Course: TIMG 5204A: Responsible AI Ethics [S23]

Assignment #2: Analyzing Bias

Topic: Facebook Ad's algorithm

By Group #2:

Ahnaf Mohsin

Arezoo Khojasteh Abbasi

Manas Agarwal

Maryam Agha Hosseinalishirazi

Rajdeep Jinegar

Shahrzad Kahrizi

Date of Submission: June 12th, 2023

Table of Content

Section 1: Introduction

Through extensive collaborative research, our group has explored the inner workings of Facebook's ad algorithm, unearthing concerning evidence of inherent bias within its system that has created controversy. This introductory overview aims to provide a glimpse into our findings, shedding light on the discriminatory nature of Facebook's ad algorithm and the potential implications it carries for targeted advertising.

Facebook's ad algorithm serves as a powerful tool for advertisers to reach their desired audience with precision and effectiveness. Operating on machine learning principles, the algorithm analyzes vast amounts of user data, including demographics, interests, and behaviors, to deliver tailored ads that maximize engagement and conversion rates [1][4]. Our collective research has uncovered significant biases entrenched within Facebook's ad algorithm, warranting attention and concern. These biases manifest in ways that perpetuate discrimination across critical domains such as housing, employment, and credit [1]. Despite its aim to deliver personalized ads, the algorithm's microtargeting capability inadvertently creates exclusionary effects, compromising the principles of fairness and equal opportunity.

The biases exhibited by Facebook's ad algorithm encompass various dimensions. Primarily, demographic biases arise when the algorithm targets or excludes individuals based on factors like age, gender, race, or ethnicity [2][3]. Additionally, interest-based biases occur as the algorithm tends to reinforce existing user preferences, potentially limiting exposure to diverse perspectives. Lastly, feedback loop biases emerge due to the algorithm's reliance on historical data, perpetuating discriminatory patterns and exacerbating inequality [1]. The presence of bias within Facebook's ad algorithm raises profound ethical concerns. Discriminatory practices in ad targeting can reinforce societal inequalities, hinder access to vital services, perpetuate stereotypes, and impede progress towards a more inclusive society. Moreover, such biases can undermine regulatory efforts aimed at combating discrimination in critical areas such as housing and employment.

In light of these findings, it is crucial for Facebook to take proactive measures to address algorithmic bias. Enhancing transparency by providing clearer guidelines for advertisers and users regarding ad targeting is a vital step. Regular audits and assessments of the algorithm's fairness and bias must be implemented to ensure accountability. Furthermore, incorporating diverse

perspectives into the algorithm's development and monitoring processes can aid in identifying and rectifying potential biases.

After research, our group has identified the biases within Facebook's ad algorithm, exposing the potential for discrimination in targeted advertising. Recognizing and rectifying these biases is essential for fostering fairness, equal opportunity, and a more inclusive digital advertising landscape. By prioritizing transparency, accountability, and ongoing evaluation, Facebook can strive towards an algorithm that empowers users while upholding ethical standards and promoting a diverse and equitable society. The rest of our report will dissect the data being collected, the biases that were found in Facebook's ad algorithm, and mitigation strategies for these biases.

Section 2: Data and Mechanism of Bias

In this section, we will explore the data and mechanisms of bias in Facebook's ad algorithm. We will identify and explain the existing biases and any additional ones discovered through research and analyze how they impact the outcomes of the algorithm and decision-making processes.

Types of Data Used: Facebook's ad algorithm relies on a vast amount of data to make informed decisions about ad delivery. This data includes users' demographic information, such as age, gender, and location, as well as their interests, behaviors, and past interactions on the platform. Advertisers can also provide specific targeting criteria to reach their desired audience. The algorithm utilizes this data to optimize ad delivery and maximize user engagement [1].

Existing Biases in the Data/System: Research, (such as the paper by M. Ali et al. [1]), has highlighted several biases in Facebook's ad algorithm. These biases can stem from various sources, such as biased training data, inherent societal biases, and the optimization process itself. Some of the identified biases include [1]:

1. Demographic Biases: The algorithm may disproportionately deliver ads based on factors like race, gender, and age, even when advertisers do not explicitly specify such preferences. This can result in certain demographic groups being overexposed or excluded from particular ads.

2. Behavioral Biases: The algorithm may learn from users' past behaviors and interactions, which can inadvertently perpetuate existing biases. For example, if certain groups have historically been underrepresented in ad targeting, the algorithm may perpetuate this bias by delivering ads predominantly to the majority population.
3. Proxy Biases: The algorithm may use proxy attributes that are correlated with sensitive characteristics, leading to discriminatory outcomes. For instance, if the algorithm relies on zip codes as a proxy for race, it can indirectly lead to discriminatory ad delivery based on racial demographics.
4. Contextual Biases: The algorithm may fail to adequately consider the context in which ads are delivered, leading to inappropriate or offensive ad placements. For example, an ad promoting a product or service may be displayed alongside sensitive or controversial content, causing harm to both users and advertisers.
5. Cultural Biases: The algorithm may not sufficiently account for cultural differences and preferences, resulting in ads that are irrelevant, offensive, or insensitive to certain cultural groups. This can contribute to the perpetuation of cultural stereotypes and misunderstandings.
6. Algorithmic Fairness Biases: The algorithm may exhibit biases that violate principles of fairness, such as treating similar individuals or groups differently without justifiable reasons. This can result in discriminatory outcomes and unequal treatment of users.
7. Accessibility Bias: Accessibility bias occurs when the algorithm inadvertently favors or excludes individuals with certain disabilities or limitations. For example, if the algorithm primarily relies on visual cues or content, it may not adequately cater to users with visual impairments, thus limiting their access to relevant ads. This bias can result in unequal treatment and hinder inclusive advertising practices.

Machine Learning Techniques Deployed: Facebook uses a combination of machine learning algorithms and techniques to show specific ads to specific users. The exact details of the algorithm's architecture and methodology are proprietary, but they are likely to involve a combination of supervised learning, reinforcement learning, and recommendation systems [1][4]. These techniques enable the algorithm to learn from historical data and user interactions, identifying patterns and making predictions about user preferences and engagement. The primary algorithm used for ad targeting on Facebook is called the Facebook Ads Auction. The Facebook

Ads Auction employs a variety of signals to determine which ads are most relevant to a particular user. These signals include demographic information, interests, behavior, and engagement history on the platform. The algorithm analyzes these signals along with the ad campaign's targeting parameters and the bidding strategies of advertisers to determine which ad to show to a specific user. Facebook also employs various machine learning techniques, such as predictive modeling and clustering, to better understand user preferences and predict the likelihood of engagement with specific ads [1]. This helps optimize the ad delivery process by showing ads that are more likely to resonate with individual users. It's important to note that Facebook's algorithms are proprietary, and the specific details of their implementation are not publicly disclosed. Therefore, the exact algorithms and techniques used by Facebook for ad targeting may evolve and change over time as they continue to refine their systems [4].

Consequences of Biases on Human Life: The biases present in Facebook's ad algorithm can have significant consequences for individuals and society as a whole and can also influence the algorithm's ad delivery decisions and shape the targeting and exposure of ads to users. Some potential consequences include [1]:

1. Reinforcement of Stereotypes: Biased ad delivery can reinforce societal stereotypes by perpetuating differential treatment based on characteristics such as race, gender, or age [2]. This can contribute to the marginalization and exclusion of certain groups, hindering their access to opportunities and resources.
2. Economic Inequities: Biased ad delivery can result in economic disparities as it may favor or exclude certain groups from targeted advertising. This can impact business opportunities, employment prospects, and access to relevant products or services, further exacerbating existing socioeconomic inequalities.
3. Discriminatory Practices: Biased ad delivery can enable discriminatory practices in areas like housing, employment, and credit opportunities. Advertisers may inadvertently or intentionally target specific demographics while excluding others, perpetuating discrimination and hindering equal access to essential services and opportunities.
4. Psychological Effects: Biased ad delivery can have psychological effects on individuals who consistently receive ads that reinforce stereotypes or exclude them from certain

opportunities. This can lead to feelings of exclusion, reduced self-esteem, and a sense of unfair treatment, impacting their overall well-being.

5. Impact on User Perception and Engagement: Contextual biases can lead to ads being placed in inappropriate or offensive contexts, which can negatively impact user perception and engagement with the ads, potentially reducing their effectiveness.
6. Reduced Ad Effectiveness and User Satisfaction: Cultural biases can result in the delivery of ads that are irrelevant or offensive to certain cultural groups. This can lead to reduced ad effectiveness and user satisfaction, as the ads may fail to resonate with the targeted audience.
7. Unequal Treatment and Discriminatory Outcomes: Algorithmic fairness biases, such as treating similar individuals or groups differently without justifiable reasons, can lead to discriminatory outcomes. This unequal treatment can result in harm to affected individuals or communities and perpetuate inequalities in the access to services, opportunities, or resources.
8. Exclusion and Digital Inequities Outcomes : Accessibility bias can have significant consequences on the effectiveness and fairness of the ad algorithm. By not accounting for the needs of individuals with disabilities, the algorithm may exclude them from accessing relevant advertisements, thus limiting their opportunities and experiences. This can perpetuate digital inequities and further marginalize individuals with disabilities, hindering their ability to engage with products, services, and opportunities that are essential for their well-being and participation in society.

Addressing Biases in the System: Facebook has taken steps to address biases in its ad algorithm. The company has committed to improving transparency and accountability by providing advertisers with more information about how their ads are delivered and introducing tools to test and verify the fairness of ad campaigns, and also incorporating accessibility guidelines into the algorithm, prioritizing features like alternative text, screen reader support, and compatibility with assistive technologies. They have also established partnerships with external organizations to conduct audits and assess the potential biases in their systems. However, it is important to note that mitigating biases in algorithmic systems is a complex task. Ongoing efforts are required to ensure continuous evaluation, monitoring, and improvement of the ad algorithm to minimize biases and promote fairness [1].

Section 3: Mitigation

The article [1] addresses the biases in Facebook's ad delivery process and provides insights into the mechanisms behind and impact of these biases. While the paper [1] focuses on highlighting the issue rather than proposing specific mitigation strategies, it sheds light on the potential discriminatory outcomes created by advertising platforms. Nevertheless, based on the information provided, we can consider potential mitigation approaches [1]:

1. Transparency and Accountability: One way to address biases is to improve transparency and accountability in the ad delivery process. Advertising platforms like Facebook can provide advertisers with more detailed information about how their ads are delivered. By sharing insights into the factors that influence ad delivery, advertisers can better understand the potential biases and make informed decisions. In addition to providing detailed information about ad delivery, advertising platforms can establish clear guidelines and policies on fair and inclusive ad delivery practices. By setting transparent standards, platforms can hold themselves accountable and provide advertisers with a framework for creating non-discriminatory ad campaigns.
2. Fairness Testing Tools: Introducing fairness testing tools can help advertisers assess the fairness of their ad campaigns. By providing advertisers with the ability to test and verify the potential biases in their ads, they can identify and rectify any unintended discriminatory outcomes before launching the campaigns. These tools can also enable advertisers to optimize their ads to reach a more diverse and inclusive audience. Alongside fairness testing tools, advertising platforms can offer educational resources and guidelines to help advertisers understand and navigate potential biases in ad delivery. These resources can provide insights into best practices for creating inclusive ad campaigns, including suggestions for diverse targeting and messaging strategies.
3. External Audits: Collaboration with external organizations to conduct audits can be another mechanism to address biases. By inviting independent third-party organizations to assess the potential biases in their systems, advertising platforms can gain valuable insights and recommendations for improving fairness. External audits can provide an objective evaluation of the algorithms and help identify areas for improvement. To enhance the effectiveness of external audits, advertising platforms can collaborate with auditors to

develop standardized methodologies and metrics for evaluating fairness in ad delivery. This collaboration can help establish industry-wide benchmarks for fairness and enable platforms to benchmark their performance against established standards.

4. Algorithmic Improvements: Continuously evaluating and improving the ad delivery algorithms is crucial to minimizing biases. Advertising platforms can invest in research and development to enhance the fairness of their algorithms. This may involve refining the machine learning models, incorporating additional fairness metrics, and conducting ongoing analysis of the outcomes to detect and mitigate biases. In addition to refining machine learning models, advertising platforms can invest in research to understand the underlying causes of biases in ad delivery algorithms. By identifying specific features or patterns that contribute to biased outcomes, platforms can develop mitigation techniques, such as debiasing algorithms or adjusting weighting factors, to reduce discrimination.
5. Ethical Considerations: Considering ethical implications throughout the ad delivery process is essential. Ad platforms should prioritize ethical decision-making, ensuring that the algorithms and processes are designed to be fair and inclusive. This involves taking proactive measures to identify potential biases, engaging in regular discussions on ethical considerations, and promoting a culture of fairness within the organization. Furthermore, promoting ethical considerations involves incorporating fairness and inclusivity as core principles in the design and implementation of ad delivery algorithms. This can include adopting algorithmic fairness techniques, such as counterfactual fairness or demographic parity, to ensure that ad delivery does not disproportionately disadvantage or exclude certain groups based on protected attributes.

By implementing these mitigation approaches, advertising platforms like Facebook can take steps towards addressing the biases in their ad delivery process and strive for more equitable outcomes. It is important for policymakers, platforms, and advertisers to collaborate in order to create a fair and inclusive digital advertising ecosystem.

References

[1] M. Ali et al., "Discrimination through optimization: How facebook's ad delivery can lead to skewed outcomes," arXiv.org, https://arxiv.org/abs/1904.02095 (accessed Jun. 11, 2023).

[2] Person and P. Dave, "Study flags gender bias in Facebook's ADS tools," Reuters, https://www.reuters.com/technology/study-flags-gender-bias-facebooks-ads-tools-2021-04-09/ (accessed Jun. 11, 2023).

[3] C. Mello-Klein, "Facebook's ad delivery algorithm is discriminating based on race, gender and age in photos, northeastern researchers find," Northeastern Global News, https://news.northeastern.edu/2022/10/25/facebook-algorithm-discrimination/#:~:text=The%20researchers%27%20previous%20work%20showed,Black%20women%2C%20according%20to%20Mislove (accessed Jun. 11, 2023).

[4] D. M. V. A. S. Jennie Lyon, "The machine learning strategy that makes Facebook ads more profitable," LinkedIn, https://www.linkedin.com/pulse/machine-learning-strategy-makes-facebook-ads-more-jennie (accessed Jun. 11, 2023).