All Things Unequal: Measuring Disparity of Potentially Harmful Ads on Facebook

Muhammad Ali
Northeastern University
mali@ccs.neu.edu

Angelica Goetzen
Max Planck Institute for Software Systems
agoetzen@mpi-sws.org

Alan Mislove
Northeastern University
amislove@ccs.neu.edu

Elissa Redmiles
Max Planck Institute for Software Systems
eredmiles@mpi-sws.org

Piotr Sapiezynski
Northeastern University
p.sapiezynski@northeastern.edu

***Abstract*—Due to their powerful targeting capabilities and low barrier to entry, online advertising platforms are now used by a wide variety of advertisers and businesses to reach consumers. This situation raises concerns that these technologies could be used by malicious actors to advertise content that might be harmful to users, such as outright scams, or other deceptive offers that are more nuanced in their negative impacts on users. This paper aims to measure the prevalence of such potentially harmful advertisements on Facebook, as well as whether there are any biases in how the advertising system delivers them. To do so, we conduct a longitudinal study of a diverse panel of Facebook users ($n = 41$), balanced on key demographics (age, gender, ethnicity/race, educational attainment), who contribute their Facebook advertisements to us. Leveraging the 3,200 ads we collect, we provide insight into the different kinds of ads that comprise each participant's feed, with a focus on characterizing ads that are potentially harmful. Our contributions include qualitatively building a granular understanding of the content users see in their Facebook ads, and quantitatively understanding how the distribution of these different ad types changes across users.**

I. INTRODUCTION

Responding to the growing calls for increased transparency in online advertising, Internet platforms such as Facebook, Google, Twitter, and others have made libraries of ads available to the public for inspection. These initiatives have already enabled a number of investigations, for example into strategies for political advertising [8], misinformation and scam ads [19], [20], inconsistencies in the enforcement of community standards [23], and shortcomings of the ad libraries themselves [10]. Nevertheless, the current transparency model via ad libraries is limited in scope, both in terms of the content, as not all ads are made available, and with respect to questions that can be answered using the published meta information, as critical information is often omitted.

Most notably, the ad libraries fail to capture the variety of unique experiences of *individual* users. Researchers have shown that many harmful phenomena online, which appear negligible in aggregate, can be focused on a small number of users, for whom they can constitute a major part of the online experience [14]. Only by dis-aggregating the data and inspecting the online experience of users as individuals can we learn about the extent of the uncommon but consequential harms [15].

In this paper, we use an alternative approach to investigating social media ads that addresses this shortcoming. We recruit an age-, gender-, education-, and race-balanced sample of participants ($n = 41$) who are active Facebook users, and we collect the "diet" of advertising they are shown. We observe a number of recurrent themes which have a potential to cause individual and societal harms. For example, overexposure to health related content may trigger or aggravate the user's anxiety [22]; skewed delivery of opportunity ads may contribute to exacerbating gender inequality in the job market [1], or racial segregation in housing [5]; personalization of political content may lower the diversity of viewpoints users are presented with [2]. Further, the content of some ads may create harm for certain users, for example by triggering existing mental health challenges (such as with alcohol, gambling, or body image), instabilities (such as with finances), or by misleading them (for example with clickbait and misinformation).

Our measurements show that a vast majority of the ads ($\approx$78%) our participants received pertain to everyday products and services. The remainder, however, fall into potentially consequential categories such as ads for financial services, healthcare, job opportunities, or clickbait. Importantly, ads in these categories are not distributed randomly; we observe that the ad diets—and therefore the online experience—of individual users vary drastically. For example, one of our participants received an ad "diet" that consisted of one of every four ads being identified as clickbait, compared to one in 35 on average across our sample. For another participant, healthcare-related ads constituted 23% of all ads they saw, more than twice the rate for other participants.

Our results provide a first look into how individual experiences on modern ad platforms can vary, where certain users might be more vulnerable to potentially harmful content. This raises novel research questions about the origin of these differences, whether they are due to advertiser targeting, or the platform's ad delivery choices; as well how these differences are perceived by users themselves. Our work lays the foundation for future work to tackle these questions, to better understand the impact of potentially harmful ads on users.

II. Background and Related Work

In this section, we provide a brief overview of how advertising platforms on social media services work, and the research literature on their privacy, security, and fairness properties.

Social media services such as Facebook and Twitter are primarily funded by advertising. These services use the vast amount of data they collect on platform users to build powerful advertising systems that deliver ads to users based on both advertiser-provided constraints (called *targeting*) as well as platform choices (called *delivery*).

During the ad targeting phase, the advertiser uploads their ad creative (text, image, link, etc), and chooses how they want the platform to deliver their ads. Importantly, the advertiser also chooses *which* platform users are eligible to receive their ad. Platforms typically give advertisers a variety of choices of how they can target users, including based on demographics (age, gender, location), interests, or behaviors. Platforms also provide more advanced features, such as the ability to target users based on uploading lists of personally-identifiable information (PII), or by deploying web tracking pixels to third-party sites.

Once an advertiser submits their ad during the targeting phase, the ad delivery phase begins. Each time the platform has an opportunity to show a user an ad, the platform's algorithms must make a choice about which advertiser "wins". Historically, the choice was made primarily based on the bid the advertiser placed; however, modern ad platforms take other factors, such as "relevance" of a particular ad to a particular user, into account in determining the ads that users are shown. Therefore, studying modern advertising platforms is naturally an inquiry both into how advertisers use these systems, as well as how these systems view their users.

Prior work in understanding advertising systems and their impact on consumers has focused on several aspects of the ecosystem, ranging from understanding the data sharing and targeting infrastructure [6], [7], [12], [17], [21], [27], to privacy and transparency concerns [4], [26], as well as algorithmic issues around discrimination and misinformation in ads [1], [3], [11], [24], [25], [28].

Our work, in particular, is closely related to recent efforts in understanding the prevalence of problematic, deceptive, or "bad" ads in the marketplace [28], [29]. The various forms of existing problematic advertising, and their prevalence within contextual web ads has been documented recently [28]. By specifically crawling misinformation sites and using qualitative coding, Zeng el al. provide a look into the many kinds of ads that users might find problematic, such as ads for health supplements and insurance products. Further, there have been efforts towards building a more systematic understanding of which ads are potentially problematic according to users [29]. By conducting a user study, Zeng et al. establish a taxonomy of reasons users find ads problematic or untrustworthy. We leverage their existing taxonomy and focus on a subset of the potentially harmful ad types they identify, which we find evidence of in our Facebook data. Our contribution extends this existing state-of-the-art in our focus on individual-level trends and ads collected from an intentionally demographically diverse participant pool, and is also distinct in our focus on a social media platform rather than contextual web ads.

Variable	Value	n	%
Gender	Female + Non-binary	26	63.41
	Male	15	36.59
Ethnicity/Race	White	23	56.1
	Black	7	17.07
	Hispanic	6	14.63
	Asian	4	9.72
	Native American + Pacific Islander	1	2.44
Education	High School	5	12.2
	Some College	11	26.83
	College	18	43.9
	Graduate School	7	17.07
Income	$< \$25,000$	4	9.76
	\$25,000 - \$50,000	13	31.71
	\$50,000 - \$75,000	5	12.2
	\$75,000 - \$100,000	6	26.83
	\$100,000+	11	26.83
	No Answer	2	4.88

TABLE I
DEMOGRAPHIC BREAKDOWN OF CURRENT PANEL.

Our work is also motivated by results from algorithm auditing and fairness literature that document how the ad delivery algorithms—in an attempt to show users "relevant" content—deliver opportunity ads [1] and political ads [2] in a way that is skewed along gender, race, and political affiliation. In this work we ask whether similar effects are observed for potentially harmful advertising. We therefore reason about the prevalence of harmful advertising not just in terms of their presence on Facebook, but also in how they are distributed across the users.

III. Methodology

Recall that our goal is to conduct a longitudinal study with a diverse panel of participants ($n = 41$), collecting real Facebook advertisements targeted to these participants. To do so, we extend existing methodology [11], [13] to allow participants to donate their individual advertising and targeting data. We then recruited a diverse panel of participants, balanced along multiple demographic variables, to contribute their data. Finally, we analyzed these ads by building a consistent annotation codebook and manually coding over 3,000 ads. The following section describes the steps of our methodology.

A. Panel Recruitment

Ultimately, we aim to recruit a panel of 180 participants with demographics balanced along four axes:

- *Age:* We group participants into two generational cohorts: Born before 1980 (roughly Generation X and older), and Born in or after 1980 (roughly Millennial and younger).
- *Gender:* Based on self-reported gender, we group participants into two cohorts: (1) women and non-binary and (2) men.

- *Education Level:* To account for participants' socioeconomic status (SES), we use participants' self-reported education level as this is a known correlate with SES [9]. We group participants into two cohorts: higher education level (earned at least a Bachelor's degree) or lower education level (did not earn a Bachelor's degree).
- *Race/Ethnicity:* Based on self-reported data, we group participants into four race/ethnicity groups: white, Black, Hispanic, and Asian. We place multi-racial participants into their non-white racial category (if one of their races was white) or into both racial categories (if not).

Thus far, we have recruited 41 participants who have contributed their ads over a period of 2–5 weeks. Ultimately, participants will contribute ads for at least 12 weeks. The demographics of the 41 participants whose data is used in this work is given in Table I.

We recruit participants through two methods: Prolific, and ads on Facebook. Via Prolific, we post ads of our screening survey targeted to combinations of our four chosen demographic variables. On Facebook, we target U.S.-based users who match our demographic criteria. Participants are screened for their Facebook use (must use for at least 10 minutes a day), whether they access Facebook though a desktop or laptop with Chrome or Firefox, whether they use ad blockers, and whether they use software for anonymous browsing (such as Tor or VPN). Participants who meet our screening requirements are sent another survey containing instructions on how to contribute their ads via their browser.

Participants who are chosen from our screening survey are compensated with a $5 Amazon Gift Card upon signing up. They are also compensated with $15 for every month of continued data donation, and $25 for continued contributions for the planned three months.

B. Analysis

We analyze our collected data first qualitatively, to identify the types of ads our participants see, and then quantitatively, to identify differences in the ads they see and how they were delivered.

Qualitative Analysis: Understanding Ad Types. To consistently categorize each participant's ads into a fixed set of ad types, we inductively developed a codebook from a subset of our data collection, cross-referenced against Zeng et al.'s existing categories of problematic ads [28], [29]. Specifically, we ran a pilot phase of our data collection with 7 participants (not included in the 41 participants whose data is analyzed in this paper), for a month between June and July 2021, to inform our codebook. Our complete codebook, with definitions, is described in Section IV. To ensure robustness of our annotation process, we double-code a subset of our data, where we find inter-annotator agreement through Cohen's Kappa to be $\kappa = 0.75$.

Quantitative Analysis: Identifying Differences. Once each ad from our data collection has been annotated, we characterize the prevalence and disparity of different ad types by comparing their distribution across different participants. By using individually annotated ads, our goal in the quantitative analyses is to understand what proportion of each ad type are contributed by each participant, and how that varies across the overall dataset. When comparing proportions, we employ the χ^2 test for equality of proportions for establishing difference of proportions.

C. Ethics

We take extreme caution with our data due to its ability to uniquely identify each participant's repository of contributed ads. Our dataset is kept under strict access control on our backend server, accessible only within our institution. We further use only pseudonymous identifiers to identify each participant's ads. Our project has been reviewed and approved by our institutions' ethics review board and institutional review board, respectively.

D. Limitations

Our participant sample comes from different parts of the U.S. While we originally aimed to recruit in one geographic area to control for geographic variance in ads, we were unable to recruit sufficient numbers of diverse participants with this constraint. We recognize that participants of differing demographics may have different experiences with regard to their location in the U.S. (e.g., socioeconomic backgrounds may be subjective to the cost of living in different locations), and that these differences may impact the ads they receive.

Additionally, the data used in our work comes from participants recruited over the course of two months. We acknowledge the potential effects that timing of sign-up between participants may have had on the ads they received.

Further, while we draw from Zeng et al.'s taxonomy of problematic ads, which was based on people's reactions to ads they were shown, we do not ask our participants their opinions of their ads. Thus, our characterization of these ads is based on our positionality as researchers. We aim to incorporate participants' reactions to ads in future work.

Finally, we note that when signing up the users can chose a gender identity other than "man" or "woman". Here, however we analyze gender as a binary as we currently do not have any non-binary participants in our sample.

IV. Results

In this section, we describe our qualitative codebook and our measurement results from applying it to our collected dataset of 3,200 ads.

A. Codebook of Facebook Ad Types

As part of our qualitative analysis, we develop a consistent codebook that broadly categorizes the different kinds of ads that we observe for our participants. We define our codes below; a more detailed description of each code alongside examples is provided in Table III in the Appendix.

- *Neutral:* Ads that simply seek to advertise a product, service or apolitical message, which we have no reason to believe would harm users.

- *Opportunity:* Ads that present any sort of financial, employment, housing, or other miscellaneous opportunity to users.
- *Sensitive:* Ads that might be sensitive or triggering for some users, such as body image (weight loss), gambling, or alcohol ads.
- *Potentially Prohibited:* Ads that, according to Facebook's "Prohibited Content" ad policy,[1] may not actually be allowed on the website.
- *Clickbait:* Ads that omit information to entice users, are unclear about the advertised product, and/or use sensational and loud language in the description, or images with very dense text.
- *Financial:* Ads that contain products, services or messages related to managing finances, loans, building credit, and other financial processes. This code excludes product or service ads that may reference purchases or savings, but whose central message is not about a user's finances.
- *Healthcare:* Ads that contain products, services or messages related to healthcare, fitness, mental and physical wellness, or physical appearance.
- *Class Action Lawsuit:* Ads that contain information related to class action lawsuits that may be applicable to users.
- *Political:* Ads that contain any overt references to political subject matters.

We note that codes such as *Class Action Lawsuit* and *Political* are used mainly to distinguish ads that might be out of scope for our investigation of potential harms. *Opportunity* and *Financial* include a wide spectrum of content, ranging from benign and useful to potentially problematic. *Neutral* is used as the default/baseline code to separate out content that does not meet any of our other categories. Ads can be annotated with multiple codes, except when they are marked as *Neutral*; in our dataset of 3,200 ads, a total of 134 ads have multiple codes.

B. Prevalence of Potentially Harmful Advertising

Ad Type	Frequency	%
Neutral	2610	78.1
Healthcare	312	9.34
Opportunity	207	6.19
Clickbait	69	2.06
Financial	41	1.23
Sensitive	34	1.02
Political	28	0.84
Potentially Prohibited	23	0.69
Class Action Lawsuit	18	0.54

TABLE II
PREVALENCE OF EACH AD TYPE (FROM OUR CODEBOOK) IN OUR COLLECTED DATASET.

We first try to understand the composition of our dataset in terms of our identified ad types. Table II shows the distribution of ad types in our collected dataset of 3,200 ads. As expected we see that the largest fraction (≈78%) of our dataset are labeled *Neutral* (i.e., benign product and services ads). However, the remaining 22% fall into at least one of the other categories.

Healthcare ads (e.g. for supplements and fitness products) comprise around 9% of our dataset, making them the second most common ad type our participants saw. Next, 6% of our data consists of *Opportunity* ads, and a little over 1% of *Financial* ads. We qualitatively find that ads for financial services and education/job opportunities are quite diverse in their content, and some of these may be predatory. For example, an ad from California Debt Relief that links to a nondescript information form, and no information about the purported program itself; debt relief services in general may be exploitative and are viewed as being "risky" to engage with by the U.S. Consumer Financial Protection Bureau.[2] Others may be beneficial, such as ads for budgeting apps.

Next, we observe that ads with *Clickbait*, content that is *Potentially Prohibited* by Facebook's advertising rules, and *Sensitive* ads (e.g. gambling and dieting) together constitute nearly 4% of our dataset. Finally, ads for *Political* content and *Class Action Lawsuits* comprise a very small fraction of our data, and thus we do not consider them for the remainder of the paper.

Overall, these results show that while a vast majority of our collected Facebook ads are for benign products and services, there is a noticeable presence of content that may be harmful to the user, or there is potential for the platform's personalization to create harm by overexposing users to certain content (for example healthcare), or unequally distributing it (e.g. in the case of opportunity ads). This highlights the need to better understand how these ads are distributed across our participants: how individual participants' experiences might vary, and whether some may receive a disproportionate fraction of potentially harmful ads.

C. Ad differences across participants

To understand how the distribution of potentially harmful ads changes compared to more benign content across participants, we measure the ad type distribution (or "ad diet") for each participant. Figure 1 shows distribution of ad types across our panel. While non-*Neutral* ads constitute a small minority of our overall dataset (Table II), they are *not* a small minority of ads for every participant. We see in Figure 1 that the proportion of non-*Neutral* content differs significantly (omnibus test for equality of proportions: $\chi^2 = 472.10$; $p \ll 0.05$) across our participants.

We also observe that participants can be susceptible to different kinds of potential harms through these ads. For example, in Figure 1, only ∼40% of Participant 31's content is *Neutral*, compared to ∼97% for Participant 0. Similarly, we find that more than 20% of Participant 26's ads are *Clickbait*, and close to 10% of Participant 27's ads are *Financial*; both of these

[1] https://www.facebook.com/policies/ads/prohibited_content

[2] https://www.consumerfinance.gov/ask-cfpb/what-are-debt-settlementdebt-relief-services-and-should-i-use-them-en-1457/.

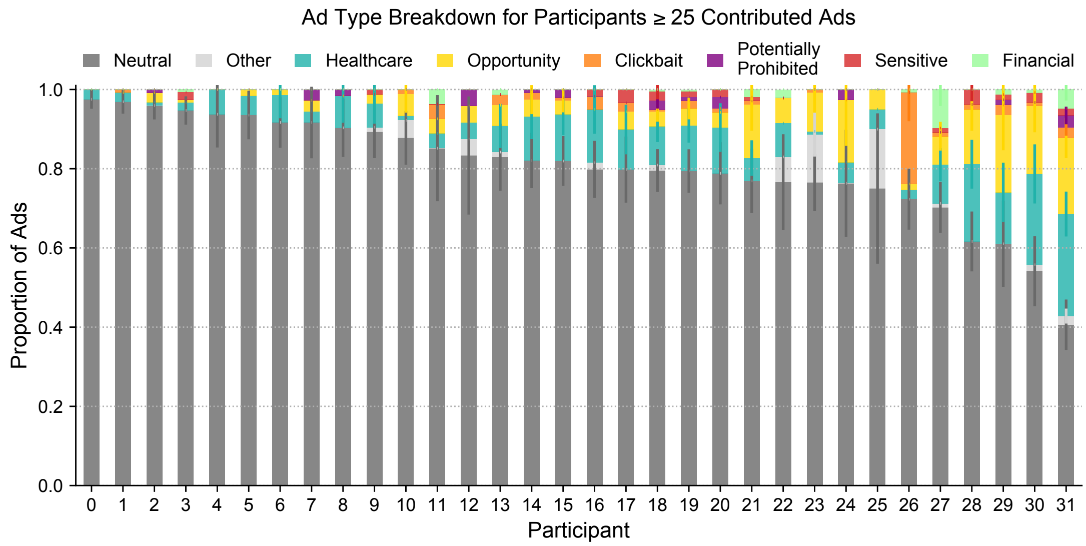

Fig. 1. Difference in ad types, in terms of proportion of all ads contributed, across participants in our panel. Error bars (binomial confidence intervals) are computed via the normal approximation method. Participants are sorted left-to-right in descending order of their fraction of non-*Neutral* ads. Error bars are not shown when number of ads in category are less than 5.

disproportionate compared to other participants. These differences highlight the importance of understanding individual user experiences, which might be significantly different than the general prevalence of potentially problematic advertising on Facebook's platform.

D. Disparity in exposure to potentially harmful ads

Given the presence of potentially harmful ads in our dataset, and the fact that individual participants' ad compositions can vary vastly, we next look deeper into the individual categories to see how "evenly" they are spread across our participants. Specifically, for each ad type, we observe the Lorenz curve to understand the magnitude of inequality that exists in its distribution over our panel. A Lorenz curve [18]—conventionally a helpful tool in understanding inequality in wealth—can also be useful for understanding inequality and disparities on online platforms [16]. It visualizes the cumulative fraction of wealth against the cumulative population size, helping illustrate what fraction of wealth (or ads in our case) is saturated for what fraction of the population. The Lorenz curve is also closely related to the Gini coefficient, which is a scalar measure of inequality, and can be computed directly from the area between the curve and the hypothetical line of equality (which represents a uniform distribution of wealth). Recall that a Gini coefficient of 0 represents a perfectly even distribution across participants, while a Gini coefficient of 1 represents perfect inequality (all ads are shown to a single participant and none to others). Figure 2 (a) shows the Lorenz curves, and Figure 2 (b) shows the Gini coefficients for the potentially harmful ad types identified via our codebook.

We see that potentially harmful ad types such such as *Sensitive*, *Clickbait* and *Potentially Prohibited* have a significantly higher level of inequality in contributions compared to the *Neutral* ads. Figure 2 (a) shows how roughly 80% of the participants observe only ≈20% of the *Sensitive*, *Clickbait* and *Potentially Prohibited* ads, and the top 20 percentile of participants see the large remainder of these ads. This effect is

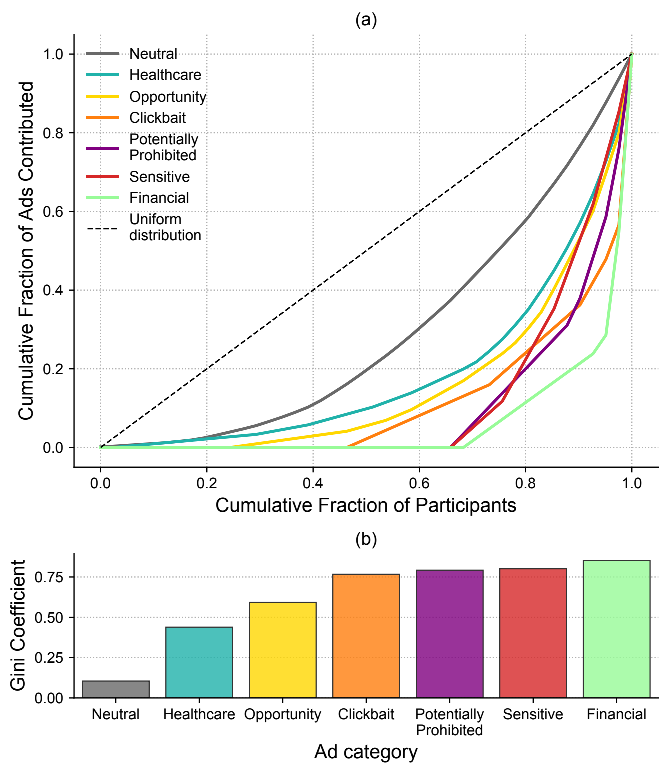

Fig. 2. **(a)** Lorenz curves for *neutral* and potentially harmful ad types, showing cumulative fraction of contributed ads as a function of the fraction of population that contributed them; curves closer to the line of equality represent a more even distribution over participants. **(b)** Gini coefficients for each ad type, computed over the fraction of each participants' ad contributions in that category, a Gini coefficient of zero represents equality, suggesting that the ad type is evenly distributed over participants in the dataset.

even more pronounced for *Financial* ads, which are saturated for only the top 10 percentile of participants. Figure 2 (b) also shows the high Gini coefficients for these potentially harmful ad types. Note that these disparities exist in contrast to the *Neutral* ads, which are much more evenly distributed across our panel, and whose Lorenz curve is closer to the uniform distribution.

These disparities in ad type distributions confirm what we observed in the previous section: that different categories of potentially harmful ads constitute a significant fraction of the ad diet for a small subset of participants, rather than being randomly or evenly distributed.

V. Discussion and Conclusion

Throughout this work we recognize that there are a number of content categories in social media ads that can be harmful to the users. We also stress that in some cases it is not the ad content that makes for a potential harm—the harms may also stem from an uneven distribution opportunities, or overexposure to triggering content.

Taken together, our results provide a first look into individual-level distribution of potentially harmful advertisements on Facebook. While the overall prevalence of such ads is low, we find that the distribution of these ads across our participants is not even, with a few participants receiving a majority of ads that we classified as potentially harmful. Looking across these potentially harmful categories, we observe a consistent pattern: the ads in these categories are distributed in a more unequal manner compared to benign ads.

Our results also indicate that further work is necessary. *First*, we need to properly understand *why* different users have such different experiences. Specifically, do users from different demographics tend to be targeted more than others? How about users in different geographies? The differences in ad distribution could be due to advertiser targeting, platform delivery choices, or a complex interaction between both of these processes. We must scale up our user population, as well as include the advertiser-defined targeting information in our analysis in order to be able to answer these and other critical questions. *Second*, our results are currently based only on our own classifications. So far, we have not yet involved the actually affected population—the Facebook users—directly. To fully understand the impacts and potential harms that these ads are having on users, we need to ask the participants themselves.

We hope our work lays the groundwork for future investigations in understanding potential harms, and informing better content policies on advertising platforms.

Acknowledgements

The authors thank Karl Satinitigan for running pilot survey experiments, and doing qualitative coding for an earlier version of this study. This work was funded by NSF grants CNS-1916020, CNS-1955227, and Mozilla.

References

[1] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead To Biased Outcomes. In *ACM Conference on Computer Supported Cooperative Work*, Austin, Texas, USA, November 2019.

[2] Muhammad Ali, Piotr Sapiezynski, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. Ad Delivery Algorithms: The Hidden Arbiters Of Political Messaging. In *ACM International Conference on Web Search and Data Mining*, Jerusalem, Israel, 2021.

[3] Muhammad Ali, Piotr Sapiezynski, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. Ad Delivery Algorithms: The Hidden Arbiters Of Political Messaging. In *ACM International Conference on Web Search and Data Mining*, Jerusalem, Israel, 2021.

[4] Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. Investigating Ad Transparency Mechanisms In Social Media: A Case Study Of Facebook's Explanations. In *Network and Distributed System Security Symposium*, San Diego, California, USA, February 2018.

[5] Joshua Asplund, Motahhare Eslami, Hari Sundaram, Christian Sandvig, and Karrie Karahalios. Auditing Race And Gender Discrimination In Online Housing Markets. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 24–35, 2020.

[6] Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, and Christo Wilson. Tracing Information Flows Between Ad Exchanges Using Retargeted Ads. In *Proceedings of Usenix Security*, Austin, TX, August 2016.

[7] Muhammad Ahmad Bashir and Christo Wilson. Diffusion Of User Tracking Data In The Online Advertising Ecosystem. *Proceedings on Privacy Enhancing Technologies (PoPETS)*, 2018(4), July 2018.

[8] NaLette Brodnax and Piotr Sapiezynski. From Home Base To Swing States: The Evolution Of Digital Advertising Strategies During The 2020 Us Presidential Primary. *Political Research Quarterly (Forthcoming)*, 2022.

[9] Paul T Cirino, Christopher E Chin, Rose A Sevcik, Maryanne Wolf, Maureen Lovett, and Robin D Morris. Measuring socioeconomic status: reliability and preliminary validity for different approaches. *Assessment*, 9(2):145–155, 2002.

[10] Laura Edelson, Tobias Lauinger, and Damon McCoy. A Security Analysis Of The Facebook Ad Library. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 661–678. IEEE, 2020.

[11] Laura Edelson, Shikhar Sakhuja, Ratan Dey, and Damon McCoy. An Analysis Of United States Online Political Advertising Transparency. https://arxiv.org/abs/1902.04385.

[12] Steven Englehardt, Dillon Reisman, Christian Eubank, Peter Zimmerman, Jonathan Mayer, Arvind Narayanan, and Edward W. Felten. Cookies That Give You Away: The Surveillance Implications Of Web Tracking. In *International World Wide Web Conference*, Florence, Italy, April 2015.

[13] Facebook Political Ad Collector Searchable Interface. https://projects.propublica.org/facebook-ads/.

[14] Dorota Glowacka and Karolina Iwanska. Algorithms Of Trauma: New Case Study Shows That Facebook Doesn't Give Users Real Control Over Disturbing Surveillance Ads. https://en.panoptykon.org/algorithms-of-trauma.

[15] David Lazer, Eszter Hargittai, Deen Freelon, Sandra Gonzalez-Bailon, Kevin Munger, Katherine Ognyanova, and Jason Radford. Meaningful Measures Of Human Society In The Twenty-first Century. *Nature*, 595(7866):189–196, 2021.

[16] Tomo Lazovich, Luca Belli, Aaron Gonzales, Amanda Bower, Uthaipon Tantipongpipat, Kristian Lum, Ferenc Huszar, and Rumman Chowdhury. Measuring disparate outcomes of content recommendation algorithms with distributional inequality metrics. *arXiv preprint arXiv:2202.01615*, 2022.

[17] Mathias Lecuyer, Riley Spahn, Yannis Spiliopolous, Augustin Chaintreau, Roxana Geambasu, and Daniel Hsu. Sunlight: Fine-grained Targeting Detection At Scale With Statistical Confidence. In *ACM Conference on Computer and Communications Security*, 2015.

[18] Max O Lorenz. Methods of Measuring the Concentration of Wealth. *Publications of the American statistical association*, 9(70):209–219, 1905.

[19] Jeremy B. Merrill and Marshall Allen. "Trumpcare" Does Not Exist. Nevertheless Facebook And Google Cash In On Misleading Ads For "garbage" Health Insurance. https://bit.ly/trumpcarePropublica.

[20] Jeremy B. Merrill and Hanna Kozlowska. A Precious-metals Scheme Used Fear And Facebook To Trick Older Conservatives Out Of Their Savings. https://qz.com/1749695/facebook-ads-targeted-fox-news-fans-for-shady-silver-coin-scheme/.

[21] Hooman Mohajeri Moghaddam, Gunes Acar, Ben Burgess, Arunesh Mathur, Danny Yuxing Huang, Nick Feamster, Edward W Felten, Prateek Mittal, and Arvind Narayanan. Watching you watch: The tracking ecosystem of over-the-top tv streaming devices. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 131–147, 2019.

[22] Madhumita Murgia. Time To Turn Off Facebook's Digital Fire Hose. https://www.ft.com/content/d5dcfece-4e3c-4937-81ac-20dc736c4c27.

[23] Craig Silverman and Ryan Mac. Facebook Promised To Label Political Ads, But Ads For Biden, The Daily Wire, And Interest Groups Are Slipping Through. https://www.buzzfeednews.com/article/craigsilverman/facebook-biden-election-ads.

[24] Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. On The Potential For Discrimination In Online Targeted Advertising. In *Conference on Fairness, Accountability, and Transparency*, New York, New York, USA, February 2018.

[25] Latanya Sweeney. Discrimination In Online Ad Delivery. In *Social Science Research Network Working Paper Series*, 2013.

[26] Giridhari Venkatadri, Yabing Liu, Athanasios Andreou, Oana Goga, Patrick Loiseau, Alan Mislove, and Krishna P. Gummadi. Privacy risks with facebook's pii-based targeting: Auditing a data broker's advertising interface. In *IEEE Symposium on Security and Privacy*, San Francisco, California, USA, May 2018.

[27] Giridhari Venkatadri, Piotr Sapiezyński, Elissa M. Redmiles, Alan Mislove, Oana Goga, Michelle Mazurek, and Krishna P. Gummadi. Auditing Offline Data Brokers Via Facebook's Advertising Platform. In *International World Wide Web Conference*, San Francisco, California, USA, May 2019.

[28] Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. Bad news: Clickbait and deceptive ads on news and misinformation websites. In *Workshop on Technology and Consumer Protection*, 2020.

[29] Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. What makes a "bad" ad? user perceptions of problematic online advertising. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–24, 2021.

Appendix

Code	Definition	Examples
Neutral	Ads that simply seek to advertise a product, service or apolitical message	**SiriusXM:** "Experience the limited-run Prince Channel FREE on SiriusXM..." **Kate Spade:** "everything's up to 75% off! shop surprise sale now."
Opportunity	Ads that present any sort of financial, employment, housing, or other miscellaneous opportunity to users. For example, ads for degree programs, jobs or gig-work, fellowships, scholarships etc.	**SMU Boot Camps:** "The first step to a new career in cybersecurity starts here — apply today!" **Google:** "Free Virtual Workshop: Create a Career Plan to Get the Job You Want"
Potentially Prohibited	Ads that may be prohibited according to Facebook's prohibited content policies. For example, Tobacco, drugs, unsafe dietary supplements, multi-level marketing, weapons etc.	**Tru Niagen:** "Think Differently About Cellular Recovery; Healthy Aging from the Inside Out" **Healthy Mama:** "After reading this, I am so relieved I'm giving my kids the right type of vitamin"
Clickbait	Ads that omit information to entice users, are unclear about the advertised product, and/or use sensational and loud language in the description, or images with very dense text.	**Minting Nickels:** "When rent goes above $800 per month, you've got to make these changes..." **Dad Patrol:** "20+ Times People Were Actually Satisfied With How Well Things Turned Out"
Financial	Ads that contain products, services or messages related to managing finances, building credit, and other financial processes. This code excludes product or service ads that may reference purchases or savings, but whose central message is not about a user's finances.	**Upstart:** "Pay off $1,000-$50,000 today. Checking your rate doesn't hurt your credit score!" **American Express Business:** "Earn up to 180,000 Hilton Bonus Points. Terms apply."
Healthcare	Ads that contain products, services or messages related to healthcare, fitness, mental and physical wellness, or physical appearance.	**Cerebral:** "Get the care you need from the comfort of your home. Affordable mental health is here for you." **The Pill Club:** "Did you know? You can get birth control prescribed online AND delivered for free!"
Class Action Lawsuit	Ads that contain information related to class action lawsuits that may be applicable to users.	**Cough Syrup Lawsuit:** "Purchased Non-Drowsy Cough Syrup? You may have a claim for compensation". **Zinus Mattress Lawsuit:** "Seek justice and compensation for: coughing, wheezing, itchy skin..."
Political	Ads that contain any overt references to political subject matters.	**Unite for Reproductive and Gender Equity:** "Anti-abortions politicians in Texas are doing everything they can to cut off access to abortion and shut down clinics..." **NC Warn:** "Climate disasters are devastating NC communities. Stop Duke Energy's climate-wrecking gas expansion of power plants."

TABLE III
CODEBOOK DEFINITIONS, WITH EXAMPLES FROM COLLECTED DATASET.