \ldots
Isolating algorithm’s influence: ads with “default” targeting

Advertiser has no preference whatsoever

Linear Regression (as before) on subset of default targeting ads

e.g. fraction_clickbait \sim woman + hispanic + older + ...

<table>
<thead>
<tr>
<th>Variable</th>
<th>Problematic</th>
<th>Pot. Prohibited</th>
<th>Deceptive</th>
<th>Clickbait</th>
<th>Sensitive: Financial</th>
<th>Sensitive: Other</th>
</tr>
</thead>
</table>
Isolating algorithm’s influence: ads with “default” targeting
Advertiser has no preference whatsoever

<table>
<thead>
<tr>
<th>Variable</th>
<th>Problematic</th>
<th>Pot. Prohibited</th>
<th>Deceptive</th>
<th>Clickbait</th>
<th>Sensitive: Financial</th>
<th>Sensitive: Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.191***</td>
<td>[0.13, 0.26]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender: Woman</td>
<td>-0.059*</td>
<td>[-0.11, -0.01]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race: Black</td>
<td>0.01</td>
<td>[-0.05, 0.07]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>race: Asian</td>
<td>-0.019</td>
<td>[-0.1, 0.06]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity: Hispanic</td>
<td>0.017</td>
<td>[-0.08, 0.12]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education: college and above</td>
<td>-0.033</td>
<td>[-0.09, 0.02]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age: Gen-X and older</td>
<td>0.077**</td>
<td>[0.02, 0.13]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Even within ads with the broadest possible targeting:

**Older participants** (still) see **7.7 pp more** Problematic ads.
Participants identifying as **women** (still) see **5.9 pp fewer** Problematic ads.
Isolating algorithm’s influence: ads with “default” targeting
Advertiser has no preference whatsoever

Even within ads with the broadest possible targeting:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Problematic</th>
<th>Pot. Prohibited</th>
<th>Deceptive</th>
<th>Clickbait</th>
<th>Sensitive: Financial</th>
<th>Sensitive: Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.191***</td>
<td>0.013***</td>
<td>0.014*</td>
<td>0.023</td>
<td>0.133***</td>
<td>0.009*</td>
</tr>
<tr>
<td></td>
<td>[0.13, 0.26]</td>
<td>[0.01, 0.02]</td>
<td>[0.03]</td>
<td>[-0.01, 0.05]</td>
<td>[0.08, 0.18]</td>
<td>[0, 0.02]</td>
</tr>
<tr>
<td>Gender: Woman</td>
<td>-0.059*</td>
<td>-0.006*</td>
<td>-0.007</td>
<td>-0.003</td>
<td>-0.046*</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>[-0.11, -0.01]</td>
<td>[-0.01, 0]</td>
<td>[-0.02, 0]</td>
<td>[-0.03, 0.02]</td>
<td>[-0.09, 0]</td>
<td>[0, 0.01]</td>
</tr>
<tr>
<td>Race: Black</td>
<td>0.011</td>
<td>0.002</td>
<td>0.007</td>
<td>0.011</td>
<td>-0.007</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>[0.05, 0.07]</td>
<td>[0, 0.01]</td>
<td>[0.01, 0.02]</td>
<td>[-0.02, 0.04]</td>
<td>[-0.06, 0.04]</td>
<td>[-0.01, 0]</td>
</tr>
<tr>
<td>race: Asian</td>
<td>-0.019</td>
<td>-0.005</td>
<td>-0.003</td>
<td>-0.007</td>
<td>-0.003</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>[-0.1, 0.06]</td>
<td>[-0.01, 0]</td>
<td>[-0.02, 0.01]</td>
<td>[-0.04, 0.03]</td>
<td>[-0.07, 0.06]</td>
<td>[-0.01, 0.01]</td>
</tr>
<tr>
<td>Ethnicity: Hispanic</td>
<td>0.017</td>
<td>-0.009</td>
<td>0.028**</td>
<td>-0.021</td>
<td>0.027</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>[-0.08, 0.12]</td>
<td>[-0.02, 0]</td>
<td>[0.01, 0.05]</td>
<td>[-0.06, 0.02]</td>
<td>[-0.05, 0.11]</td>
<td>[-0.02, 0]</td>
</tr>
<tr>
<td>Education: college and above</td>
<td>-0.033</td>
<td>-0.002</td>
<td>0</td>
<td>0.005</td>
<td>-0.036</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>[-0.08, 0.02]</td>
<td>[-0.01, 0]</td>
<td>[-0.01, 0.01]</td>
<td>[0.02, 0.03]</td>
<td>[-0.08, 0.01]</td>
<td>[-0.01, 0.01]</td>
</tr>
<tr>
<td>Age: Gen-X and older</td>
<td>0.077**</td>
<td>-0.003</td>
<td>0.011</td>
<td>0.041**</td>
<td>0.034</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>[0.02, 0.13]</td>
<td>[-0.01, 0]</td>
<td>[0, 0.02]</td>
<td>[0.02, 0.06]</td>
<td>[-0.01, 0.08]</td>
<td>[-0.01, 0]</td>
</tr>
</tbody>
</table>

Older participants (still) see 7.7 pp more Problematic ads—4.1 pp more Clickbait content. Participants identifying as women (still) see 5.9 pp fewer Problematic ads—largely due to lower exposure to Financial ads. New effect: Hispanic participants see 2.8 pp more Deceptive than non-Hispanic participants.
Summary + takeaways

- First study of real user experiences with problematic ads—provides an understanding of disparate exposure through lived experiences

- Malicious advertisers are aware of vulnerable populations, and do use tools at their disposal to run ads

- Even if advertisers are not aware, personalization will roll out the red carpet

- Personalization and malicious advertisers together can expose vulnerable users to harmful content

- In addition to moderation, platforms might need to limit optimization as well—proposal: stop personalization altogether for problematic content

- Transparency is valuable, despite platforms being resistant to studies
Thank you, USENIX Security!

More results + discussion in full paper!

Questions?

ali.muh@northeastern.edu
@lukshmichowk
Backup Slides
## Panel Demographics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Recruited n</th>
<th>Recruited %</th>
<th>Active n</th>
<th>Active %</th>
<th>Census %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Female</td>
<td>96</td>
<td>52.17</td>
<td>71</td>
<td>53.79</td>
<td>50.5</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>86</td>
<td>46.74</td>
<td>59</td>
<td>44.70</td>
<td>49.5</td>
</tr>
<tr>
<td></td>
<td>Non-binary</td>
<td>2</td>
<td>1.09</td>
<td>2</td>
<td>1.52</td>
<td>–</td>
</tr>
<tr>
<td>Age</td>
<td>Younger than Gen-X</td>
<td>134</td>
<td>72.83</td>
<td>88</td>
<td>66.67</td>
<td>33.6</td>
</tr>
<tr>
<td></td>
<td>Gen-X and older</td>
<td>50</td>
<td>27.17</td>
<td>44</td>
<td>33.33</td>
<td>47.8</td>
</tr>
<tr>
<td>Race / Ethnicity</td>
<td>White</td>
<td>105</td>
<td>57.07</td>
<td>82</td>
<td>62.12</td>
<td>75.8</td>
</tr>
<tr>
<td></td>
<td>Latino/Hispanic</td>
<td>21</td>
<td>11.41</td>
<td>16</td>
<td>12.12</td>
<td>18.9</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>53</td>
<td>28.80</td>
<td>32</td>
<td>24.24</td>
<td>13.6</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
<td>21</td>
<td>11.41</td>
<td>16</td>
<td>12.12</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>3</td>
<td>1.63</td>
<td>3</td>
<td>2.27</td>
<td>–</td>
</tr>
<tr>
<td>Education</td>
<td>Below Bachelor’s</td>
<td>72</td>
<td>39.13</td>
<td>51</td>
<td>38.64</td>
<td>58.5</td>
</tr>
<tr>
<td></td>
<td>Bachelor’s or above</td>
<td>112</td>
<td>60.87</td>
<td>81</td>
<td>61.36</td>
<td>32.9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>184</strong></td>
<td></td>
<td></td>
<td><strong>132</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Demographics of panel participants.
Survey Instrument

Q1. How would you describe the advertised product/offer’s relevance to you?

[Completely Irrelevant] [Irrelevant] [Neutral] [Relevant] [Completely Relevant]

Q2. Which of the following, if any, describe your reasons for disliking this ad?

- It is irrelevant to me, or does not contain interesting information.
- I do not like the design of the ad.
- It contains clickbait, sensationalized, or shocking content.
- I do not trust this ad, it seems like a scam.
- I dislike the advertiser.
- I dislike the type of product being advertised.
- I find the content uncomfortable, offensive, or repulsive.
- I dislike the political nature of the ad.
- I find the ad pushy or it causes me to feel anxious.
- I cannot tell what is being advertised. (unclear)
- I do not dislike this ad.

Q3. Which of the following, if any, describe your reasons for liking this ad?

- The content is engaging, clever or amusing
- It is well designed or eye-catching.
- I am interested in what is being advertised.
- It is clear what product the ad is selling.
- I trust the ad, it looks authentic or trustworthy.
- I trust the advertiser.
- It is useful, interesting, or informative.
- It clearly looks like an ad and can be filtered out.
- I do not like this ad.

[Q2 and Q3 from Zeng et. al., CHI ’21]