Why am I Still Seeing This: Measuring the Effectiveness Of Ad Controls and Explanations in AI-Mediated Ad Targeting Systems

Jane Castleman, Aleksandra Korolova

Princeton University
janeec@princeton.edu, korolova@princeton.edu

Abstract

Recently, Meta has shifted towards AI-mediated ad targeting mechanisms that do not require advertisers to provide detailed targeting criteria. The shift is likely driven by excitement over AI capabilities as well as the need to address new data privacy policies and targeting changes agreed upon in civil rights settlements. At the same time, in response to growing public concern about the harms of targeted advertising, Meta has touted their ad preference controls as an effective mechanism for users to exert control over the advertising they see. Furthermore, Meta markets their "Why this ad" targeting explanation as a transparency tool that allows users to understand the reasons for seeing particular ads and inform their actions to control what ads they see in the future.

Our study evaluates the effectiveness of Meta's "See less" ad control, as well as the actionability of ad targeting explanations following the shift to AI-mediated targeting. We conduct a large-scale study, randomly assigning participants the intervention of marking "See less" to either *Body Weight Control* or *Parenting* topics, and collecting the ads Meta shows to participants and their targeting explanations before and after the intervention. We find that utilizing the "See less" ad control for the topics we study does not significantly reduce the number of ads shown by Meta on these topics, and that the control is less effective for some users whose demographics are correlated with the topic. Furthermore, we find that the majority of ad targeting explanations for local ads made no reference to location-specific targeting criteria, and did not inform users why ads related to the topics they requested to "See less" of continued to be delivered. We hypothesize that the poor effectiveness of controls and lack of actionability and comprehensiveness in explanations are the result of the shift to AI-mediated targeting, for which explainability and transparency tools have not yet been developed by Meta. Our work thus provides evidence for the need of new methods for transparency and user control, suitable and reflective of how the increasingly complex and AI-mediated ad delivery systems operate.

Introduction

In the first quarter of 2024, Meta brought in $35.6 billion in ad revenue (Meta 2024b), with their targeted advertising and matching systems helping advertisers reach new and existing audiences for whom the ads are particularly relevant (Meta 2022a). However, extensive prior work now details the harms that such targeted ads, pushed to users' feeds based on information collected about them without their explicit control, can have on users. For example, ads on subjects such as body weight control and image can increase anxiety and threaten users' health (Gak, Olojo, and Salehi 2022). Queer users have pointed out the dangers of targeted advertising in violating their privacy and control over their identity, and expressed the need for better explanations and controls to mediate the information and assumptions used to target their ads (Sampson, Encarnacion, and Metaxa 2023). Users with particular health conditions note the helplessness and trauma they feel when ads related to their health are forced onto their feed, with no way to completely stop these ads or prevent targeting based on inferred traits (Wu et al. 2023).

In response to these harms, Facebook (currently Meta) built a suite of user-facing tools such as ad preferences and individual ad targeting explanations, aimed to give users control and agency over the ads they see. A key *ad personalization control tool* is a button that allows users to mark "See less" to specific ad topics in Ad Preferences. Facebook promises that after marking "See less" for a topic, advertisers won't be able to target the user by specifying an interest in that topic for all future ads (shown in Figure 2) (Facebook Help Center 2024). A key *transparency control tool* is the "Why am I seeing this ad?" interface (shown in Figure 1) at the top of each ad (Meta 2019), which Facebook promises details the advertiser choices and user activity that informed that ad's delivery. Furthermore, the explanations are aimed to help users inform their ad control choices, as they provide a key entry point to the ad preferences.

Previous work has already found these controls to have limited accessibility (Hsu et al. 2020) and the ad explanations to not be fully transparent (Andreou et al. 2018). **Our main focus is understanding whether the effectiveness of these controls and explanations are further hampered by the recent shift to AI-mediated targeting on Meta**.

The shift to AI-mediated targeting has been a result of the introduction of new privacy protections and restrictions on targeting due to regulatory, competition, civil rights and political considerations. In particular, new data privacy policies, such as Apple's App Tracking Transparency, have pres-

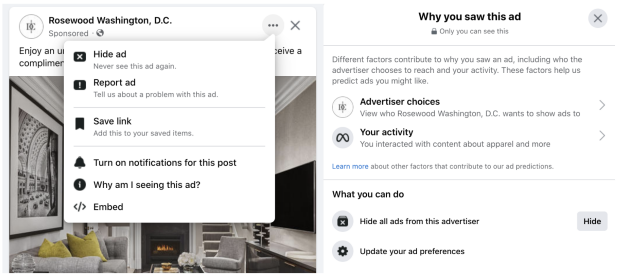

Figure 1: The options presented to users after clicking into the "Why am I seeing this ad?" interface.

sured Facebook to develop mechanisms to find audiences to deliver ads to while having less direct access to user data, especially from the advertisers. Furthermore, in response to a 2018 civil rights settlement on discrimination in employment advertising (ACLU 2018), a 2022 settlement on discrimination in housing advertising (The US Department of Justice 2022), and continued public pressure on the potential harms of political advertising (Shane 2017; Ribeiro et al. 2019; Weintraub 2019), Facebook has removed thousands of targeting categories from its advertiser-facing tools from 2022 to 2024 (Meta 2021, 2024c; Sapiezynski et al. 2024) and now aims to decrease skew in delivery of housing and employment ads (Austin 2022, 2023; Bogen et al. 2023; Timmaraju et al. 2023).

At the same time, to continue delivering on its message of the value of personalized advertising to users and the value of helping the advertisers reach the right audiences, Facebook has introduced new AI-mediated targeting and audience selection tools and new server-side tracking tools (El fraihi et al. 2024). One such AI-mediated audience selection tool, Advantage+ Audiences (Meta 2022a), promises to improve advertiser experiences and reduce costs (Meta 2020, 2022b, 2024c), using AI to identify relevant users without explicit targeting criteria such as user demographics and interests provided by the advertiser. Instead, Advantage+ relies on data about users' past conversions and browsing behaviors, and ad interactions of users that are algorithmically deemed similar, to narrow audiences on the advertisers' behalf (Meta 2024a).

Much of recent AI-based technology suffers from lack of interpretability and explainability, both due to its algorithmic complexity and due to the complexity of the data it relies upon (Lipton 2018; Poursabzi-Sangdeh et al. 2021). We thus hypothesize that the shift to AI-mediated targeting brings new challenges to the effectiveness of controls and explanations related to advertising personalization, as those tools likely have not kept pace with the shift to AI and are likely still built around categorical (i.e. demographic-, interest-, and activity-based) audience creation practices. We hypothesize that since Meta's suite of new Advantage+ tools no longer requires advertisers to supply precise targeting criteria, there is no clear pathway for users to see particular interests in their "Why this ad?" explanations, or for the ad delivery algorithm to act upon the "See less" user actions for specific topics. The result is lack of effectiveness of ad controls and explanations.

Our work and contributions We perform a large-scale sociotechnical audit to measure how Facebook's ad matching algorithm responds to user-initiated changes in ad controls on the topics of *Body Weight Control* and *Parenting* in early 2024. We recruit a generic group of users and a group of users whose age and gender are correlated with interest in each topic to understand whether it is more difficult for users belonging to certain demographic groups to effectively opt-out of certain ad topics. We randomly assign our recruited participants to mark "See less" to the ad topics and measure whether there is an observable change in the number of ads those users subsequently receive about those topics. Finally, to provide insight into the impact of Advantage+ audiences on ad explanations, we collect the targeting explanations attached to each ad and quantify the misalignment between the ad content and explanations.

Our findings provide new quantitative evidence of the ineffectiveness of marking "See less" to the topics of *Body Weight Control* and *Parenting*. They also show that ad targeting explanations do not productively inform ad control changes, including for ads relating to topics that users have marked to "See less" of. We summarize our findings as follows:

- We find large-scale, quantitative evidence of the ineffectiveness of Facebook's "See less" ad control that was built around categorical targeting of demographics, interests, and behaviors.
- We compare the effectiveness of "See less" ad controls across demographics, and find the controls are less effective for some users belonging to demographic groups that are more likely to have actual experiences with the topic they are trying to "See less" of.
- In our analysis of ad targeting explanations we find that the majority fail to be informative and do not connect to existing ad controls. In particular, the misalignment of generic geographic targeting explanations with evidently local ad content offers insight into the effects of Advantage+ audience targeting on actionability of ad explanations.

These results validate our hypothesis that AI-mediated targeting mechanisms negatively impact the effectiveness of ad controls and the actionability of ad delivery decisions, leading to a substantial decrease in meaningful transparency and control over advertising personalization for the users. Our work thus motivates new research questions in explainability of AI, particularly in reconciling AI-mediated targeting and matching with the need for faithful, actionable explanations and controls.

Related Work

Existing research outlines the importance of user control, transparency, and algorithmic audits in supporting users and holding platforms accountable for their targeted advertising systems.

User Control Providing users with control over the ads they see is crucial for maintaining their agency and safety. Ad delivery algorithms continue to show clickbait, untrustworthy, and distasteful ads despite including ad quality in their ad ranking process (Zeng, Kohno, and Roesner 2021). Ads can also cause psychological distress, loss of autonomy, changes to user behavior, and marginalization or traumatization (Wu et al. 2023).

Our study of Facebook's ad controls focuses on ad topic controls, inferred from users' activities and used in ad explanations. For users with histories of disordered eating, targeted advertising worsened anxiety surrounding food and exercise and negatively impacted self-esteem (Gak, Olojo, and Salehi 2022). For years, women have noted their inability to see fewer ads related to pregnancy and children's health, increasing anxiety and forcing them to revisit trauma (Brockell 2018; Contreras 2022). While users cannot fully remove assigned ad topics, they can mark "See less" to ads relating to these topics. Given users' expressed desires to exert control over ads relating to *Body Weight Control* and *Parenting*, we focus on these topics in our study of effectiveness of Facebook's "See less" ad control.

Despite Facebook's communications regarding the power and value of its ad personalization and transparency control tools (Meta 2019; Facebook Help Center 2024), research continues to find that current systems have ineffective algorithmic controls, opaque and misleading explanations, and poorly designed interfaces, all reducing user agency (Chromik et al. 2019). Other studies have emphasized the lack of usability of existing ad interfaces, (Habib et al. 2020, 2022; Leon et al. 2012), arguing that controls are not sufficiently accessible nor are they aligned with user needs.

Algorithmic Transparency Given the complexity of AI-mediated ad targeting systems, algorithmic transparency is crucial for users' understanding of how their data is used to make ad delivery decisions and inform ad control changes. When users do not fully understand AI systems, they build their own inferences about how the algorithm functions (Yuan et al. 2023) and expect improvement (Smith-Renner et al. 2020), which can lead to misinformed attempts at control and frustration. Users also note the importance of accurate ad explanations for ads leveraging personal data, interest targeting, and custom audiences to maintain a sense of control over their data (Lee et al. 2023; Wei et al. 2020), and their desire to understand how algorithms use their information to make inferences (Dolin et al. 2018; Eslami et al. 2018). Transparent explanations and opportunities for feedback through ad controls help inform productive choices over AI-mediated decisions and user data (Smith-Renner et al. 2020).

Recent research by (Chouaki et al. 2022) suggests that changes to Facebook's ad platform have driven a shift away from advertiser-driven microtargeting, finding that the majority of ads did not use microtargeting in comparison to a 2018 study (Cabañas, Cuevas, and Cuevas 2018). We hypothesize that shift towards generic, AI-mediated targeting threatens transparency, since ad delivery algorithms can no longer rely on advertiser targeting criteria to explain ad delivery, creating new transparency challenges.

Additionally, many problematic ads use no targeting, indicating that the ad delivery algorithm perceives them as relevant to users rather than delivering them based on audience targeting (Ali et al. 2023). When ad explanations are incomplete or incorrect, users may still attempt to infer explanations and are hesitant to blame algorithmic error, misinforming future actions to control their ads (Eslami et al. 2018; Rader and Gray 2015). Complete, accurate, and actionable ad explanations are thus necessary for expressing what data was used to inform ad delivery and provide users with information to make productive ad control changes.

Algorithmic Audits Algorithmic audits are essential tools for investigating the mechanisms of black-box ad delivery systems to uncover their impacts on privacy, fairness, polarization, and user experiences. Previous investigations into Facebook's ad targeting revealed that it led to privacy threats to users' personal information (Faizullabhoy and Korolova 2018; Korolova 2011); the Meta Pixel sent sensitive medical information to advertisers (Feathers et al. 2022), and that adversaries could uncover users' phone numbers and site visits by reconstructing custom audience data (Venkatadri et al. 2018). Audits also revealed discriminatory ad delivery, with algorithms inequitably delivering harmful ads (Ali et al. 2022), housing and employment ads (Ali et al. 2019; Nagaraj Rao and Korolova 2023; Imana, Korolova, and Heidemann 2021), and education ads (Imana, Korolova, and Heidemann 2024), and contributing to echo chambers in political advertising (Ali et al. 2021). Researchers have also found that custom audiences can introduce bias (Sapiezynski et al. 2022) and amplify existing bias (Speicher et al. 2018) in advertiser audiences. Large-scale user audits provide insight into user experiences with ad systems, such as issues with ad controls (Datta, Tschantz, and Datta 2015; Habib et al. 2022), perceptions of problematic ads (Zeng, Kohno, and Roesner 2021), and preferred ad explanations (Lee et al. 2023; Wei et al. 2020).

Using a large-scale user audit, our study addresses a gap in existing research by providing quantitative evidence of the ineffectiveness of Facebook's ad topic controls and the poor actionability of their targeting mechanism, prompted by the shift to AI-mediated targeting, marketed as Advantage+ audiences.

Facebook's Commitments to Users

Ad Controls

Facebook introduced the "See less" ad control in 2021 followed by an update in 2022, just four months before the release of Advantage+ audiences (Meta 2022a), promising that the ad topic control allows users to restrict the interest targeting categories used to reach them and the exert control over the ad content they see by opting to see fewer ads relating to certain topics (Meta 2021). Figure 2 shows their claim to users that "you won't get as many ads about that topic and advertisers can't target you based on an interest in it" (Facebook Help Center 2024).

When we tested the "See less" control in 2024, Facebook returned a popup message indicating that an ad control set-

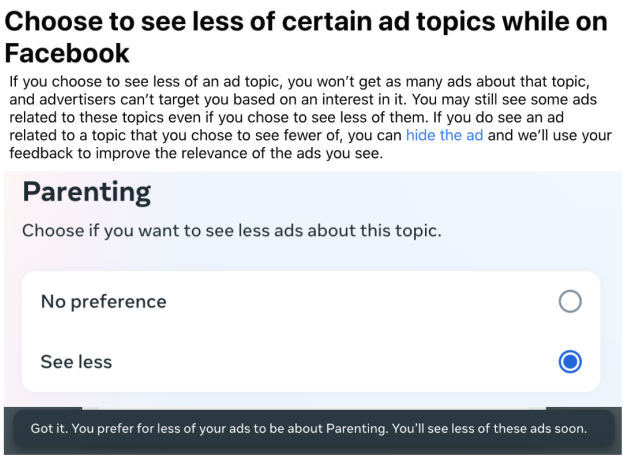

Figure 2: Facebook's description of the outcome of marking "See less" (Facebook Help Center 2024), and the message returned when the "See less" control is changed for the topic *Parenting*.

ting had been changed, shown in Figure 2. The commitment to show fewer related ads "soon" is vague, but we still expect users to see fewer related ads to a topic they have marked to "See less" of over a period of a couple weeks.

Ad Explanations

One of Facebook's five pillars of AI (Pesenti 2021) is "Transparency & Control." One mechanism to uphold this pillar is the "Why am I seeing this ad?" interface attached to each ad, initially debuted in 2014 (Meta 2019). The "Why am I seeing this ad?" interface includes an "Advertiser Choices" section listing an advertiser's targeting mechanisms, such as profile information, custom audiences, interests, and location (Meta 2019). A 2023 update adds a "Your activity" section listing previous user activity that influenced ad delivery, outputted by machine learning models similar to the ones that inform ad delivery (shown in Figure 6) to further increase transparency (Meta 2023). To be fully transparent, we expect these explanations to contain all of the "Advertiser Choices" that led to a user being targeted.

Methodology

In this section, we describe the structure of our user survey to perform a large-scale sociotechnical audit of Facebook's ad controls and ad explanations for two topics: *Parenting* and *Body Weight Control*. Our goal is to evaluate whether they uphold Facebook's commitments to users made in their descriptions of ad explanations and ad controls.

Participant Recruitment

We recruit participants for our study on Prolific (Prolific 2024) and use Prolific's built-in screeners based on participants' self-identified characteristics to choose two sets of Prolific users: one set with users who may be more likely to have ads related to *Parenting* and one set with users who may be more likely to have ads related to *Body Weight Control*. Given the inherently limited scope of our study, we chose these topics due to previously expressed concerns about ads worsening trauma related to parenting (Brockell 2018; Contreras 2022) or exacerbating food- and exercise-related anxiety (Gak, Olojo, and Salehi 2022). Participants in each set match the following screening characteristics:

- *Parenting*: Has a child less than 10 years old, currently living with their child, uses Facebook.
- *Body Weight Control*: Has gone on a diet in the past, marked "Health & Fitness" as a hobby, exercises "Sometimes" or "Often," uses Facebook.

Furthermore, to test our hypothesis that it is more difficult for users belonging to demographic groups that Facebook deems correlated with the topics to effectively change the ads they see via the provided "See less" controls we develop the following *correlated demographics* for each interest, used in combination with our previous screeners:

- *Parenting*: Women aged 25 to 45, since they are most likely to be new parents (Bui and Miller 2018).
- *Body Weight Control*: Women aged 18 to 59 and men aged 40 to 59, since these groups are most likely to have experiences with dieting and weight loss (Martin et al. 2019).

In sum, we recruit 201 participants, 110 from the *Parenting* screener, of whom 68 match the *Parenting* correlated demographic, and 91 from the *Body Weight Control (BWC)* screener, of whom 31 match the correlated demographic. Table 1 shows the demographic breakdown of our participants. Our participants' demographics are skewed towards women and individuals under 50, in line with the skewed demographics of Prolific users (Charalambides 2021).

Study Design

We asked all participants to collect the ads they see on Facebook and share them in three separate rounds of the study, separated by approximately one week each, in a process outlined in Figure 3.

Variable	**Value**	***Parenting***		***BWC***	
		n	**%**	**n**	**%**
Gender	Woman	92	83.6	57	63.3
	Man	18	16.4	34	36.7
Age	18-29	23	22.1	17	18.7
	30-49	70	67.3	46	50.6
	50-69	12	11.5	25	27.5
	> 69	0	0	3	3.3
Correlated	1 (yes)	68	61.8	31	34.1
	0 (no)	42	38.2	60	65.9
Total		110		91	

Table 1: Demographics of participants, stratified by topic.

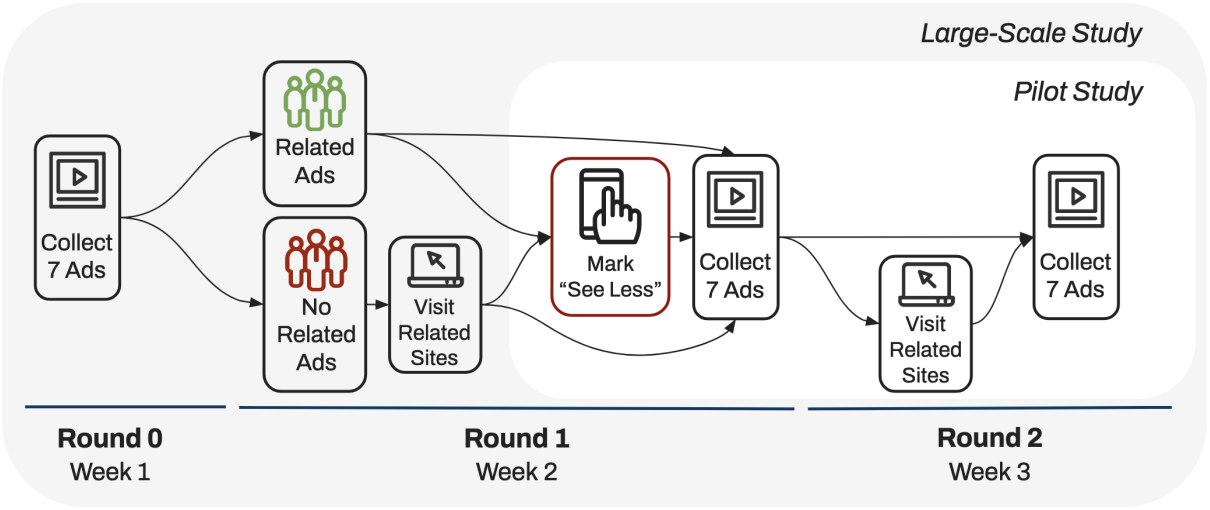

Figure 3: Timeline of our user survey to collect feed ads with randomized ad control intervention.

After analyzing the set of shared ads from Round 0, we asked the 62 participants for the topic *Parenting* and 79 participants for the topic *Body Weight Control* who did not see any ads related to their assigned topic to visit the websites of advertisers related to their topic. During Round 1, we asked a randomly chosen half of participants from each topic group, whom we subsequently call our intervention group, to mark "See less" to their assigned topic, then share the ads they saw immediately after that. For the topic *Parenting*, 46 participants exercised the "See less" ad control and for the topic *Body Weight Control*, 28 participants exercised the "See less" ad control. The other participants, whom we subsequently call our Control group, were not asked to perform any action and were only asked to share the ads they saw, as in Week 1. During Round 2, we again asked the groups of participants who did not see any ads related to their assigned topic to visit the websites of advertisers related to their topic.

We selected the websites for participants without related ads to visit from stereotypical advertisers in *Parenting* and *Body Weight Control*, prioritizing selection using the number of advertisements the advertiser was actively running as listed by the Facebook Ad Library. Then, we verified their websites use Meta Pixels to track participant data using The Markup's BlackLight tool (Mattu and Sankin 2020). We did this to simulate users with continued web activity to sites related to topics they marked to "See less" of, which should not reduce the effectiveness of the "See less" ad control.

For each ad, we collected the advertiser, the "Why am I seeing this ad?" information attached to each ad, and a link to the ad. The large-scale study consisted of one pre-intervention round (Round 0), collecting 8 feed ads per participant, followed by 2 post-intervention rounds (Round 1 and Round 2), collecting 7 feed ads per participant. The pilot study consisted of Round 1 and Round 2 from the large-scale study, including the randomized intervention and the sorting of participants into groups with related and unrelated ads.

Parenting	Body Weight Control
Goodnites	Planet Fitness
Primrose Schools	Noom
Care.com	WeightWatchers
Pampers	Nutrisystem
Graco Baby	OrangeTheory
Kindercare	WHOOP
BuyBuyBaby	MyFitnessPal
The Nok Box	

Table 2: Advertisers related to participants' topic of interest.

Measuring Ad Control Effectiveness

We first study the effectiveness of the "See less" control by measuring how the average number of ads related to topic marked "See less" changes over time for the intervention versus control groups. We split our data into two datasets, one for participants assigned the topic *Parenting* and one for *Body Weight Control*. Then, we create a codebook to assign each ad a score based on its relatedness to a participant's assigned topic, calculating the average relatedness for the intervention and control groups in each dataset. We find that the intervention and control groups experience similar rates of related ads over time, suggesting that the "See less" intervention may not be effective.

Classifying Related Ads

We manually code the relatedness of each ad to the participant's assigned topic, either *Parenting* or *Body Weight Control*, along the following scale: "Not related" = 0, "Somewhat related" = 0.5" and "Related" = 1. We assign each ad j with a numeric $ad\_score[j]$ using this relatedness scale.[1]

[1]The statistical significance of our results is robust to changes in the magnitude of our ad relatedness scale; values from 0 to 1 were chosen for simplicity.

Figure 4: From left to right, ads coded with relatedness = 0, 0.5, 1 to *Parenting*.

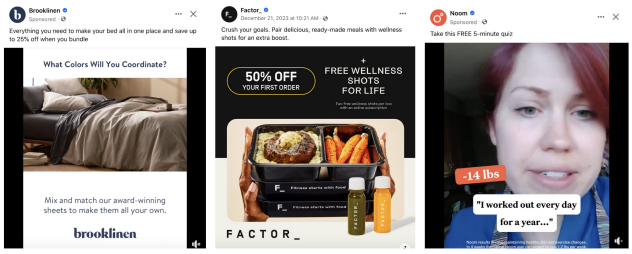

Figure 5: From left to right, ads coded with relatedness = 0, 0.5, 1 to *Body Weight Control*.

- Score of 0: The text or image does not contain information related to the blocked topic.
- Score of 0.5: The text or image contains information somewhat related to the blocked topic. *Body Weight Control*: diet-focused meal planning services, exercise equipment, and ad images referencing fitness and nutrition. *Parenting*: children's activities, toys, and images of children with references to parenting.
- Score of 1: The text or image contains information directly related to the blocked topic. *Body Weight Control*: fitness products and services, diet plans, and ad images promoting weight loss and working out. *Parenting*: childcare services, childcare products, and children's healthcare.

Example ads for *Parenting* and *Body Weight Control* are shown in Figures 4 and 5, respectively.

Measuring Change in Related Ads

For each topic, we combine participant data from the pilot and large-scale studies by round. We calculate $r_u^{(i)}$, the proportion of ads related to their assigned topic shown to a user u in Round i, by summing the ad scores for each ad j they encountered in Round i, and dividing that sum by the total number of ads for that user in that round, n_{u_i}:

$$r_u^{(i)} = \frac{1}{n_{u_i}} \sum_{j=1}^{n_{u_i}} ad\_score[j]$$

For a subset of participants S belonging to a particular demographic or intervention group, we calculate the average proportion of related ads in each round.

	Parenting		Body Weight Control	
Round	Control	Intv.	Control	Intv.
0	0.130	0.163	0.029	0.038
1	0.123	0.108	0.120	0.060
2	0.191	0.139	0.161	0.114

Table 3: Average proportion of ads $\mu_S^{r^{(i)}}$ related to the topics *Parenting* and *Body Weight Control* seen by users with and without marking "See less."

$$\mu_S^{r^{(i)}} = \frac{1}{|S|} \sum_{u \in S} r_u^{(i)}$$

We present the results broken down by round, topic and group in Table 3. For the topic *Parenting*, the control group sees an increase in the proportion of related ads from Round 0 to Round 2, while the intervention group sees a decrease overall, but a slight increase from Round 1 to Round 2. However, for the *Body Weight Control* group, both the control and intervention groups see an increase in the proportion of related ads, suggesting the "See less" ad control is ineffective.

Testing for Significance in the Change in Relatedness

To establish whether the observed change in the proportion of related ads over time is statistically significant, we use the Mann-Whitney U test with the significance level $\alpha = 0.1$ (Mann and Whitney 1947). Specifically, we compare the median change in related ads over time between the control and intervention groups. Our analysis did not show a statistically significant difference in the decrease of related ads between the control and intervention groups for *Parenting* or *Body Weight Control*, suggesting that the "See less" ad control does not meaningfully reduce the number of ads related to the blocked topic.

First, we calculate the change in relatedness from Round 0 to Round 2 for user u as δ_u:

$$\delta_u = r_u^{(2)} - r_u^{(0)}.$$

Then, we find the median change in the proportion of related ads experienced by a subset of participants in group S, m_δ^S:

$$m_\delta^S = median(\{\delta_u\}, \text{for } u \in S).$$

We compare m_δ for the control and intervention groups, denoted $m_\delta^{P,C}, m_\delta^{P,I}$ for the topic *Parenting* and $m_\delta^{B,C}, m_\delta^{B,I}$ for the topic *Body Weight Control*. Our null hypotheses are $m_\delta^{P,C} = m_\delta^{P,I}$ and $m_\delta^{B,C} = m_\delta^{B,I}$. Table 4 lists the results of these tests.

As the p-values for both hypotheses are greater than the significance level of $\alpha = 0.1$, we conclude that there is no evidence that the intervention group saw a significantly greater decrease in the proportion of related ads from Round 0 to Round 2 in comparison to the control group for both *Parenting* and *Body Weight Control*.

Null Hypothesis	p-value	Adjusted p-value
$m_\delta^{B,C} = m_\delta^{B,I}$	0.129	0.258
$m_\delta^{P,C} = m_\delta^{P,I}$	0.229	0.305

$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Table 4: p-values for Mann-Whitney U Tests of Intervention vs. Control.

Comparison of Demographic Type

We compare the effectiveness of the "See Less" intervention for our topics of interest for all users in the intervention group, split into correlated versus non-correlated demographics. We hypothesize that the effectiveness of the intervention decreases for participants belonging to demographics correlated with our topic of interest. Table 5 shows the average proportion in related ads seen by round and demographic type, split by topic.

The impact of participant demographic on the average proportion of related ads shown after the intervention is mixed for the topics *Body Weight Control* and *Parenting*. We find that for the topic *Body Weight Control*, participants in non-correlated demographics see a stronger increase in the average proportion of related ads shown than participants in correlated demographics, while for the topic *Parenting*, participants in correlated demographics see an increase in the average proportion of related ads shown while those in non-correlated demographics see a decrease.

	Parenting		Body Weight Control	
Round	Corr.	Not Corr.	Corr.	Not Corr.
0	0.155	0.187	0.051	0.010
1	0.132	0.048	0.072	0.028
2	0.183	0.041	0.093	0.214

Table 5: Average proportion of ads related to the topics *Parenting* and *Body Weight Control* seen by users in correlated demographics and those not in correlated demographic.

Testing for Significance across Demographic Types We also compare the median change in related ads over time between participants who marked "See less" belonging to correlated demographics versus those not belonging to correlated demographics using a Mann-Whitney U test. We compare m_δ for correlated and non-correlated demographics, denoted $m_\delta^{P,COR}, m_\delta^{P,NCOR}$ for the topic *Parenting* and $m_\delta^{B,COR}, m_\delta^{B,NCOR}$ for the topic *Body Weight Control*. Our null hypotheses are $m_\delta^{P,COR} = m_\delta^{P,NCOR}$ and $m_\delta^{B,COR} = m_\delta^{B,NCOR}$. Table 6 lists the results of this test.

For the topic *Parenting*, participants in correlated demographics see significantly less of a decrease in related ads than participants in non-correlated demographics. This provides statistically significant evidence at $\alpha = 0.1$ level the that it is more difficult for users in our defined correlated demographics to see fewer ads related to *Parenting* than users not in correlated demographics. Conversely, we find no significant difference in the median change in the proportion of related ads between participants who marked "See less" in correlated versus non-correlated demographics for the topic *Body Weight Control*.

Null Hypothesis	p-value	Adjusted p-value
$m_\delta^{B,COR} = m_\delta^{B,NCOR}$	0.995	0.995
$m_\delta^{P,COR} = m_\delta^{P,NCOR}$	0.021^{**}	0.085^{*}

$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Table 6: p-values for Mann-Whitney U Tests between Demographic Types.

Measuring Actionability of Targeting Explanations

Facebook does not publish information about the number or types of advertisers using the suite of Advantage+ products, nor is this information included in the "Why am I seeing this?" interface. Therefore, to gain circumstantial evidence of the prevalence of Advantage+ audience use, we first measure the frequency of location-specific criteria in the ad targeting information of local ads. We then measure whether topic-specific ads indicate that they were targeted using that particular topic. We hypothesize that ads delivered using Advantage+ audience selection only return generic information about their location and interest targeting in the explanations, and thus the above measurements would be good proxies. We find that the majority of location-specific ads do not reference location information in targeting explanations, and the majority of ads related to specific interests only return generic targeting information. The findings support our hypothesis that the shift to Advantage+ audience selection reduces ad explanation faithfulness and actionability.

Defining Local Ads and Targeting Explanation Types

We collected 3,868 ads from 201 participants over three rounds. We now detail how we label ads as location-specific and classify targeting information into three types: generic, activity-based, and interest-based.

Labeling Local Ads To classify ads as *local*, we establish specific criteria based on the ad creative's content. An ad is labeled as local if it references products and services that are geographically constrained, meaning they are only accessible to users within a particular area. We use the following criteria to determine whether an ad is local:

- Geographic references: ad explicitly references a specific state, city, town, or neighborhood, excluding vacation ads.
- Location-based services: ad references services that are inherently local, such as real estate listings, local events, or health services.
- Contact Information: ad lists local phone number, address, or directions indicating a specific location.

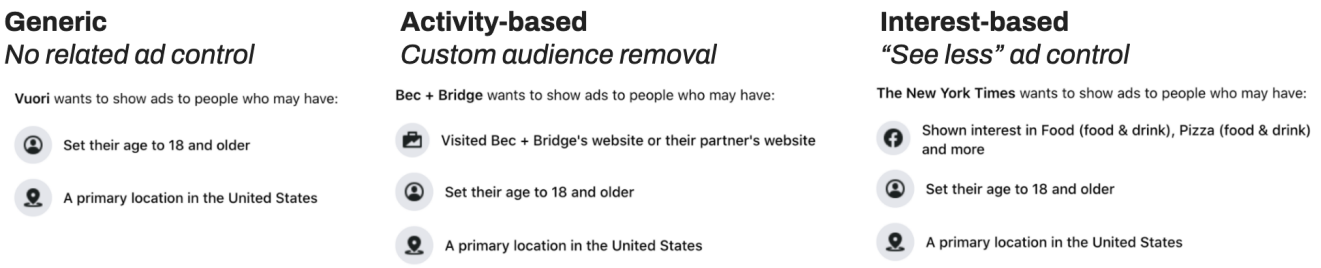

Figure 6: From left to right, "Why am I seeing this ad?" explanations classified as generic, interest-based, and activity-based.

Targeting Info.	Generic	Activity	Interest
Age	✓	-	-
Location (broad)	✓	-	-
Language	✓	-	-
Gender	✓	-	-
Interest	✗	-	✓
Site Visit	✗	✓	-
Hashed List	✗	✓	-
Lookalike Audience	✗	✓	-

Table 7: Coding Ads by Targeting Information

Labeling Targeting Explanations Using the advertising targeting information from the uploaded "Why am I seeing this ad?" explanation, we label each targeting explanation as *generic*, *interest-based*, or *activity-based*, with examples shown in Figure 6.

We label ad explanations that only included a location as the United States, age, language, and gender as *generic*, since users cannot change these attributes. Generic explanations do not correspond to an existing ad control, and do not give information as to why the ad was delivered to a specific user since they are so broad. Ads using Advantage+ audience selection return generic ad explanations since they do not rely on manual targeting criteria.

We label explanations listing an interest as *interest-based*, which connect inferences of users interests to ad delivery and inform changes to the "See less" ad control. We label ad explanations listing web activity, including similar activity to existing customers, as *activity-based*, which connect online activity to ad delivery and inform changes to custom audience preferences. Ad explanations can be both interest-based and activity-based. Our full criteria is listed in Table 7. A "✓" indicates included information and a "✗" indicates excluded information. A label is added to an explanation when all included and excluded information is satisfied.

Targeting Explanations by Ad Type

Local Ads We now present our findings for local ads and their explanations (see Figure 7). To uphold actionability and faithfulness to ad content, we expect local ads to specify either location-specific targeting or activity-related information in the explanations. We find that this is not the case – 67% of local ads only included generic targeting information, while 17% listed interest-related targeting information and 16% listed activity-related targeting information. Thus, for the vast majority of local ads, the explanations are not faithful or actionable, as we would expect local ads to either list local targeting or activity-based targeting as reasoning why it was delivered to specific users. We cannot reliably tell whether the underlying reason for generic targeting is Advantage+, but AI-mediated audience selection is one likely explanation for such a prevalence of missing local or activity information. The lack of faithfulness for local ads makes it challenging for users to understand what controls to leverage.

Interest-Specific Ads We now present our analysis of the targeting information attached to the 222 related ads delivered in Round 2, one week after the "See less" intervention. We focus on Round 2 since these ads were delivered after we would expect the intervention to be fully effective. Therefore, we expect faithful explanations to inform participants why an ad was delivered despite the control, rather than returning generic targeting explanations that do not reference the user's expressed preferences to "See less."

We calculate the proportion of generic, interest-based, and activity-based targeting information for the control and intervention groups, shown in Figure 8. We also include the proportion of ads from advertisers that participants with no related ads were instructed to visit ("Pixeled" advertisers) from Table 2.

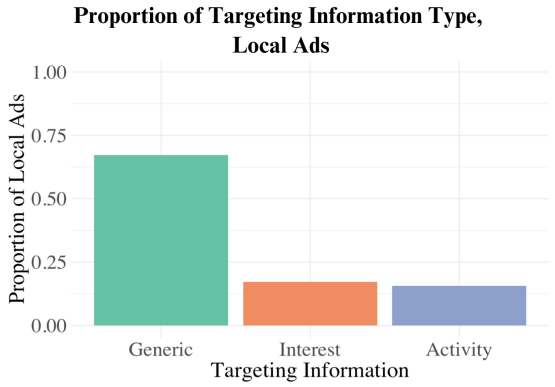

Figure 7: Proportion of targeting information type for local ads.

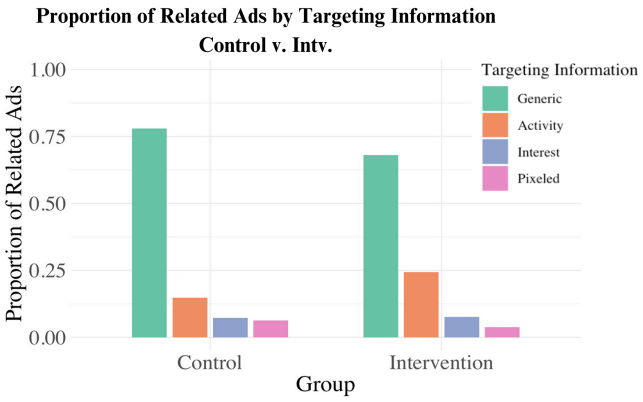

Figure 8: Proportions of targeting information type for related ads, post-intervention.

Similar to our findings with respect to explanations for local ads, we find that the majority of targeting explanations for interest-specific ads are generic; moreover, the intervention group sees an increase in activity-based targeting explanations. Both of these findings are consistent with what one would expect from AI-mediated targeting and delivery, since AI-mediated targeting is not aligned with the interest-based "See less" ad control, so users are targeted despite the control, returning a majority of generic targeting explanations.

In summary, the prevalence of generic targeting information reduces the actionability of explanations, since they do not help users understand why ads were delivered to them or express their ad preferences to change the ads they see.

Discussion

We demonstrated the ineffectiveness of Facebook's "See less" ad topic control and the lack of actionability in ad explanations, evidenced by the prevalence of generic ad targeting explanations. Along with Chouaki et al., we attribute these issues in user-facing controls and transparency tools to the recent shift from advertiser microtargeting to AI-mediated audience selection (Meta 2022a).

Broadly, our findings illustrate the significant challenges in preserving transparency, actionability, and user control over complex AI systems. Previous research has highlighted the difficulty of balancing data deletion with explainable decisions in AI-mediated recommendation systems (Pawelczyk et al. 2023). As new data privacy regulations and governance protections emerge, it is crucial to investigate how AI systems can support users' rights to transparency and control while employing AI-mediated targeting mechanisms.

More specifically, our findings reflect a fundamental misalignment between Facebook's existing ad controls and explanations with their new ad delivery system. Advantage+ audience selection shows how AI can improve ad targeting while reducing the collection of user data, potentially improving privacy, but this improvement currently comes at the cost of effectiveness of user-facing control and faithfulness in the transparency of ad targeting decisions. Therefore, future research, especially from platforms, is necessary to develop new mechanisms for transparency and user control that align with the shift to AI-mediated ad targeting and delivery systems.

Limitations & Future Work

First, our study collected a relatively low number of ads, collecting a maximum of 22 ads per participant, instead aiming to attract and retain a high number of participants. Future research could conduct a longitudinal study similar to the work done by the Panoptykon Foundation (Sapieżyński 2023), but with a larger number of participants. Additionally, due to our limited budget for recruiting participants and the difficulty of finding participants with high proportions of related ads, we asked participants to artificially change their web activity by visiting sites related to their blocked topics. The construction of surveys and collection of data could be streamlined with a browser extension hosting surveys and automatically collecting participants' ad data as in Lam et al..

Ethics

Because of the potentially private nature of the ads users see, to minimize potential harm, we collected the minimum data necessary for our research goals and made the data sharing with us optional when possible. Our study and data collection and storage practices were approved by our Institutional Review Board. We ensured participants were fully briefed on the scope of the study prior to consenting and paid them between $12/hr and $20/hr, increasing payments for more time- and data-intensive surveys. We believe that participation in our study, such as clicking on "Why this ad?" information did not affect users' experiences on the Facebook platform and exercising the "See less" controls both had limited effect and was transparent to users and they could have chosen not to follow this instruction, without penalty. We did not use any automated scraping to collect the data.

Conclusion

In our large-scale study, we found that participants marking "See less" to *Parenting* and *Body Weight Control* did not see a significant decrease in the proportion of ads related to these topics over time. Furthermore, some users in correlated demographics with their blocked topics saw a higher proportion of related ads, on average, than users not in correlated demographics. Finally, the majority of local ads and topic-related ads delivered to users post-intervention included only generic targeting information, failing to guide users as to what controls they could further exercise to decrease the proportion of ads on these topics. Overall, our study demonstrates that the push for advertisers to use Advantage+ audiences and the resulting AI-mediated user targeting, without the corresponding updates to the technology behind explanations and controls, leads to the prevalence of generic targeting information and the ineffectiveness of the "See less" ad controls. Our work motivates future research into transparency and control interfaces that are effective when the primary targeting process is AI- rather than advertiser- driven.

Acknowledgements

We thank Andrés Monroy-Hernández of the Princeton HCI Lab for his thoughtful and insightful feedback. This work was supported in part by funding from Princeton's McIntosh Independent Work/Senior Thesis Fund, which was used towards compensation of our survey participants and by the National Science Foundation grants CNS-1956435 and CNS-2344925.

References

ACLU. 2018. ACLU and Workers Take On Facebook for Gender Discrimination in Job Ads. https://www.aclu.org/press-releases/aclu-and-workers-take-facebook-gender-discrimination-job-ads.

Ali, M.; Goetzen, A.; Mislove, A.; Redmiles, E.; and Sapiezynski, P. 2022. All Things Unequal: Measuring Disparity of Potentially Harmful Ads on Facebook. In *Proceedings of the 2022 Workshop on Consumer Protection*.

Ali, M.; Goetzen, A.; Mislove, A.; Redmiles, E. M.; and Sapiezynski, P. 2023. Problematic Advertising and its Disparate Exposure on Facebook. In *Proceedings of the 32nd USENIX Security Symposium*, arXiv:2306.06052. arXiv. ArXiv:2306.06052 [cs].

Ali, M.; Sapiezynski, P.; Bogen, M.; Korolova, A.; Mislove, A.; and Rieke, A. 2019. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): 1–30.

Ali, M.; Sapiezynski, P.; Korolova, A.; Mislove, A.; and Rieke, A. 2021. Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging. In *14th ACM International Conference on Web Search and Data Mining (WSDM)*.

Andreou, A.; Venkatadri, G.; Goga, O.; Gummadi, K. P.; Loiseau, P.; and Mislove, A. 2018. Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations. In *Proceedings 2018 Network and Distributed System Security Symposium*. San Diego, CA: Internet Society. ISBN 978-1-891562-49-5.

Austin, R. L. 2022. Expanding Our Work on Ads Fairness. https://about.fb.com/news/2022/06/expanding-our-work-on-ads-fairness/.

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

Bogen, M.; Tripathi, P.; Timmaraju, A. S.; Mashayekhi, M.; Zeng, Q.; Roudani, R.; Gahagan, S.; Howard, A.; and Leone, I. 2023. Toward fairness in personalized ads.

Brockell, G. 2018. Perspective — Dear tech companies, I don't want to see pregnancy ads after my child was stillborn. *Washington Post*.

Bui, Q.; and Miller, C. C. 2018. The Age That Women Have Babies: How a Gap Divides America. *The New York Times*.

Cabañas, J. G.; Cuevas, A.; and Cuevas, R. 2018. Unveiling and quantifying facebook exploitation of sensitive personal data for advertising purposes. In *Proceedings of the 27th USENIX Conference on Security Symposium*, SEC'18, 479–495. USA: USENIX Association. ISBN 9781931971461.

Charalambides, N. 2021. We recently went viral on TikTok - here's what we learned. https://www.prolific.com/resources/we-recently-went-viral-on-tiktok-heres-what-we-learned.

Chouaki, S.; Bouzenia, I.; Goga, O.; and Roussillon, B. 2022. Exploring the Online Micro-targeting Practices of Small, Medium, and Large Businesses. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2): 1–23.

Chromik, M.; Eiband, M.; Völkel, S. T.; and Buschek, D. 2019. Dark Patterns of Explainability, Transparency, and User Control for Intelligent Systems. In *Joint Proceedings of the ACM IUI 2019 Workshops*.

Contreras, B. 2022. https://www.latimes.com/business/technology/story/2022-05-25/for-pregnant-women-the-internet-can-be-a-nightmare.

Datta, A.; Tschantz, M. C.; and Datta, A. 2015. Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. In *Proceedings on Privacy Enhancing Technologies*, arXiv:1408.6491. arXiv. ArXiv:1408.6491 [cs].

Dolin, C.; Weinshel, B.; Shan, S.; Hahn, C. M.; Choi, E.; Mazurek, M. L.; and Ur, B. 2018. Unpacking Perceptions of Data-Driven Inferences Underlying Online Targeting and Personalization. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. Montreal QC Canada: ACM. ISBN 978-1-4503-5620-6.

El fraihi, A.; Amieur, N.; Rudametkin, W.; and Goga, O. 2024. Client-side and Server-side Tracking on Meta: Effectiveness and Accuracy. In *Proceedings on Privacy Enhancing Technologies*.

Eslami, M.; Krishna Kumaran, S. R.; Sandvig, C.; and Karahalios, K. 2018. Communicating Algorithmic Process in Online Behavioral Advertising. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 1–13. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-5620-6.

Facebook Help Center. 2024. Choose to see less of certain ad topics while on Facebook. https://www.facebook.com/help/353660662271696.

Faizullabhoy, I.; and Korolova, A. 2018. Facebook's Advertising Platform: New Attack Vectors and the Need for Interventions. In *IEEE Workshop on Technology and Consumer Protection (ConPro '18)*.

Feathers, T.; Fondrie-Teitler, S.; Waller, A.; and Mattu, S. 2022. Facebook Is Receiving Sensitive Medical Information from Hospital Websites – The Markup. https://themarkup.org/pixel-hunt/2022/06/16/facebook-is-receiving-sensitive-medical-information-from-hospital-websites.

Gak, L.; Olojo, S.; and Salehi, N. 2022. The Distressing Ads That Persist: Uncovering The Harms of Targeted Weight-Loss Ads Among Users with Histories of Disordered Eating. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2): 1–23.

Habib, H.; Pearman, S.; Wang, J.; Zou, Y.; Acquisti, A.; Cranor, L. F.; Sadeh, N.; and Schaub, F. 2020. "It's a scavenger hunt": Usability of Websites' Opt-Out and Data Deletion Choices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–12. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-6708-0.

Habib, H.; Pearman, S.; Young, E.; Saxena, I.; Zhang, R.; and Cranor, L. F. 2022. Identifying User Needs for Advertising Controls on Facebook. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1): 1–42.

Hsu, S.; Vaccaro, K.; Yue, Y.; Rickman, A.; and Karahalios, K. 2020. Awareness, Navigation, and Use of Feed Control Settings Online. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–13. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-6708-0.

Imana, B.; Korolova, A.; and Heidemann, J. 2021. Auditing for Discrimination in Algorithms Delivering Job Ads. In *The Web Conference (WWW)*.

Imana, B.; Korolova, A.; and Heidemann, J. 2024. Auditing for Racial Discrimination in the Delivery of Education Ads. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 2348–2361. ArXiv:2406.00591 [cs].

Korolova, A. 2011. Privacy Violations Using Microtargeted Ads: A Case Study. *Journal of Privacy and Confidentiality*, 3(1): 27–49.

Lam, M. S.; Pandit, A.; Kalicki, C. H.; Gupta, R.; Sahoo, P.; and Metaxa, D. 2023. Sociotechnical Audits: Broadening the Algorithm Auditing Lens to Investigate Targeted Advertising. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2): 360:1–360:37.

Lee, H.-P. H.; Logas, J.; Yang, S.; Li, Z.; Barbosa, N.; Wang, Y.; and Das, S. 2023. When and Why Do People Want Ad Targeting Explanations? Evidence from a Four-Week, Mixed-Methods Field Study. In *2023 IEEE Symposium on Security and Privacy (SP)*, 2903–2920. San Francisco, CA, USA: IEEE. ISBN 978-1-66549-336-9.

Leon, P.; Ur, B.; Shay, R.; Wang, Y.; Balebako, R.; and Cranor, L. 2012. Why Johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 589–598. Austin Texas USA: ACM. ISBN 978-1-4503-1015-4.

Lipton, Z. C. 2018. The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3): 31–57.

Mann, H. B.; and Whitney, D. R. 1947. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1): 50–60.

Martin, C.; Herrick, K.; Sarafrazi, N.; and Ogden, C. 2019. Attempts to Lose Weight Among Adults in the United States, 2013–2016. https://www.cdc.gov/nchs/products/databriefs/db313.htm.

Mattu, S.; and Sankin, A. 2020. How We Built a Real-time Privacy Inspector – The Markup. https://themarkup.org/blacklight/2020/09/22/how-we-built-a-real-time-privacy-inspector.

Meta. 2019. Why Am I Seeing This? We Have an Answer for You. https://about.fb.com/news/2019/03/why-am-i-seeing-this/.

Meta. 2020. Simplifying Targeting Categories. https://www.facebook.com/business/news/update-to-facebook-ads-targeting-categories.

Meta. 2021. Removing Certain Ad Targeting Options and Expanding Our Ad Controls. https://www.facebook.com/business/news/removing-certain-ad-targeting-options-and-expanding-our-ad-controls.

Meta. 2022a. Introducing New Automation Tools to Increase Sales and Drive Growth. https://about.fb.com/news/2022/08/introducing-new-automation-tools-to-increase-sales-and-drive-growth/. Accessed: 2024-05-04.

Meta. 2022b. Preparing for Upcoming Removal of Certain Ad Targeting Options. https://www.facebook.com/government-nonprofits/blog/preparing-for-upcoming-removal-of-certain-ad-targeting-options.

Meta. 2023. Increasing Our Ads Transparency. https://about.fb.com/news/2023/02/increasing-our-ads-transparency/.

Meta. 2024a. About Advantage+ audience. https://www.facebook.com/business/help/273363992030035.

Meta. 2024b. Meta Reports First Quarter 2024 Results. https://investor.fb.com/investor-news/press-release-details/2024/Meta-Reports-First-Quarter-2024-Results/default.aspx.

Meta. 2024c. Updates to detailed targeting. https://www.facebook.com/business/help/458835214668072.

Nagaraj Rao, V.; and Korolova, A. 2023. Discrimination through Image Selection by Job Advertisers on Facebook. In *2023 ACM Conference on Fairness, Accountability, and Transparency*, 1772–1788. Chicago IL USA: ACM. ISBN 9798400701924.

Pawelczyk, M.; Leemann, T.; Biega, A.; and Kasneci, G. 2023. On the Trade-Off between Actionable Explanations and the Right to be Forgotten. In *International Conference on Learning Representations*, arXiv:2208.14137. arXiv. ArXiv:2208.14137 [cs].

Pesenti, J. 2021. Facebook's five pillars of Responsible AI. https://ai.meta.com/blog/facebooks-five-pillars-of-responsible-ai/.

Poursabzi-Sangdeh, F.; Goldstein, D. G.; Hofman, J. M.; Wortman Vaughan, J. W.; and Wallach, H. 2021. Manipulating and Measuring Model Interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 1–52. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-8096-6.

Prolific. 2024. Easily find vetted research participants and AI taskers at scale. https://www.prolific.com/.

Rader, E.; and Gray, R. 2015. Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, 173–182.

New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-3145-6.

Ribeiro, F. N.; Saha, K.; Babaei, M.; Henrique, L.; Messias, J.; Benevenuto, F.; Goga, O.; Gummadi, K. P.; and Redmiles, E. M. 2019. On microtargeting socially divisive ads: A case study of russia-linked ad campaigns on facebook. In *Proceedings of the conference on fairness, accountability, and transparency*, 140–149.

Sampson, P.; Encarnacion, R.; and Metaxa, D. 2023. Representation, Self-Determination, and Refusal: Queer People's Experiences with Targeted Advertising. In *2023 ACM Conference on Fairness, Accountability, and Transparency*, 1711–1722. Chicago IL USA: ACM. ISBN 9798400701924.

Sapiezynski, P.; Ghosh, A.; Kaplan, L.; Rieke, A.; and Mislove, A. 2022. Algorithms that "Don't See Color": Comparing Biases in Lookalike and Special Ad Audiences. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 609–616. ArXiv:1912.07579 [cs].

Sapiezynski, P.; Kaplan, L.; Korolova, A.; and Mislove, A. 2024. On the Use of Proxies in Political Ad Targeting. In *Proceedings of the ACM on Human-Computer Interaction*.

Sapieżyński, P. 2023. Algorithms of Trauma 2. How Facebook Feeds on Your Fears — Fundacja Panoptykon. https://panoptykon.org/algorithms-of-trauma-2-how-facebook-feeds-on-your-fears.

Shane, S. 2017. These Are the Ads Russia Bought on Facebook in 2016. https://www.nytimes.com/2017/11/01/us/politics/russia-2016-election-facebook.html.

Smith-Renner, A.; Fan, R.; Birchfield, M.; Wu, T.; Boyd-Graber, J.; Weld, D. S.; and Findlater, L. 2020. No Explainability without Accountability: An Empirical Study of Explanations and Feedback in Interactive ML. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. Honolulu HI USA: ACM. ISBN 978-1-4503-6708-0.

Speicher, T.; Ali, M.; Venkatadri, G.; Ribeiro, F. N.; Arvanitakis, G.; Benevenuto, F.; Gummadi, K. P.; Loiseau, P.; and Mislove, A. 2018. Potential for Discrimination in Online Targeted Advertising. *Proceedings of Machine Learning Research*.

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. https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known.

Timmaraju, A. S.; Mashayekhi, M.; Chen, M.; Zeng, Q.; Fettes, Q.; Cheung, W.; Xiao, Y.; Kannadasan, M. R.; Tripathi, P.; Gahagan, S.; Bogen, M.; and Roudani, R. 2023. Towards Fairness in Personalized Ads Using Impression Variance Aware Reinforcement Learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4937–4947. ArXiv:2306.03293 [cs].

Venkatadri, G.; Andreou, A.; Liu, Y.; Mislove, A.; Gummadi, K. P.; Loiseau, P.; and Goga, O. 2018. Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface. In *2018 IEEE Symposium on Security and Privacy (SP)*, 89–107. San Francisco, CA: IEEE. ISBN 978-1-5386-4353-2.

Wei, M.; Stamos, M.; Veys, S.; Reitinger, N.; Goodman, J.; Herman, M.; Filipczuk, D.; Weinshel, B.; Mazurek, M. L.; and Ur, B. 2020. What twitter knows: characterizing ad targeting practices, user perceptions, and ad explanations through users' own twitter data. In *Proceedings of the 29th USENIX Conference on Security Symposium*, SEC'20. USA: USENIX Association. ISBN 978-1-939133-17-5.

Weintraub, E. L. 2019. Don't abolish political ads on social media. Stop microtargeting. https://www.washingtonpost.com/opinions/2019/11/01/dont-abolish-political-ads-social-media-stop-microtargeting/.

Wu, Y.; Bice, S.; Edwards, W. K.; and Das, S. 2023. The Slow Violence of Surveillance Capitalism: How Online Behavioral Advertising Harms People. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 1826–1837. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701924.

Yuan, C. W. T.; Bi, N.; Lin, Y.-F.; and Tseng, Y.-H. 2023. Contextualizing User Perceptions about Biases for Human-Centered Explainable Artificial Intelligence. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, 1–15. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9421-5.

Zeng, E.; Kohno, T.; and Roesner, F. 2021. What Makes a "Bad" Ad? User Perceptions of Problematic Online Advertising. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–24. Yokohama Japan: ACM. ISBN 978-1-4503-8096-6.