Technological Forecasting & Social Change 201 (2024) 123204

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Artificial intelligence and algorithmic bias? Field tests on social network with teens☆

G. Cecere[a], C. Jean[b,*], F. Le Guel[c], M. Manant[c]

[a] Institut Mines-Télécom, Business School, LITEM, 9, rue Charles Fourier, 91000 Évry-Courcouronnes, France
[b] Grenoble Ecole de Management, 38000 Grenoble, France
[c] Université Paris-Saclay, RITM, 54, boulevard Desgranges, 92330 Sceaux, France

ARTICLE INFO

JEL classification:
L02
L05
M3

Keywords:
Artificial intelligence
Algorithmic bias
Field tests
Audit
Online social network

ABSTRACT

Artificial intelligence (AI) is a general purpose technology that is used in many sectors. However, automated decision-making powered by AI algorithms can lead to unintended outcomes, especially in the context of online platforms. The lack of transparency related to AI algorithms and their categorization methods make practical insights into effective management of the risks associated to their utilization of crucial importance. We address these issues through two field tests aimed at mitigating biases in online science, technology, engineering, and mathematics (STEM) education-related ads targeting teenagers. We conducted online ad campaigns involving gender-unspecific, women-specific, and gender-neutral ads targeted at young social network users. Our findings show that inclusion in the ad of a gender-oriented message tends to alleviate algorithmic gender bias but also reduced overall ad visibility. Our research shows also that text length has a significant impact on ad visibility, and that gender-oriented messages influence the display of the ad based on gender.

1. Introduction

Artificial intelligence (AI), considered a disruptive technology, has brought significant transformations to technological contexts (Pietronudo et al., 2022; Dwivedi et al., 2023). Unlike traditional technologies, AI algorithms are able to process very large amounts of data which generate valuable insights for businesses (Brynjolfsson et al., 2019). They are used frequently by online platforms with the aim of improving market interactions by matching users to firms. However, there is some recent research which shows that use of AI can lead to unintended biases in the display of content (Cowgill and Tucker, 2017). Algorithmic bias occurs when "*the output of an algorithm benefit or disadvantage some individuals or groups more than others without a justified reason for these unequal impacts*" (Kordzadeh and Ghasemaghaei, 2022, page 388). In particular, algorithmic biases on digital platforms raise important and complex social issues related to discrimination and fairness in access to information (Rambachan et al., 2020). It is critical for practitioners and policy makers to better understand how these biases arise in practice and whether a categorical approach could reduce them.

More generally, the technological evolution based on AI involves some disadvantages that could offset the potential benefits of this technology, affecting many digital market business operations. An

☆ This paper was previously circulated under the title 'STEM and Teens: An Algorithmic Bias on Social Media.' We extend our special thanks to the Efrei Paris-Sud and Epitech Paris schools for their generous support. We are grateful to Ananya Sen, Maya Bacache-Beauvallet, Anna D'Annunzio, Babur De los Santos, Regis Renault, Helena Perrone, Amedeo Piolatto, and Emanuel Vespa for their valuable comments and suggestions. A prior version of this paper was presented at various conferences, including the 2023 Digital Transformation Society Conference, the 2020 Advances With Field Experiment Conference, the 46th Annual Conference of the European Association for Research in Industrial Economics (EARIE), the 2019 INFORMS Marketing Science Conference, the 4th Economics of Media Bias Workshop, the 2019 IO Clemson Workshop, the 2017 and 2018 Boston MIT Conference on Digital Experimentation (CODE), the 2018 TSE Biennial Postal Conference, the 10th ICT Paris Conference, the 67th Annual Meeting of the French Economic Association (AFSE), the 16th ZEW Conference on the Economics of Information and Communication Technologies, the 29th European Conference on Operational Research (EURO), the 2018 Research Seminar Università di Torino, and the 2nd Annual Conference of the French Association of Law and Economics (AFED). We acknowledge financial support from MSH Paris-Saclay under grants 16-OT-6, 18-EM-08 and 22-EM-06, Carnot TSN, Horizon 2020 under grant 101120657, Association Pascaline, and Institut DATAIA under grant YPPOG. All errors remain the sole responsibility of the authors.

* Corresponding author.
E-mail address: clara.jean@grenoble-em.com (C. Jean).

https://doi.org/10.1016/j.techfore.2023.123204
Received 15 July 2023; Received in revised form 26 November 2023; Accepted 30 December 2023
Available online 18 January 2024

example is the empirical evidence of algorithmic bias in the context of ad distribution, which challenges ad effectiveness and gives rise to ethical and fairness issues. The algorithm of TikTok faces allegations against its recommendation systems which reinforce the filter bubble.[1] Using data from Twitter, the article by Huszár et al. (2022) indicates that algorithms on this social media platform amplify extremely left and extremely right political groups more than moderate ones. In 2023, Global Witness conducted an online experiment and showed that Facebook (now Meta) failed to prevent discriminatory job post targeting. Global Witness ran a series of ad campaigns on Facebook across different countries, promoting real job vacancies.[2] It found that 91 % of those exposed to mechanics vacancies identified as male and 79 % of the audience targeted for pre-school teaching posts identified as female. These results highlight the gender disparity in ad targeting and raise concerns about potential algorithmic bias disadvantaging young women's job opportunities.[3] These concerns have resulted in different approaches to tackle algorithmic bias, including regulatory intervention. Recent regulation such as the Digital Service Act (DSA)[4] in Europe and the American Data Privacy and Protection Act (ADPPA)[5] in the US have drawn attention to the accountability and transparency of artificial intelligence tools.[6]

Given the opacity of AI algorithms in terms of their categorizations, we need more in depth knowledge about algorithmic decision-making bias – its identification and management. However, work on methods to evaluate algorithmic systems to detect and reduce bias is limited. We propose an advertising industry-centric method involving field tests to audit and tweak algorithms in digital markets. We rely on the theoretical algorithmic audit framework (Sandvig et al., 2014) aimed at assessing algorithmic decision-making to detect bias (Orphanou et al., 2022). We employ this framework to study whether using an advertising content-based methodological approach could mitigate algorithmic bias in online advertising. We provide evidence related to how to detect algorithmic bias, and how ad design can be leveraged to address this issue.

There is evidence suggesting the presence of algorithmic gender biases in several different sectors in the context of ads related to science, technology, engineering, and mathematics (STEM) careers. We conducted two field tests in France, in the form of ad campaigns on Facebook targeting young college students. The tests were aimed at understanding the impact of ad text adaptation and length on display patterns. In the first field test, we investigated the effect of adapting ad text on ad display. In 2017, over a two-week period, we ran 101 simultaneous ad campaigns on behalf of an engineering school which offers STEM education, targeting students in 101 French high schools. To mitigate algorithmic bias, we used two different advertising messages targeting two distinct but statistically similar groups of high schools: a women-specific text (the treatment group) and a gender-unspecific text (the control group). We run a second field test to tweak algorithm helping us to detect the sources of algorithmic bias. This second field test was conducted in 2020 over a five days period to study how ad text influences the perceived quality of the ad by the algorithm for a larger population and the relative consequences in terms of ad display. The ad campaign was for a computer science school and included an additional gender-neutral condition which allows us to disentangle the effects of text length and text adaptation.

We found that algorithmic bias can be mitigated by accurate ad text design. In particular, our results show that the women-specific ad eliminated the difference in ad display between women and men targets. This result was consistent across both our field tests related to the problems involved in algorithmic auditing. By comparing ad text length and ad text adaptation, we found that ad text length is associated with reduction in ad display and emphasizes a trade-off which advertisers need to consider. Our examination of algorithm behavior in response to different ad texts (second field test) allowed us to find that algorithmic bias persists in the case of gender-neutral text but disappears in the context of women-specific text.

This paper focuses on a specific topic of ad content distributed on social media and our main research question is related to the design of an algorithmic bias audit method. Our field tests were focused on tweaking algorithmic decision-making in an environment where ensuring algorithmic transparency and accountability is challenging. In practice, the process underlying the algorithm's decisions is not accessible. However, the present paper proposes a method for auditing AI algorithms, where we study ad delivery outcomes and test different ad content. Our analysis sheds light on the complex dynamics of ad display patterns and the potential impact of ad text adaptation and text length on algorithmic decision-making. We build on recent work which explores techniques such as prompting algorithms to test their behavior (Horton, 2023) and employing experimental approaches to measure the impact of providing the algorithm with specific information (Dujeancourt and Garz, 2023; Haaland et al., 2023) to prevent risks associated with its use.

The paper is organized as follows. Section 2 describes the theoretical background and reviews the relevant literature. Section 3 discusses the process of advertising on Facebook and the design of the first field test. Section 4 presents descriptive statistics for the first field test and Section 5 presents the empirical results. Section 6 discusses potential alternative explanations for our results. Section 7 describes the second field test. We discuss results of both experiments in Section 8. The conclusion follows.

2. Theoretical background and literature review

AI is considered a general-purpose technology (Agrawal et al., 2019b). AI algorithms are having a profound influence in numerous sectors including but not limited to manufacturing, supply chain management, healthcare, and product and service customization (OECD, 2019). AI is defined as "*the theory and development of computer systems able to perform tasks normally requiring intelligence*" (Agrawal et al., 2019c, page 3). In this sense, AI algorithms have the capacity to emulate human-like capabilities and perform tasks accordingly. The use of algorithms can have unintended ambivalent and countervailing effects. On the one hand, their use to enable a range of tasks from automation to prediction has resulted in significant performance improvements from a business perspective (Brynjolfsson et al., 2018; Cui et al., 2021). On the other hand, algorithms can generate and perpetuate discrimination against minorities and women which is questioning trust in the technology (Omrani et al., 2022). This article contributes to two streams of inter-related research highlighting the drawbacks associated with technological developments in AI. First, it contributes to work on algorithmic decision-making. Second, it adds works on the auditing of AI algorithms. The integration of AI is providing significant benefits for the advertising industry in the realm of data analysis in particular which highlights the need for a framework to evaluate and assess algorithmic fairness and efficiency.

[1] https://www.wired.co.uk/article/tiktok-filter-bubbles, last retrieved November 2023.

[2] They run the ad campaigns in UK, Netherlands, France, India, Ireland, and South Africa.

[3] https://www.globalwitness.org/en/campaigns/digital-threats/new-evidence-of-facebooks-sexist-algorithm/, last retrieved November 22, 2023.

[4] https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package, last retrieved November 22, 2023.

[5] https://www.commerce.senate.gov/services/files/9BA7EF5C-7554-4DF2-AD05-AD940E2B3E50, last retrieved November 22, 2023.

[6] For example, the DSA prohibits advertising directly targeting children or specific categories based on personal data such as ethnicity, political views, or sexual orientation and requires increased transparency for online advertising. The ADPPA is trying to regulate how organizations keep and use consumer data.

2.1. Algorithmic decision-making in digital markets

The need to manage vast amounts of real-time data generated by online behaviors has led to the use AI algorithms in firms' big data analysis activities. In the context of advertising, previous research has explored the determinants of an effective online advertising strategy (Goldfarb and Tucker, 2011), and especially the collection and use of personal data to enable ad targeting (Lau, 2020).

AI algorithms and algorithmic decision-making, can improve automated prediction (Agrawal et al., 2019a; Möhlmann et al., 2021) but can also produce unexpected outcomes. Vlačić et al. (2021) highlights the challenges posed by AI and the need for more research on the advantages and contributions of this technology for improving products and services. In the advertising industry, algorithms can enhance advertising effectiveness, reduce waste, and lower prediction costs, thereby increasing value (Agrawal et al., 2018). However, O'Neil (2016) argues that algorithms may generate bias. Algorithms can reproduce apparent discrimination or stereotypes observed in society and learned from biased individual data (Marjanovic et al., 2021). Tucker (2023) proposed the idea of the 'algorithmic exclusion,' where algorithmic predictions are impeded by bad or missing data. This raises the issue of the type of information that should be available to algorithms (Kokshagina et al., 2023). In the digital platforms cases, Sweeney (2013) observed that compared to white-identifying names, black-identifying names received more displays of an ad for criminal records services. Other studies confirm the existence of biases against women in algorithms (Datta et al., 2015; Ali et al., 2019).

However, in practice much remains to be learned about how to reduce algorithmic bias and work on mitigation of algorithmic bias and related externalities is limited. To investigate this issue, we investigate the influence of ad content on ad display (Ali et al., 2019) using a framework which addresses the following theoretical and practical question: "*How can we reduce distortion in algorithmic content exposure?*" This issue has implications for technology management and the problems encountered by managers striving to achieve specific objectives or understanding trade-offs linked to the elimination of algorithmic bias.

Our paper is related to the work of Lambrecht and Tucker (2019) who conducted a field test in 190 countries involving STEM career ads on a social network; it revealed algorithmic bias against women. The authors interpret this bias as a market failure related to use of AI by online platforms. Our work differs as we focus on how to reduce algorithmic bias in practice. Additionally, we also consider other factors that could influence algorithmic decision-making in the context of STEM higher education ads.

Work on algorithmic decision-making highlights instances where algorithms exhibit bias and this strand of research underscores the importance of identifying bias. We need to evaluate AI systems to ensure fairness. AI algorithm auditing responds to calls for transparency and accountability in algorithmic decision-making.

2.2. Auditing AI algorithms

The increase in algorithmic harm has prompted digital market stakeholders to investigate 'algorithmic audits' to assess algorithm behaviors. Algorithmic audits are defined as "*a specific subset of audit studies focused on studying algorithmic systems and content*" (Metaxa et al., 2021) which often require empirical investigation to identify potential problematic behavior (Bandy, 2021). Algorithmic audits can assess algorithm output based on their inputs without the need to access the underlying code. However, the dynamic environments in which algorithms are employed makes algorithmic auditing particularly complex.

Algorithmic auditing has been applied to a range of sectors including news (Kwak et al., 2021; Zou and Schiebinger, 2018), ridesharing (Cheng et al., 2022; Greenwood et al., 2022; Mejia and Parker, 2021), and advertising (Kingsley et al., 2020). To better understand the design and use of algorithms, Lyytinen et al. (2021) recommend diverse experiments to test how humans delegate tasks to machines to reduce undesired behaviors. Development of a thorough algorithmic audit involves numerous issues (Imana et al., 2023) including the choice of variables since the wrong choice could result in inaccurate measurements (Mullainathan and Obermeyer, 2017). Also, the lack of transparency perceived by the targets of algorithms and use of proxies for demographic attributes constitute additional hurdles (Fischer et al., 2020; Matter et al., 2022). Finally, since most research focuses on a specific case or context, generalizability is a major limitation (Vecchione et al., 2021).

The literature in this area includes two strands of work. The first strand focuses on algorithmic changes on social media platforms to audit algorithmic decision-making. Garz and Szucs (2023) use the print editions of German newspapers as a counterfactual to show that changes to Facebook's algorithm resulted in a 30 % increase in political posts compared to print media. Reuning et al. (2022) studied the effects of the 2018 Facebook algorithm change and found that in terms of reaching local communities, it benefited Republicans more than Democrats. The opacity surrounding algorithms can potentially skew message distribution by automatically targeting specific audiences (Riemer and Peter, 2021). The second strand of work examines how online platforms prioritize content in the context of the media industry. Khan et al. (2022) emphasize that algorithms play a major role in the selection and curation of information on social media. To examine search engines, Fischer et al. (2020) conducted an audit on Google News which showed that the algorithm prioritized national news over local news. Similarly, Matter et al. (2022) employed an original experiment focused on the 2020 US Elections which showed that the Google algorithm prioritized previously visited websites and not websites representative of the user's ideology. Also, evidence of systematic distortion in the information presented through Twitter's algorithmic curation was provided by Bartley et al. (2021) who conducted a sock-puppet audit on social media to examine black-box social media systems.

Currently, there is an urgent need for algorithm audits to detect discrimination and bias by automated systems on online platforms (Sandvig et al., 2014). This should extend beyond technical considerations and include the implications for broader social welfare (Rambachan et al., 2020). In the present paper, we propose some practical guidelines and provide empirical evidence showing how field tests can be used to audit algorithms and reduce algorithmic bias. Our experimental design is aimed at addressing the issues discussed above. It should be noted that our proposed external audit method is aimed at assessing AI algorithms on social media platforms in the context of reducing content distortions and fostering fairer access to information (Mikalef et al., 2022).

3. The setting

3.1. Advertising on Facebook Ads Manager

We use Facebook as it was one of the most used social networks by young people in France at the time of the field test.[7] Facebook Ads Manager allows firms to directly design and launch ad campaigns on its platform. Facebook's ad algorithm is designed to match ad content to users' interests optimizing advertisers' budgets. Less emphasis on any of these features could discourage advertisers from running ads, or reduce the amount of time spent by users on the platform. Facebook like many other platforms, allows advertisers to compete for chunks of space on its platform and hit users' eyeballs. Each impression is the result of an auction which is a modified version of a second price auction (Gordon et al., 2019). To reach the right audience, Facebook offers advertisers different audience criteria, and in particular, allows them to target

[7] https://www.weforum.org/agenda/2017/03/most-popular-social-networks-mapped/, last retrieved June 18, 2023.

specific audiences by selecting different demographic characteristics such as users' location, age, gender, language, or marital status.

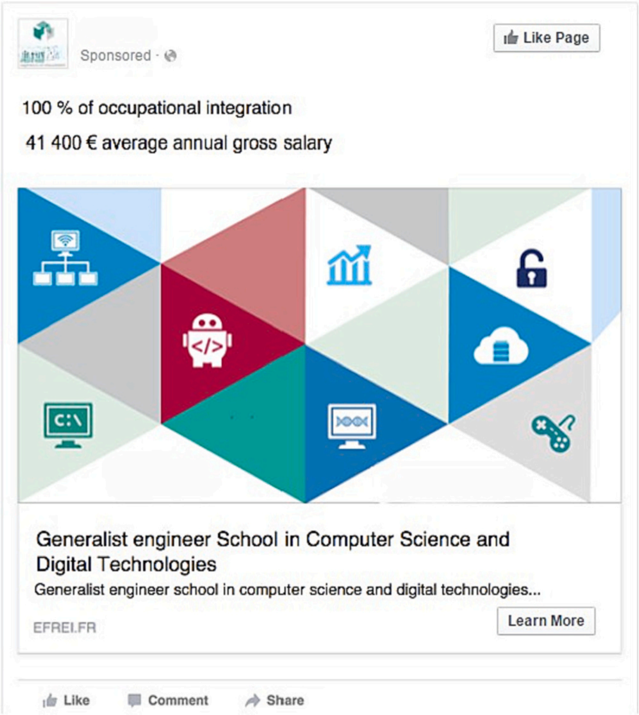
Algorithmic evaluation of ad content occurs for two reasons. First, Facebook's ad algorithm optimizes the matching of millions of ads per day to targeted users. Ads displayed to users are those with the highest total value scores calculated by the platform's ad algorithm. Total value scores are based on a range of measures: bid, estimated action rate, ad quality. According to Facebook, estimated action rate is an estimation of "how likely a person is to take actions required to get the results you're aiming at" based in part on the previous actions of the target audience. Ad quality is based on predictive analysis which estimates the likelihood that the consumer will click on a particular ad (Athey and Nekipelov, 2010). Ad quality is calculated based on the advertiser's account history and the ad content. Thus, algorithmic evaluation of the ad text is important for determining ad quality score and thus ad display. Second, the content of each ad is evaluated to verify its compliance with Facebook's general advertising policy.[8] Given the millions of ads launched on the platform each day, the ad platform uses algorithmic based control to review ad content, supervised by some human control (Cecere et al., 2020). The ads comprising our experiment have no restrictions since they propose educational content to students aged between 16 and 19 years. Educational content is not subject to particular policy restrictions. However, ads targeting individuals under 18 are subjected to several restrictions if the ads are related to gambling, birth control, financial issues, cosmetics, and social casino games.[9] This suggests that algorithmic automated control might not be the same for students aged under 18 suggesting potential different ad outcomes for youngest users.

3.2. Design of the ad campaigns: first field test

In the first field test, we investigate the effect of gender-oriented message on ad display aiming at reducing algorithmic gender bias. This field test was conducted on Facebook in 2017. We ran 101 simultaneous ad campaigns over a two-week period from March 12 to March 25, 2017. Each ad campaign was paired with a French high school with a unique Facebook profile associated to a Facebook URL, and was aimed at female and male students aged between 16 and 19 years. We target this particular age group as it corresponds to the target audience of the engineering school, which predominantly enrolls students with high school diplomas in France. The engineering school belongs to "post-bac" types of schools i.e. school to which students can candidate after obtaining their baccalaureate. The 16–19 age interval corresponds to the age range of students in high schools. On the Facebook ad platform, this age range defines two distinct age categories: 13–17 and 18–24 years. For each ad campaign, we optimized ad display by number of impressions, that is, we asked for optimization the number of times the ad was displayed to individuals. We set a daily budget of 2 euros for each ad campaign,[10] and the bid auction was fixed at the CPM because the aim was to reach the maximum number of students.

The field test is based on one treated group and one control group of targeted schools. We created two groups of high schools using a randomization procedure on administrative data. The treatment and the control groups include respectively 49 and 52 high schools. We run the advertising campaign using the usual marketing message used in the engineering school. High schools in the control group were exposed to an ad with the following gender-unspecific text: "100% of occupational integration. e41,400 average annual gross salary". High schools in the treatment group were exposed to an ad with women-specific text: "100% of occupational integration. 41,400 euros average annual gross salary for women" (Figs. 1 and 2).[11] The high school students in each of the two groups were exposed to the same images but two different texts - gender-unspecific (the control ad), and women-specific (the treatment ad). We choose the expression "for women" as the simplest message which would be easily interpretable by the algorithm as intended for women. It should be noted that the salary 41,400 euros average annual gross quoted in the ads is the actual salary earned by outgoing graduate students in this school, and there is no significant difference in the salaries paid to women and men after graduation of that engineering school.[12]

The choice to target French high schools is based on three reasons. First, the ad campaigns target students in the process of selecting their future tertiary education, specifically those under 19 years old. Second, we want to reach an audience relevant to the engineering school thus high schools. Third, we find it particularly relevant to investigate whether an algorithmic gender bias can be reduced in the context of STEM ads where younger people are targeted. The field test cannot be replicated since it is no longer possible to target students at a particular high school. A change in accessible targets was imposed on September 20, 2017 following several high-profile scandals. This prompted Facebook to reduce the target types in its Facebook Ads Manager.[13]



100 % of occupational integration

41 400 € average annual gross salary

Generalist engineer School in Computer Science and Digital Technologies

Generalist engineer school in computer science and digital technologies...

EFREI.FR

Fig. 1. Gender-unspecific ad (control).

[8] https://www.facebook.com/policies/ads/, last retrieved November 30, 2021.

[9] https://www.facebook.com/policies/ads/restricted$_$content, last retrieved November 30, 2021.

[10] Given our targeting options, this amount has been suggested by the algorithm as the minimum budget that advertisers place in the Facebook Ad Manager.

[11] Both ads were run in French.

[12] The engineering school has reported that women and men earn the same average salary after graduation. This aligns with official statistics, indicating that in 2017 in France, the average gross national salary for engineers after graduation was approximately 33,000 euros for women and 34,800 euros for men. https://start.lesechos.fr/societe/egalite-diversite/pour-les-femmes-ingenieures-les-inegalites-debutent-des-le-diplome-1177620, last retrieved February 13, 2022.

[13] See Fig. 16 for more information on the targeting on the platform and Fig. 17 for more details on the Facebook Ad Manager.

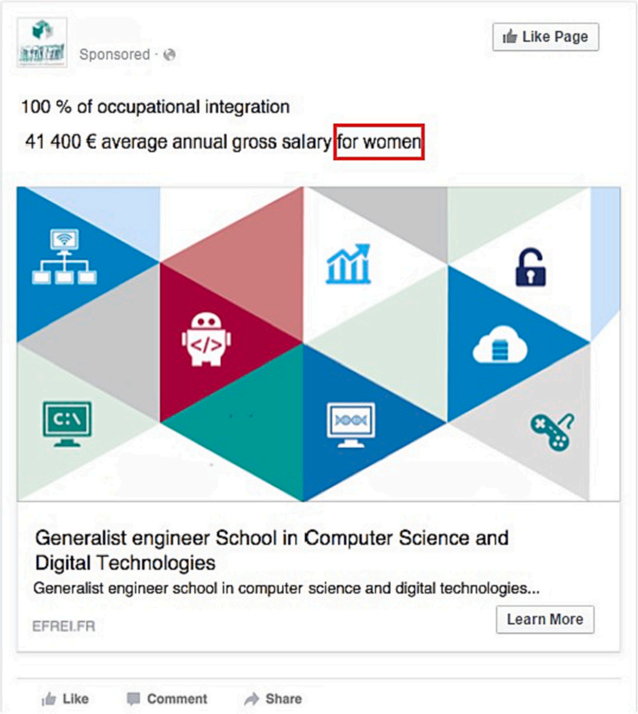

Sponsored ·
Like Page
100 % of occupational integration
41 400 € average annual gross salary for women
Generalist engineer School in Computer Science and Digital Technologies
Generalist engineer school in computer science and digital technologies...
EFREI.FR
Learn More
Like
Comment
Share

Fig. 2. Women-specific ad (treatment).

3.3. Facebook ad campaign data

Facebook Ads Manager provides detailed information on ad campaign performance broken down by age and gender. We have derived five primary measures of ad campaign performance outcomes, commonly used to assess marketing campaign by key performance indicators (KPIs) and referenced in the marketing literature (Goldfarb, 2014; Tucker, 2014; Yang et al., 2016; Erdmann and Ponzoa, 2021): *Impressions*, *Click through-rate* (*CTR*), *Reach*, *Reach cost* and *Cost per thousand impressions* (*CPM*).

Our main dependent variable is *Impressions* defined as the number of times an ad is displayed to a given group. Note here that a user receiving the ad who scrolls down and up the ad, counts as one impression. *CTR* is the ratio of the number of clicks to the number of impressions. This variable measures ad effectiveness as it measures users' behavior. *Reach* refers to the number of different users who saw the ad and *Reach cost* is the cost of reaching the targeted users. The *CPM* represents the cost of displaying an ad to a specific audience, calculated per 1000 impressions.

3.4. Age and gender of Facebook users

We check whether women and men are equally represented on the platform. To do so, we rely on the Facebook Audience Insight which provides statistics related to the percentage of male and female users for each age group (from 18 to 65 and over, divided into 6 cohorts) in a specific country. Fig. 3 depicts the statistics for Facebook users in France. We observe equal representation on the platform of women and men aged between 18 and 24. This is consistent with the gender breakdown for the overall Facebook user population which is 52 % women and 48 % men. The figures related to the 16–17 age group also suggest equal representation of women and men.[14] Therefore, as there is no observable difference in terms of gender representation on the platform, it should not impact our results.

3.5. High school administrative data

We collected high school administrative data from Etalab for the 2015–2016 school year.[15] For each high school, we have information on the size of the school with the overall enrollment, the number of women enrolled, and the proportion of women enrolled in science track. We also include information on the type and location of the high school - public, located in Paris and offering general education. Table 1 summarizes the descriptive statistics for the high schools in our sample. Given our object of study - distribution to women and men of an ad for an engineering school - we were particularly interested in female and science track enrolments. On average in each high school, enrolment is about 869 students including 464 women (about 53 %). On average, 44.2 % of the students in each high school are enrolled in the science track, and 46.8 % of these are women. We also have information on high school graduation rates and social levels. The average graduation rate for students in science is 91.16 % which is in line with the national average, and 59.0 % of students have at least one parent in a high job position. And, 79.2 % of the schools are public, 74.3 % offer general education, and 32.7 % are located in Paris.

3.6. Randomization procedure

Before launching the ad campaigns, we randomly assigned high schools to either the control or the treatment group based on administrative data. To clearly identify the causal effect of our field test, we need to have a perfectly statistically balanced group of high schools between both the treatment and control groups. Table 2 presents the pre-treatment statistics. Column (1) indicates the overall means of the variables as presented in Table 1. Column (2) reports the mean variables for the control group and column (3) reports the mean variables for the treatment group. We estimated the average baseline characteristics for high schools in the treatment and control groups. To test for balanced groups, we compute the equality of the means of each characteristic for each covariate. The last column in Table 2 presents the p-values and shows no mean difference between the two groups. All observable characteristics are balanced between the control and the treatment groups at the conventional level ($p\text{-value} < 0.05$).

4. Descriptive statistics

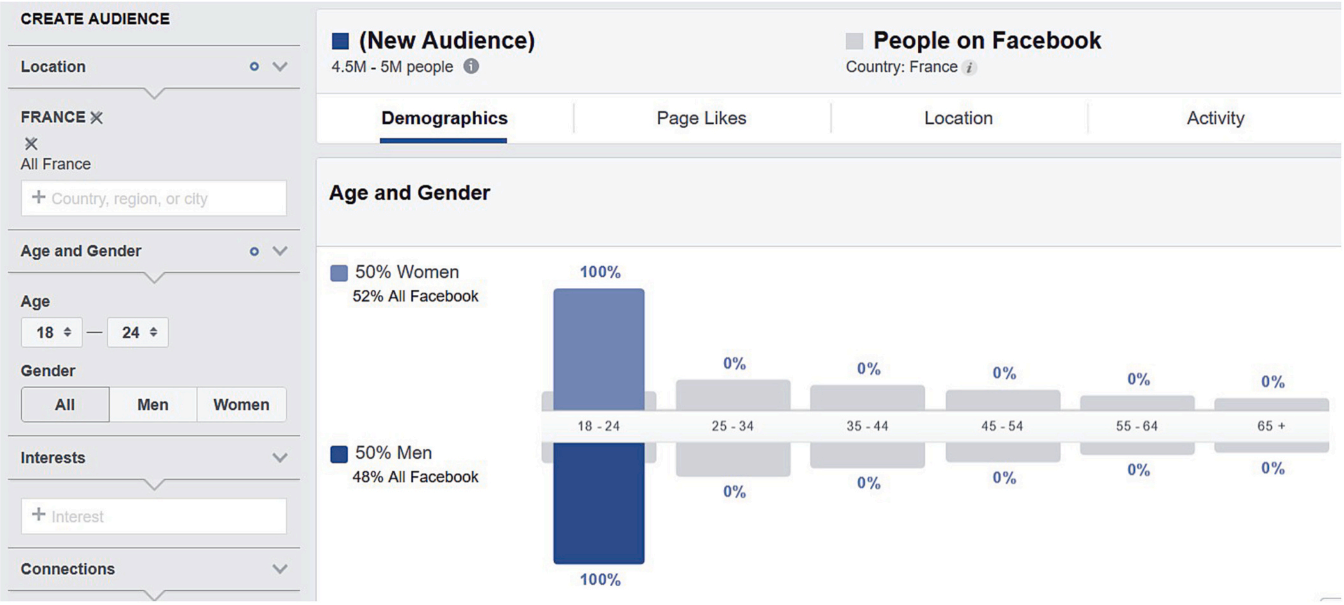
4.1. Raw data

In this section, we present the descriptive statistics of the first field test. The distributions of ad campaigns are optimized by the number of impressions, and we study the display of ads for a given school, on a given day, and for a given demographic. We only consider ad campaigns that were displayed at least once in a given day. The total sample results in 5333 rather than 5656 observations corresponding to a period of 14 days × 101 high schools × by gender (women and men) × age (16–17 and 18–19). This difference in ad display is attributable to the “choice” of the algorithm not to distribute ads to certain age and gender categories on a given day.[16] During the ad campaigns, the ad algorithm displayed a total of 1,226,929 impressions and reached 31,330 unique users with an average of 227 daily impressions and 47 people reached daily. Table 3 presents detailed descriptive statistics. Table 7 in Appendix A provides the correlation matrix.

[14] https://fr.statista.com/statistiques/574791/facebook-repartition-mondiale-par-age/ last retrieved February 11, 2022.

[15] The data are available from Etalab upon request. Etalab is a French national project which provides administrative open data for France. See https://www.etalab.gouv.fr for more information, last retrieved February 12, 2022.

[16] The estimation results do not change if we include high schools not exposed to the ad on a given day (see Appendix C).


CREATE AUDIENCE

Location

FRANCE ✕

All France

Age and Gender

Age 18 — 24

Gender: All / Men / Women

Interests

Connections

(New Audience)
4.5M - 5M people

People on Facebook
Country: France

Demographics | Page Likes | Location | Activity

Age and Gender

Fig. 3. Facebook users by gender for the 18–24 age cohort, July 2020.

4.2. *Graphical evidence*

This section provides graphical representations of the raw data. Fig. 4 displays the average number of impressions in the control and treatment groups for the field test. Overall, students enrolled in the high schools in the treatment group have fewer ads displayed to them compared to students in the control group. This difference is significant suggesting the treatment has a strong influence on the algorithmic ad display.

While the women-specific ad significantly reduces overall ad display, we observe also heterogeneous treatment effects in terms of ad display by individual demographics. Figs. 5 and 6 depict the average number of impressions in the control (gender-unspecific ad) and the treatment (women-specific ad) groups by gender for the 18–19 and 16–17 age groups.

The graphical evidence highlights two effects of the treatment. For the 18–19 age group, Fig. 5 shows that while the difference between women and men in terms of ad display is significant in the control group it is no longer significant in the treatment group - women and men saw equivalent numbers of ads. This suggests that the treatment ad is likely to reduce the ad display differences between men and women for this age group.

Fig. 6 shows the average number of impressions displayed to the 16–17 age group. Overall, we observe that the algorithm displays fewer impressions to the 16–17 age group compared to the 18–19 age group. We also observe an unexpected effect of the treatment: the treatment has an opposite effect for the 16–17 age group (compared to 18–19 age group) as it does not reduce the ad display gap between men and women. We can attribute this difference in the ad distribution pattern to Facebook's more thorough algorithmic control of ad content targeting individuals under 18 due to restrictions related to minors which aims to protect this sub-population from inappropriate content[17] (see Table 8 in Appendix B).

5. Results: does the treatment reduce algorithmic gender bias?

In this section, we adopt an econometric approach to study the effect of our treatment on impressions. As explained previously, we want to determine whether our experimental design helps reduce algorithmic gender bias captured through the number of impressions displayed to students. Our approach relies primarily on ordinary least squares (OLS) estimates and pooled data on high schools. Since the distribution is skewed we model the log of impressions for each demographic group *i* (gender and age), and each high school *j* at the time *t*. Eq. (1) captures our main econometric specification. We do not cluster the standard error at the high school levels following the work of Abadie et al. (2023) as our sample randomization occurs at the high school level. Therefore, clustering at the high school level is not necessary. The equation we estimate is as follows:

$$\begin{aligned} log(Impressions)_{ijt} = {} & \beta_0 + \beta_1 Treatment_j + \beta_2 Gender_i + \beta_3 Age1819_i \\ & + \beta_4 \left(Treatment_j \times Gender_i\right) \\ & + \lambda_t + \epsilon_{ijt}. \end{aligned} \tag{1}$$

The binary variable *Treatment* measures whether the high school belongs to the treatment group. The variable *Gender* indicates the gender of a demographic group, and takes the value 1 if it consists of women at high school *j*. *Age1819* is a variable which takes the value 1 if the age of the group is between 18 and 19 years, and zero if the age is between 16 and 17 years. λ is a vector of time fixed effects, and ϵ is the error term. Table 4 presents empirical estimates that build incrementally up to the full specification in Eq. (1).

Column (1) estimates the effect of the *Treatment*. It captures how individuals enrolled in high schools in the treatment group are exposed to the women-specific ad. The treatment significantly reduces the overall ad display in line with previous graphical evidence. The magnitude of the estimate suggests that there are 22.4 % fewer impressions in the treatment relative to the average baseline number of impressions, compared to the control group. Column (2) adds the variables *Gender* and *Age1819*. The coefficient of *Gender* is negative and significant suggesting that women receive fewer impressions (11.3 % fewer than men) for STEM education adverts, consistent with Lambrecht and Tucker's (2019) findings and underlying the persistence of algorithmic gender bias. The positive coefficient of *Age1819* suggests a higher number of ad displays to 18–19 age group. Column (3) reports the results of the main specification presented in Eq. (1). It adds an interaction term *Treatment* × *Gender* to measure the number of ads displayed to women in the treatment group. The interaction term is not

[17] See https://www.facebook.com/policies/ads/#, last retrieved November 30, 2021.

Table 1
Description of the variables used in the randomization procedure of high schools.

Variables	Description of the variables	Mean	SD	Min.	Max.
Enrollment	Number of students in a high school	868.8	(372.0)	189	2391
Women enrolled	Number of women in a high school	464.07	(217.17)	54	1164
Students in science[a]	% of students in the science track	0.442	(0.148)	0.08	0.84
Female in science[a]	% of women in the science track	0.468	(0.098)	0.18	0.75
Grad rate in science[a]	Graduation rate in the science track	91.16	(10.46)	46	100
High-incomes ratio	% of students with a parent in a high job position	0.590	(0.212)	0.122	0.980
Public school	=1 if the school is public	0.792	–	0	1
General school	=1 if the school provides general education	0.743	–	0	1
Paris school	=1 if the school is located in Paris	0.327	–	0	1
Technical school	=1 if the school provides technical education	0.030	–	0	1
Multipurpose school	=1 if the school provides general and technical ed.	0.228	–	0	1

Notes: This table reports the mean estimates with standard deviations in parentheses. The number of observations is equal to 101.

[a] The variables *Students in science*, *Female in science*, *Grad rate in science* count 94 observations because 7 high schools (2 general high schools, 2 multipurpose high schools, and 3 technical schools) do not offer a science track.

Table 3
Summary statistics for the first field test: Social network data.

Variables	Mean	SD	Min.	Max.	N
Impressions	226.899	238.093	1	1843	5333
CTR	0.002	0.007	0	0.167	5333
Reach	47.191	37.335	1	175	5333
Reach cost	10.738	6.645	0	130	5333
CPM	3.156	2.261	0	12.5	5333
Treatment	0.465	0.499	0	1	5333
Gender	0.503	0.5	0	1	5333
Age1819	0.508	0.5	0	1	5333

Notes: In this field test, one observation represents one high school, one day, one gender, and one age category. On average, the CPM surpasses our maximum bid because it represents the estimated cost for a thousand impressions, while we typically achieve on average 226.9 impressions.

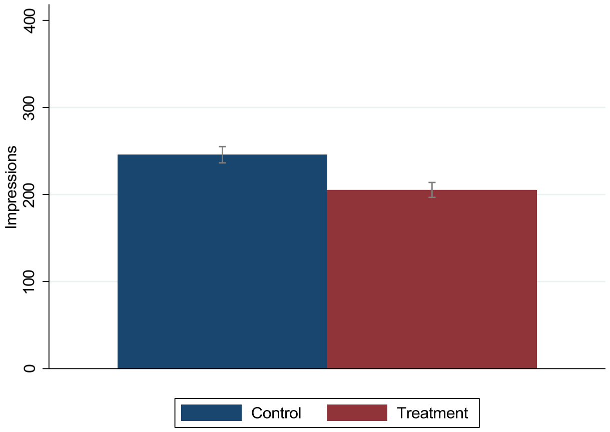


Fig. 4. Overall effect of the treatment on ad display.
Note: The confidence intervals are computed at p-value < 0.05.

Table 2
Randomization assignment to the control and treatment groups.

Variable	Overall (1)		Control (2)		Treatment (3)		p-Values
	Mean	SD	Mean	SD	Mean	SD	
Enrollment	868.8	(372.0)	897.827	(403.081)	838.000	(337.350)	0.422
Women enrolled	464.0	(217.17)	470.16	(239.24)	457.62	(193.29)	0.773
Students in science[a]	0.442	(0.148)	0.447	(0.150)	0.436	(0.148)	0.716
Female in science[a]	0.468	(0.098)	0.456	(0.093)	0.483	(0.181)	0.167
Grad rate in science[a]	91.16	(10.46)	90.902	(10.251)	91.465	(10.815)	0.796
High-incomes ratio	0.590	(0.212)	0.590	(0.208)	0.589	(0.031)	0.990
Public school	0.792	–	0.827	(0.382)	0.755	(0.434)	0.379
General school	0.743	–	0.692	(0.466)	0.796	(0.407)	0.238
Paris school	0.327	–	0.327	(0.474)	0.327	(0.474)	0.997
Technical school	0.030	–	0.019	(0.139)	0.041	(0.200)	0.528
Multipurpose school	0.228	–	0.288	(0.457)	0.163	(0.373)	0.136

Notes: This table reports the mean estimates with standard deviations in parentheses. The number of observations is equal to 101.

[a] The variables *Students in science*, *Female in science*, *Grad rate in science* count 94 observations because 7 high schools (2 general high schools, 2 multipurpose high schools, and 3 technical schools) do not offer a science track.

statistically significant, indicating that impressions did not differ in the treatment by gender. This result suggests that the display gap between men and women in the treatment group has been eliminated. Using a women-specific text removes the difference in ad display between men and women compared to the control group in which algorithmic gender bias remains.[18]

[18] Table 9 in Appendix C provides similar results with a balanced panel.

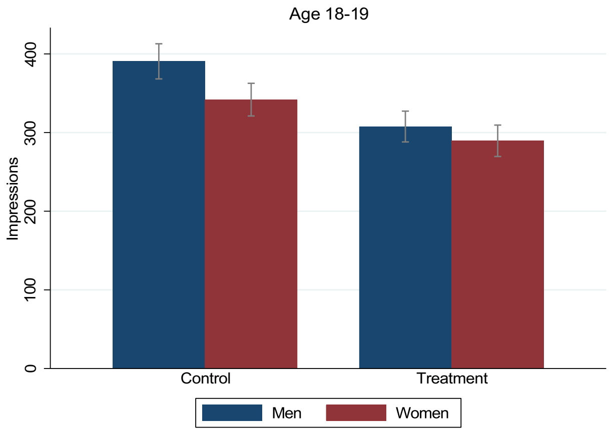


Fig. 5. Impressions for 18–19 age group.
Note: The confidence intervals are computed at p-value < 0.05.

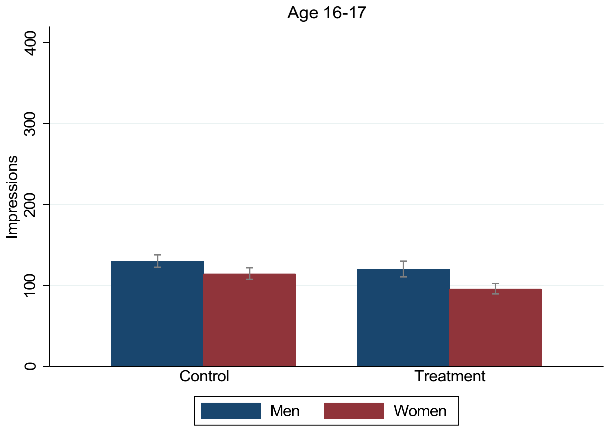


Fig. 6. Impressions for 16–17 age group.
Note: The confidence intervals are computed at p-value < 0.05.

Table 4
Effect of the treatment on impressions.

	log(Impressions)		
	(1)	(2)	(3)
Treatment	−0.224***	−0.229***	−0.287***
	(0.040)	(0.037)	(0.054)
Gender		−0.113***	−0.167***
		(0.037)	(0.050)
Age 18–19		1.112***	1.112***
		(0.037)	(0.037)
Treatment × Gender			0.117
			(0.075)
Constant	5.089***	4.591***	4.618***
	(0.080)	(0.079)	(0.081)
Time fixed effects	Yes	Yes	Yes
Mean	226.89	226.89	226.89
R-squared	0.015	0.158	0.158
Observations	5333	5333	5333

Notes: OLS estimations. The dependent variable is as indicated in the table. The number of observations is equal to 5333 rather than 5656 since we only consider observations for users exposed to ads. The estimation results do not change if we include high schools not exposed to the ad on a given day (see Appendix C). Robust standard errors are reported in parentheses.
*** Significance at 1 % level.

6. Investigating the mechanisms that can justify algorithmic decision-making

In addition to our primary findings, we delve into potential alternative mechanisms that may underline our treatment effect. First, we explore the possibility that our results are influenced by an algorithmic learning effect. Second, we examine whether the cost of impressions plays a role in the variation of ad display. Third, we investigate if other ad performance variables, such as *CTR* and *Reach*, could potentially explain the differences in ad display between the treatment and control groups.

6.1. Is there any evolution in ad display distribution?

We can expect that algorithms learn continuously and adapt their parameters accordingly which can induce changes in ad algorithm distribution patterns over time. In particular, the article by Cowgill and Tucker (2019) highlights that algorithmic feedback loops are likely to reinforce bias. Therefore, we investigate if there are any changes during the ad campaigns. Fig. 7 depicts daily ad distribution by treatment and control groups across genders. The y-axis shows the log of impressions. The ad display pattern across time is similar between the control and the treatment groups, suggesting that the ad distribution patterns are stable, and algorithmic feedback loops do not allow variation over time.

We then take a closer look at a learning effect by age category (see Figs. 14 and 15 in Appendix D). We find no evidence of learning effect over time, we observe a similar ad display pattern in both the control and the treatment groups. As shown previously, the effect of the treatment is salient for the 18–19 age group where the treatment ad is likely to eliminate the ad display difference between women and men.

6.2. What is the effect of the cost of impressions on ad display?

To explain both the decrease in ad display to the 16–17 age group and the reinforcement of algorithmic gender bias in the treatment group for this age category, we examine whether these results can be explained by the CPM. This cost reflects spillovers between advertisers (as highlighted by Lambrecht and Tucker (2019)). Fig. 8 depicts the *CPM* by age, gender, and treatment. We observe that the 16–17 age group is significantly more expensive to target compared to the 18–19 age group. An explanation may be that the platform's self-regulation which results in an additional ad content filtering for the 16–17 age group, may lead to further restriction of the distribution of ad content. This highlights a likely relationship of competition between advertisers and gender bias in the use of the AI algorithm. Another source of explanation may also be

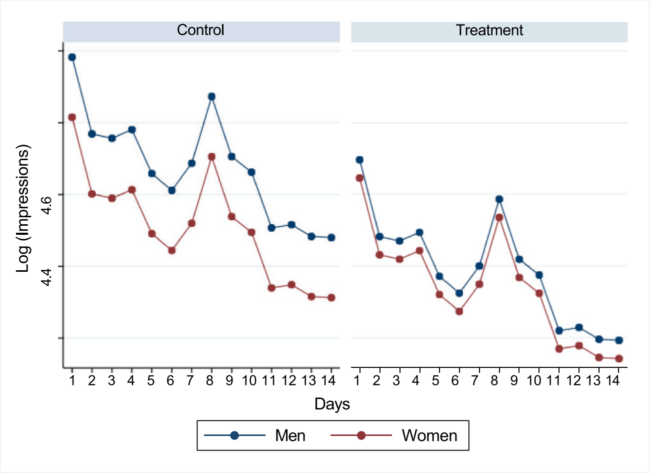


Fig. 7. Evolution of the log of impressions over time.

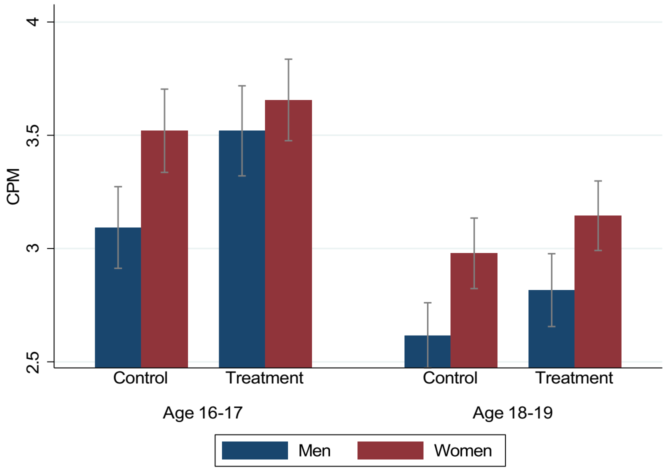


Fig. 8. CPM by age, gender, and treatment.
Note: The confidence intervals are computed at p-value < 0.05.

that this younger audience is more protected by privacy regulation (Marthews and Tucker, 2019; Cecere et al., 2022) and is therefore most expensive to reach.

We observe no statistical difference in terms of CPM between women and men in the treatment group for both age groups, but women in the control group remain a prized demographic as documented in Lambrecht and Tucker (2019). The cost of impressions explains the overall lower ad display to the 16–17 age group.

6.3. Do other performance variables explain the result?

We check also whether our treatment has an effect on other ad performance variables. Table 5 presents estimates of the main equation using alternative outcome variables: *CTR*, log(*CPM*), log(*Reach*), and *Reach cost*. We check whether the algorithmic ad distribution can be associated to individual interest in the ad (Baeza-Yates, 2018). In particular, it might be that men are less likely than women to click on the treatment ad simply because the content is likely to target women rather than men; in this case the algorithm directs more ads to women. This means that alongside making the ad more attractive to women, it becomes less attractive to men. To study this hypothesis, we look at the CTR. Column (1) presents the estimates. The interaction term *Treatment* × *Gender* is not significant, thus algorithmic ad display does not seem to be affected by the clicks of women in the treatment group. The estimates in column (2) use the log of the CPM as the dependent variable. While women are significantly more expensive to advertise than men, the insignificant interaction term *Treatment* × *Gender* shows that it is not at work in the treatment, in line with previous graphical evidences. We next investigate whether the treatment affects the number of individuals reached by the ad algorithm. The estimates are presented in column (3). While overall an equal number of men and women were reached despite differences in terms of impressions, the interaction term *Treatment* × *Gender* is positive and significant suggesting that female students enrolled in high schools in the treatment group are more likely to be reached with no differences in terms of reach cost (see column 4).

Table 5
Effect of the treatment on ad performances variables.

	CTR	log(CPM)	log(Reach)	Reach cost
	(1)	(2)	(3)	(4)
Treatment	−0.000261	0.081034***	−0.289398***	0.001589***
	(0.000288)	(0.018783)	(0.044532)	(0.000301)
Gender	−0.000292	0.098386***	−0.032694	−0.000120
	(0.000272)	(0.018220)	(0.041050)	(0.000197)
Treatment × Gender	0.000339	−0.022807	0.137569**	−0.000564
	(0.000366)	(0.026092)	(0.062036)	(0.000369)
Age 18–19	−0.000201	−0.117312***	0.721876***	0.001860***
	(0.000185)	(0.013072)	(0.030860)	(0.000180)
Constant	0.002151***	1.125467***	3.191373***	0.009599***
	(0.000379)	(0.028979)	(0.063883)	(0.000460)
Time fixed effects	Yes	Yes	Yes	Yes
R-squared	0.002	0.054	0.103	0.034
Observations	5333	5333	5333	5333

Notes: OLS estimations. The dependent variable in this table column (1) is the CTR, in column (2) it is the log(CPM), in column (3) it is the log(Reach) and in column (4) it is the reach cost. Columns (1) to (4) include time fixed effects. The number of observations is equal to 5333 since we only consider observations for users exposed to ads. Robust standard errors are reported in parentheses.

*** Significance at 1 % level.

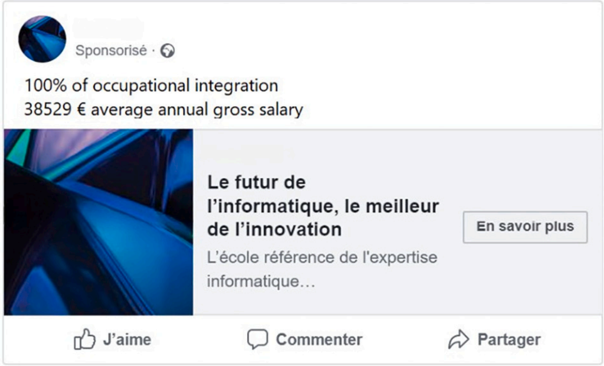
** Significance at 5 % level.

These results suggest first that the treatment eliminates the ad display gap between men and women despite overall lower ad display and lower numbers of students reached by the treatment ad. The treatment significantly increases the number of women reached. Finally, we observe also that the treatment ad reached more students in the 18–19 age groups which is in line with previous evidence.

7. Second field test: is there an effect of text length?

The primary objectives of the second field test are to assess the generalization of previous findings and to separate the impact of text length from text adaptation on ad display. Based on our previous results, while the women-specific ad (treatment) eliminates the gender ad display gap, it also significantly lowers ad display for both women and men. This could be due to a potential distortion effect related to use of the word “women” in the ad distribution. The length of the ad text and the word “women” could have affected the quality of the ad perceived by the ad algorithm. In particular, the article of Ali et al. (2019) highlights the importance of ad content in ad display. We are interested here in the effect of the words added to the ad distribution.

We used the same main message as in the first field test as shown in Figs. 9 and 11 but added a gender-neutral treatment condition (see Fig. 10). The women-specific and gender-neutral ads included respectively the words “women” (in French “femmes”) and “students” (in French “élèves”) both of which in French have the same number of characters. This set up allowed us to assess whether the meaning of the word or the length of the text was driving our results.


[redacted]
Sponsorisé ·

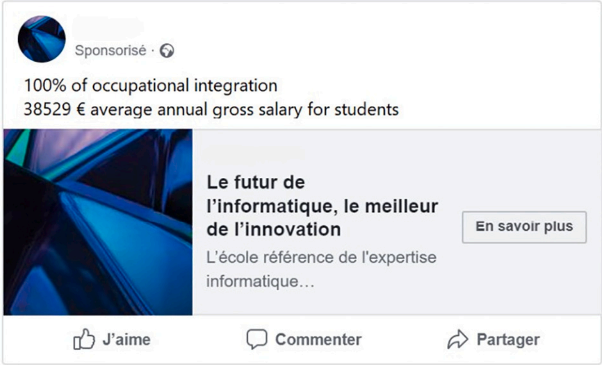
100% of occupational integration
38529 € average annual gross salary

Le futur de l'informatique, le meilleur de l'innovation

L'école référence de l'expertise informatique…

Fig. 9. Gender-unspecific ad (control).

The actual length of the ad text and the meaning of the word(s) might affect the pattern of ad distribution. Including an additional word could reduce the quality of the ad perceived by the Facebook's ad algorithm,

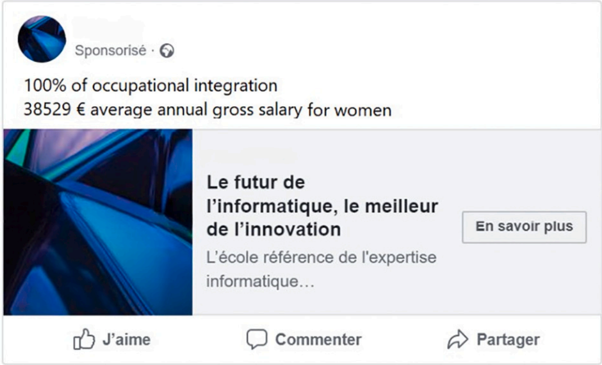

Sponsorisé ·

100% of occupational integration
38529 € average annual gross salary for students

Le futur de l'informatique, le meilleur de l'innovation

L'école référence de l'expertise informatique…

Fig. 10. Gender-neutral ad.


Sponsorisé ·

100% of occupational integration
38529 € average annual gross salary for women

Le futur de l'informatique, le meilleur de l'innovation

L'école référence de l'expertise informatique…

Fig. 11. Women-specific ad.

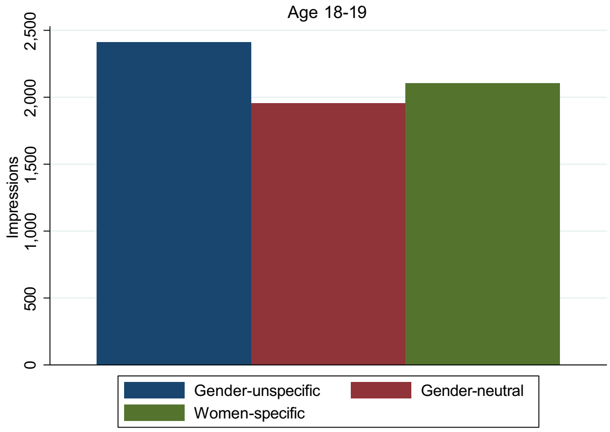


Fig. 12. Effect of adding a word on ad display.

and therefore could reduce display of the ad. According to Facebook, an ad which includes <20 % text is more likely to achieve better ad performance.[19]

We replicated the first field test with ad campaigns targeting the whole of France. Since September 2017 Facebook's ads manager no longer allows targeting of students from particular high schools. However, this does not affect our research objective, as our primary focus in this field test is on the role of text in the distribution of ads. In this field test, users were not assigned to one of the three ad campaigns. Any eligible user located in France and meeting our target criteria was open to receiving the display of the ad from any of the ad campaigns.

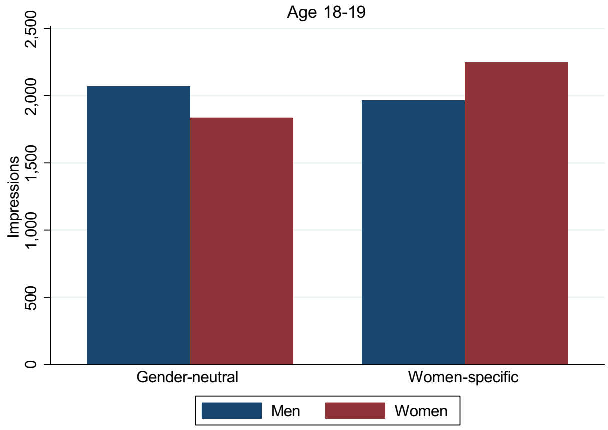


Fig. 13. Meaning of the word and gender ad display.

We conducted similar ad campaigns for another computer science school offering post-secondary education. We ran three ad campaigns simultaneously over a period of 5 days from January 18 to January 22, 2020. We target high schoolers aged between 16 and 19 years with no distinction by gender located in France. We allocated a daily budget of 5 euros per campaign following the platform recommendation given our ad setting and the bid auction was fixed at the CPM. We paid an average of 4.98 euros per day and per ad campaign.

7.1. *Descriptive statistics of the second field tests*

The total sample includes 60 observations corresponding to a period of 5 days × 3 ad campaigns × 2 genders × 2 age categories. During the ad campaigns, the algorithm displayed a total of 128,628 impressions, and reached 87,262 unique users (with an average of 2144 daily impressions and 1454 people reached daily). Table 6 presents detailed descriptive statistics. The average daily impressions is higher compared to the first field test because we targeted the whole France.

7.2. *Analysis of the main results*

To analyze the outcomes of this field test, our focus is on the 18–19 age group. This approach is adopted considering that the treatment ads might exhibit ambiguous effects in terms of ad display to minors, as previously demonstrated in the first field test. To assess whether there are difference in ad display, we estimate the statistical difference among the three conditions. Given our small sample size, we use the Wilcoxon rank-sum test. This test for matched pairs is a standard non-parametric test with a null hypothesis stating no treatment effect or that the two samples are derived from identical populations (List and Price, 2009).

Fig. 12 depicts the effect of adding a word on the ad distribution for

Table 6
Summary statistics of the second field test.

Variables	Mean	SD	Min.	Max.	N
Impressions	2143.8	286.232	1569	2832	60
CTR	0.142	0.098	0	0.418	48
CPM	1.245	0.165	0.89	1.66	60
Treatment	1	0.823	0	2	60
Gender	0.5	0.504	0	1	60
Age1819	0.5	0.504	0	1	60

Notes: One observation represents one treatment, on a day, per gender and age category.

[19] Most brands use a maximum of 14 words on Facebook ad copies. https://sproutsocial. com/insights/social-media-character-counter/, last retrieved February 26, 2020. https://www.facebook.com/business/help/980593475366490?id=1240182842783684, last retrieved February 26, 2020.

the 18–19 age group. We observe a significant drop in the number of impressions displayed to individuals in the gender-neutral treatment group compared to the gender-unspecific group (p-value of the Wilcoxon rank-sum test = 0.0019). Fig. 12 shows also that there are less impressions in the women-specific treatment group compared to the gender-unspecific group (p-value of the Wilcoxon rank-sum test = 0.0284). Therefore, adding a word significantly decreases the display of ads. This result implies that the algorithm may perceive longer ads as lower quality ads.

To explore gender-based differences in ad distribution, we focus on examining the impact of text adaptation on ad display within each gender category. Fig. 13 depicts the effect of the gender orientation of ad text on the ad distribution by gender for the 18–19 age group. On the one hand, using a gender-neutral ad perpetuates the algorithmic gender bias where the ad is significantly less displayed to women compared to men (p-value of the Wilcoxon rank-sum test = 0.0472). On the other hand, the ad with a women-specific text is more displayed to women reducing the gender imbalance in terms of ad display. The difference in ad display is no longer significant at 5 % (p-values of the Wilcoxon rank-sum test = 0.0758). Consistent with the findings from the initial field test, we observe a statistically equal number of impressions displayed to women and men within the women-specific (treatment) group, thereby mitigating algorithmic gender bias.

The results of this second field test emphasize two main findings. First, text length plays an important role in ad display since adding a word is likely to significantly reduce ad display. A potential underlying mechanism here is that in the case of longer texts, the quality of the ad determined by the algorithm is lower. Second, we retrieve the result of the first field test where a women-specific text eliminated the algorithmic gender bias while the use of a gender-neutral text perpetuates this bias.

8. Discussion

The rise in European regulation, aimed at enhancing the transparency of algorithmic decision-making, highlights the imperative to accurately identify and address algorithmic bias (Xenidis et al., 2021) Our empirical investigation of the advertising industry case used field tests to tweak algorithms in digital markets and explore the potential of a content-based approach to alleviate biases in online advertising. We studied using auditing approach (Sandvig et al., 2014) the case of a specific Internet platform i.e. Facebook and to study the impact of ad creative on ad display in the context of stereotyped ads using a content-based approach.

The results of the first field test indicate that in our experimental framework, careful design and accurate ad text can mitigate algorithmic bias. However, we found that a trade-off occurs; if the length of the ad decreased ad display, adapting the text can alleviate algorithmic bias. In particular our second field test showed that bias persisted with gender-neutral text but disappeared with women-specific text.

Our work contributes to research on algorithmic bias (Lambrecht and Tucker, 2019; Sweeney, 2013; Angwin et al., 2016; Datta et al., 2015) by proposing a strategy to reduce particularly gender-related bias. While the literature tends to emphasize the problems related to generalizing audits (Vecchione et al., 2021), our study shows that bias is consistent over time and that the impact of using ad text to mitigate algorithmic bias is also significant. This suggests that making adjustments to ad text—essentially tweaking the algorithm—could reduce algorithmic bias in online advertising.

One of the novelties of our research is that the field tests were conducted at different points in time and shows the persistence of bias in algorithmic decision-making over time. Although there might be some concern related to the fact that our ads were related to different STEM-related schools, based on field tests we show that bias in algorithmic decision-making and its mitigation implies the absence of an advertiser effect and provides external validation for our proposed methodology.

9. Conclusion

Advances in AI algorithms used to access large amounts of data on individuals can help simplify complex issues. On social network platforms, AI tools provide businesses with powerful ways to reach (and influence) segments of the population through fine targeting. However, it has been shown that algorithms can have unintended and unexpected outcomes (Cowgill and Tucker, 2017; Tucker, 2019). Thus, we need to employ advanced tools for auditing algorithms to ensure fairer and more transparent decisions (Fischer et al., 2020).

Given the evidence from the literature on algorithmic gender biases (Sweeney, 2013; Datta et al., 2018; Lambrecht and Tucker, 2019; Cowgill and Tucker, 2017), in this article, we aim to understand how ad design can reduce distortion in algorithmic content exposure. We conducted two field tests that focus on a well-known social network - Facebook - where we ran ad campaigns for engineering and computer science schools aimed at high school students. The first field test aims to uncover whether designing an appropriate ad can help mitigate algorithmic gender bias, while the second field test helps to generalize our results and provide more explanations related to how algorithms can be audited.

Through our field tests, we present compelling evidence that algorithmic gender bias can be effectively mitigated by employing tailored ad text. Specifically, in our first field test, the treatment ad eliminates discrepancies in ad display between women and men within the 18–19 age group. This result is consistent with our findings from the second field test, where we compare a gender-neutral condition to text specifically designed for women. Interestingly, the gender-neutral condition perpetuates algorithmic gender bias, while the women-specific text (treatment) successfully reduces it. Additionally, our findings reveal that the inclusion of an additional word reduces the perceived ad quality, which advertisers must bear in mind when designing their ads. By examining the algorithm's behavior in response to different ad text, the second field test not only sheds light on its dynamics but also allows for the generalization of our findings.

Finally, we identified an unexpected effect of our treatment on ads displayed to minors. The gender imbalance in terms of ad display is reinforced in the treatment group, and we attribute this result to the platform's self-regulation of ads concerning content aimed at minors. We believe that this result can be explained by a premium associated with reaching this particular audience with a specific ad message.

Our results have important implications for scholars and practitioners. Platforms and policymakers need to understand how ad algorithms work and the degree of their sensitivity. Transparency alone is not enough to ensure algorithm accountability (Marjanovic et al., 2021). Through our field tests, we manipulated the text contained in the ad displayed by the algorithm and demonstrated the different effects this could have on relative ad display. We add to the existing literature on algorithmic auditing to provide valuable insights for practitioners and policymakers, offering methods to discern and address algorithmic bias. In particular, our article corroborating previous findings in the literature related to skewed ad delivery in social media in different contexts (Ali et al., 2019). This ad design can help managers better understand the trade-offs associated with ad design to align with defined objectives. Managers must carefully weigh the trade-off between employing gender-oriented messaging and its impact on ad display, as even the inclusion of an extra word can significantly influence outcomes.

This study also highlights the value of field tests as a valuable tool for algorithm auditing. Our results provide a methodological approach to auditing online algorithms, with field tests serving as a means to achieve greater transparency in algorithmic decision-making online. Therefore, understanding the reasons behind the results should be of interest to platforms, while it should also be informative for policymakers who need to be aware of how those algorithms can be audited and the trade-offs involved in designing regulations to limit algorithmic biases.

Our work has some limitations. First, we were not able to capture the effect of competition from other advertisers during the auction process. Second, while our experiments involve real advertisers operating under authentic conditions, the field tests we conduct are rooted in highly context-specific settings. Third, algorithm auditing is challenged by the definition of algorithmic fairness standards, which is important to address to deal with discrimination and ethical issues. Nevertheless, we believe that documenting the effect of ad text on algorithmic gender bias in the context of advertising is a valuable contribution to the literature on automated algorithmic decision-making. Experts, policymakers, and citizens need to be able to verify that algorithms are not biased and identify sources of bias using independent methods.

CRediT authorship contribution statement

G. Cecere: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **C. Jean:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. **F. Le Guel:** Funding acquisition, Resources, Writing – original draft. **M. Manant:** Conceptualization, Funding acquisition, Resources, Writing – original draft.

Data availability

Data will be made available on request.

Appendix A. Matrix of correlations

Table 7
Cross-correlation.

Variables	Impressions	Clicks	CTR	Reach	Reach cost	Age1819	Gender	Treatment
Impressions	1.000							
Clicks	0.365	1.000						
	(0.000)							
CTR	−0.061	0.417	1.000					
	(0.000)	(0.000)						
Reach	0.879	0.377	−0.046	1.000				
	(0.000)	(0.000)	(0.001)					
Reach cost	−0.298	−0.146	0.004	−0.462	1.000			
	(0.000)	(0.000)	(0.750)	(0.000)				
Age1819	0.460	0.189	−0.015	0.398	0.140	1.000		
	(0.000)	(0.000)	(0.274)	(0.000)	(0.000)			
Gender	−0.060	−0.012	−0.010	−0.002	−0.030	−0.006	1.000	
	(0.000)	(0.365)	(0.470)	(0.890)	(0.031)	(0.646)		
Treatment	−0.085	−0.045	−0.007	−0.089	0.099	0.005	0.002	1.000
	(0.000)	(0.001)	(0.611)	(0.000)	(0.000)	(0.733)	(0.875)	

Appendix B. Difference in ad display for the subsample 16–17 age group for the field test 1

Table 8 presents detailed descriptive statistics on the difference in ad display for the subsample 16–17 age group for the field test 1. The difference in ad display between men and women increases in the treatment group with women receiving significantly fewer ad displays compared to men. This suggests that the platform's efforts to protect minors from potentially inappropriate content via stricter regulation may induce unintentional reinforcement of the algorithmic gender bias.

Table 8
Difference in impressions by gender and by treatment or control group for students in the 16–17 age group.

Impressions	Men		Women		Diff.	N	p-Values
	Mean	SD	Mean	SD			
Control group	130.208	102.770	114.740	96.654	15.47	1409	0.004
Treatment group	120.375	121.197	96.086	81.556	24.29	1215	0.000

Appendix C. Main result with a balanced panel

Table 9
Treatment effect on impressions with a balanced panel.

	Log(Impressions)		
	(1)	(2)	(3)
Treatment	−0.175***	−0.176***	−0.249***
	(0.047)	(0.044)	(0.064)
Gender		−0.058	−0.126**
		(0.044)	(0.061)
Age 18–19		1.155***	1.155***
		(0.044)	(0.044)

(continued on next page)

Table 9 (*continued*)

	Log(Impressions)		
	(1)	(2)	(3)
Treatment × Gender			0.146*
			(0.088)
Constant	4.736***	4.188***	4.222***
	(0.097)	(0.098)	(0.101)
Time fixed effects	Yes	Yes	Yes
R-squared	0.007	0.115	0.115
Observations	5656	5656	5656

Notes: OLS estimates. The dependent variable is the log of impressions. The number of observations is equal to 5656 as we fill missing values by zero. Robust standard errors are reported in parentheses.

*** Significance at 1 % level.

** Significance at 5 % level.

* Significance at 10 % level.

Appendix D. First field test: learning effects by age groups

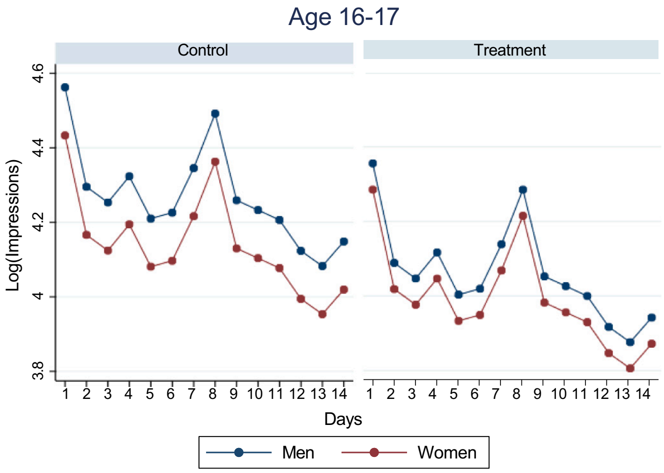


Fig. 14. Evolution of the log of impressions over time for the 16–17 age group.

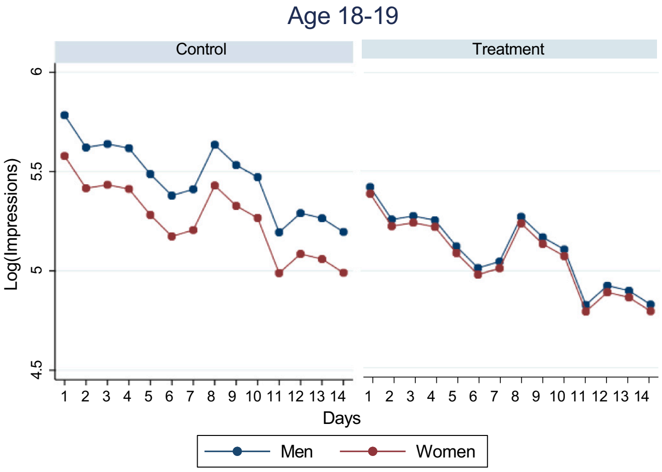
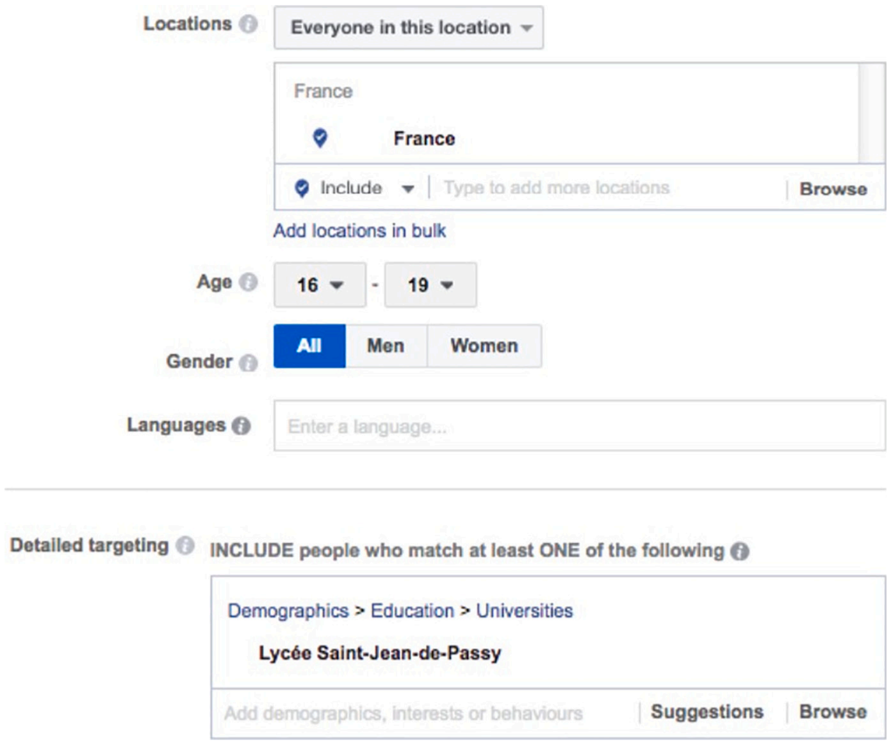


Fig. 15. Evolution of the log of impressions over time for the 18–19 age group.

Appendix E. Ad settings



Locations: Everyone in this location

France

France

Include ▾ | Type to add more locations | Browse

Add locations in bulk

Age: 16 - 19

Gender: All | Men | Women

Languages: Enter a language...

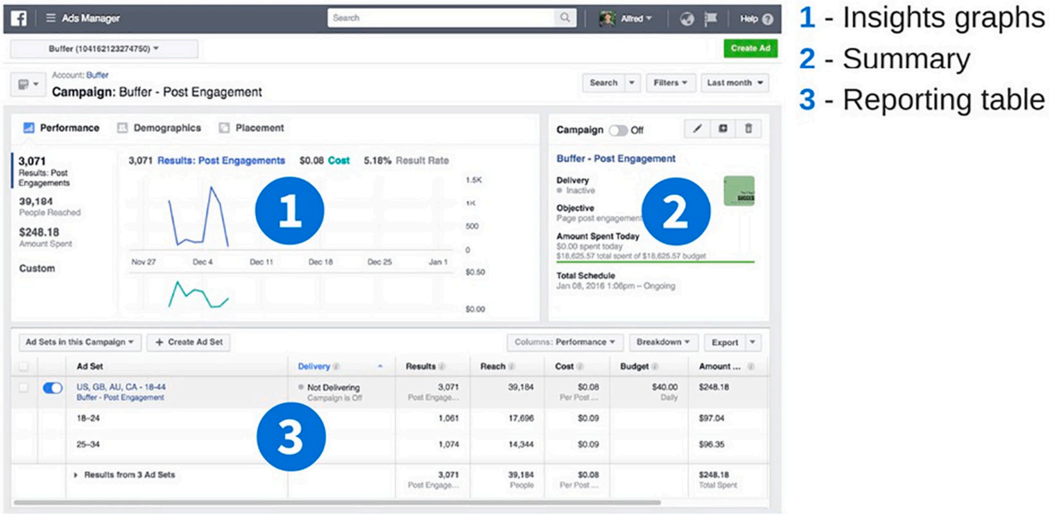
Detailed targeting: INCLUDE people who match at least ONE of the following

Demographics > Education > Universities

Lycée Saint-Jean-de-Passy

Add demographics, interests or behaviours | Suggestions | Browse

Fig. 16. First field test: Facebook Ads Manager.



Ads Manager | Search | Alfred ▾ | Help

Buffer (104162123274750) ▾ | Create Ad

Account: Buffer

Campaign: Buffer - Post Engagement

Search ▾ | Filters ▾ | Last month ▾

Performance | Demographics | Placement

3,071 Results: Post Engagements

39,184 People Reached

$248.18 Amount Spent

Custom

3,071 Results: Post Engagements $0.08 Cost 5.18% Result Rate

Campaign Off

Buffer - Post Engagement

Delivery: Inactive

Objective: Page post engagement

Amount Spent Today: $0.00 spent today; $18,625.57 total spent of $18,625.57 budget

Total Schedule: Jan 08, 2016 1:06pm – Ongoing

Ad Sets in this Campaign ▾ | + Create Ad Set | Columns: Performance ▾ | Breakdown ▾ | Export ▾

Ad Set	Delivery	Results	Reach	Cost	Budget	Amount ...
US, GB, AU, CA - 18-44 Buffer - Post Engagement	Not Delivering Campaign is Off	3,071 Post Engage...	39,184	$0.08 Per Post ...	$40.00 Daily	$248.18
18–24		1,061	17,696	$0.09		$97.04
25–34		1,074	14,344	$0.09		$96.35
▸ Results from 3 Ad Sets		3,071 Post Engage...	39,184 People	$0.08 Per Post ...		$248.18 Total Spent

1 - Insights graphs

2 - Summary

3 - Reporting table

Fig. 17. Facebook ads manager.
Source: Buffer (2018).

References

Abadie, A., Athey, S., Imbens, G.W., Wooldridge, J.M., 2023. When should you adjust standard errors for clustering? Q. J. Econ. 138 (1), 1–35.

Agrawal, A., Gans, J., Goldfarb, A., 2018. Prediction Machines: The Simple Economics of Artificial Intelligence. Harvard Business Press.

Agrawal, A., Gans, J., Goldfarb, A., 2019a. Artificial intelligence: the ambiguous labor market impact of automating prediction. J. Econ. Perspect. 33 (2), 31–50.

Agrawal, A., Gans, J., Goldfarb, A., 2019b. Economic policy for artificial intelligence. Innov. Policy Econ. 19 (1), 139–159.

Agrawal, A., Gans, J., Goldfarb, A., 2019c. The Economics of Artificial Intelligence: An Agenda. University of Chicago Press.

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., Rieke, A., 2019. Discrimination through optimization: how Facebook's ad delivery can lead to biased outcomes. In: Lampinen, A., Gergle, D., Shamma, D.A. (Eds.), Proceedings of the ACM on Human-Computer Interaction, PACMHCI, vol. 3(11). Association for Computing Machinery, New York, NY, USA, pp. 1–30.

Angwin, J., Larson, J., Mattu, S., Kirchner, L., 2016. Machine bias: there's software used across the country to predict the future crimininals and its biased against Blacks. Retrievable at, ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Athey, S., Nekipelov, D., 2010. A structural model of sponsored search advertising auctions. In: Sixth Ad Auctions Workshop, vol. 15, pp. 1–65 (September).

Baeza-Yates, R., 2018. Bias on the web. Commun. ACM 61 (6), 54–61.

Bandy, J., 2021. Problematic machine behavior: a systematic literature review of algorithm audits. Proc. ACM Human-Comput. Interact. 5 (CSCW1), 34.

Bartley, N., Abeliuk, A., Ferrara, E., Lerman, K., 2021. Auditing algorithmic bias on Twitter. In: Proceedings of the 13th ACM Web Science Conference 2021. WebSci '21. 06. Association for Computing Machinery, New York, NY, USA, pp. 65–73.

Brynjolfsson, E., Rock, D., Syverson, C., 2018. Artificial intelligence and the modern productivity paradox: a clash of expectations and statistics. In: The Economics of Artificial Intelligence: An Agenda. University of Chicago Press, pp. 23–57.

Brynjolfsson, E., Hui, X., Liu, M., 2019. Does machine translation affect international trade? Evidence from a large digital platform. Manag. Sci. 65 (12), 5449–5460.

Buffer, 2018. Facebook Advertising. The Complete Guide to Facebook Ads Manager: How to Create, Manage, Analyze Your Facebook Ads.

Cecere, G., Jean, C., Lefrere, V., Tucker, C., 2020. Trade-offs in Automating Platform Regulatory Compliance by Algorithm: Evidence From the COVID-19 Pandemic. Working Paper (SSRN).

Cecere, G., Le Guel, F., Lefrere, V., Tucker, C.E., Yin, P.-L., 2022. Privacy, Data and Competition: The Case of Apps for Young Children (Available at SSRN 4073931).

Cheng, X., Su, L., Luo, X., Benitez, J., Cai, S., 2022. The good, the bad, and the ugly: impact of analytics and artificial intelligence-enabled personal information collection on privacy and participation in ridesharing. Eur. J. Inf. Syst. 31 (3), 339–363.

Cowgill, B., Tucker, C., 2017. Algorithmic Bias: A Counterfactual Perspective. NSF Trustworthy Algorithms.

Cowgill, B., Tucker, C.E., 2019. Economics, Fairness and Algorithmic Bias. Working Paper.

Cui, T.H., Ghose, A., Halaburda, H., Iyengar, R., Pauwels, K., Sriram, S., Tucker, C., Venkataraman, S., 2021. Informational challenges in omnichannel marketing: remedies and future research. J. Mark. 85 (1), 103–120.

Datta, A., Tschantz, M.C., Datta, A., 2015. Automated experiments on ad privacy settings: a tale of opacity, choice, and discrimination. In: Proceedings on Privacy Enhancing Technologies, vol. 15. De Gruyter Open, pp. 92–112 (September).

Datta, A., Datta, A., Makagon, J., Mulligan, D.K., Tschantz, M.C., 2018. Discrimination in online advertising: a multidisciplinary inquiry. In: Friedler, S.A., Wilson, C. (Eds.), Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Proceedings of Machine Learning Research, vol. 81. PMLR, pp. 20–34 (23–24 Feb).

Dujeancourt, E., Garz, M., 2023. The effects of algorithmic content selection on user engagement with news on twitter. Inf. Soc. 39 (5), 263–281.

Dwivedi, Y.K., Sharma, A., Rana, N.P., Giannakis, M., Goel, P., Dutot, V., 2023. Evolution of artificial intelligence research in technological forecasting and social change: research topics, trends, and future directions. Technol. Forecast. Soc. Chang. 192, 122579.

Erdmann, A., Ponzoa, J.M., 2021. Digital inbound marketing: measuring the economic performance of grocery E-commerce in Europe and the USA. Technol. Forecast. Soc. Chang. 162, 120373.

Fischer, S., Jaidka, K., Lelkes, Y., 2020. Auditing local news presence on Google News. Nat. Hum. Behav. 4 (12), 1236–1244.

Garz, M., Szucs, F., 2023. Algorithmic selection and supply of political news on Facebook. Inf. Econ. Policy 62, 101020.

Goldfarb, A., 2014. What is different about online advertising? Rev. Ind. Organ. 44 (2), 115–129.

Goldfarb, A., Tucker, C., 2011. Online display advertising: targeting and obtrusiveness. Mark. Sci. 30 (3), 389–404.

Gordon, B.R., Zettelmeyer, F., Bhargava, N., Chapsky, D., 2019. A comparison of approaches to advertising measurement: evidence from big field experiments at Facebook. Mark. Sci. 38 (2), 193–225.

Greenwood, B., Adjerid, I., Angst, C.M., Meikle, N.L., 2022. How unbecoming of you: online experiments uncovering gender biases in perceptions of ridesharing performance. J. Bus. Ethics 175, 499–518.

Haaland, I., Roth, C., Wohlfart, J., 2023. Designing information provision experiments. J. Econ. Lit. 61 (1), 3–40.

Horton, J.J., 2023. Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus? Working Paper 31122. National Bureau of Economic Research.

Huszár, F., Ktena, S.I., O'Brien, C., Belli, L., Schlaikjer, A., Hardt, M., 2022. Algorithmic amplification of politics on Twitter. Proc. Natl. Acad. Sci. 119 (1), e2025334119.

Imana, B., Korolova, A., Heidemann, J., 2023. Having your privacy cake and eating it too: platform-supported auditing of social media algorithms for public interest. In: Nichols, J. (Ed.), Proceedings of the ACM on Human-Computer Interaction, vol. 7 (04). Association for Computing Machinery, pp. 1–33.

Khan, A., Brohman, K., Addas, S., 2022. The anatomy of 'fake news': studying false messages as digital objects. J. Inf. Technol. 37 (2), 122–143.

Kingsley, S., Wang, C., Mikhalenko, A., Sinha, P., Kulkarni, C., 2020. Auditing Digital Platforms for Discrimination in Economic Opportunity Advertising. arXiv preprintarXiv:2008.09656.

Kokshagina, O., Reinecke, P.C., Karanasios, S., 2023. To regulate or not to regulate: unravelling government institutional work towards AI regulation. J. Inf. Technol. 38 (2), 160–179.

Kordzadeh, N., Ghasemaghaei, M., 2022. Algorithmic bias: review, synthesis, and future research directions. Eur. J. Inf. Syst. 31 (3), 388–409.

Kwak, K.T., Lee, S.Y., Lee, S.W., 2021. News and user characteristics used by personalized algorithms: the case of Korea's News Aggregators, Naver News and Kakao News. Technol. Forecast. Soc. Chang. 171, 120940.

Lambrecht, A., Tucker, C., 2019. Algorithmic bias? An empirical study into apparent gender-based discrimination in the display of STEM career ads. Manag. Sci. 65 (7), 2966–2981.

Lau, Y., 2020. A Brief Primer on the Economics of Targeted Advertising. Technical Report. Federal Trade Commission.

List, J.A., Price, M.K., 2009. The role of social connections in charitable fundraising: evidence from a natural field experiment. J. Econ. Behav. Organ. 69 (2), 160–169.

Lyytinen, K., Nickerson, J.V., King, J.L., 2021. Metahuman systems = humans + machines that learn. J. Inf. Technol. 36 (4), 427–445.

Marjanovic, O., Cecez-Kecmanovic, D., Vidgen, R., 2021. Algorithmic pollution: making the invisible visible. J. Inf. Technol. 36 (4), 391–408.

Marthews, A., Tucker, C., 2019. Privacy Policy and Competition. Brookings Paper.

Matter, U., Hodler, R., Ladwig, J., 2022. Personalization of Web Search During the 2020 US Elections. arXiv Preprint arXiv:2209.14000.

Mejia, J., Parker, C., 2021. When transparency fails: bias and financial incentives in ridesharing platforms. Manag. Sci. 67 (1), 166–184.

Metaxa, D., Park, J.S., Robertson, R.E., Karahalios, K., Wilson, C., Hancock, J., Sandvig, C., et al., 2021. Auditing algorithms: understanding algorithmic systems from the outside in. Foundations and Trends® Hum.–Comput. Interact. 14 (4), 272–344.

Mikalef, P., Conboy, K., Lundström, J.E., Popovič, A., 2022. Thinking responsibly about responsible AI and 'the dark side' of AI. Eur. J. Inf. Syst. 31 (3), 257–268.

Möhlmann, M., Zalmanson, L., Henfridsson, O., Gregory, R.W., 2021. Algorithmic management of work on online labor platforms: when matching meets control. MIS Q. 45 (4), 1999–2022.

Mullainathan, S., Obermeyer, Z., 2017. Does machine learning automate moral hazard and error? Am. Econ. Rev. 107 (5), 476–480.

OECD, 2019. Artificial Intelligence in Society. Technical Report. OECD.

Omrani, N., Rivieccio, G., Fiore, U., Schiavone, F., Agreda, S.G., 2022. To trust or not to trust? An assessment of trust in AI-based systems: concerns, ethics and contexts. Technol. Forecast. Soc. Chang. 181, 121763.

O'Neil, C., 2016. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group, New York, NY, USA.

Orphanou, K., Otterbacher, J., Kleanthous, S., Batsuren, K., Giunchiglia, F., Bogina, V., Tal, A.S., Hartman, A., Kuflik, T., 2022. Mitigating bias in algorithmic systems—a fish-eye view. ACM Comput. Surv. 55 (5).

Pietronudo, M.C., Croidieu, G., Schiavone, F., 2022. A solution looking for problems? A systematic literature review of the rationalizing influence of artificial intelligence on decision-making in innovation management. Technol. Forecast. Soc. Chang. 182, 121828.

Rambachan, A., Kleinberg, J., Ludwig, J., Mullainathan, S., 2020. An economic perspective on algorithmic fairness. AEA Pap. Proc. 110 (05), 91–95.

Reuning, K., Whitesell, A., Hannah, A.L., 2022. Facebook algorithm changes may have amplified local republican parties. Res. Polit. 9 (2), 20531680221103809.

Riemer, K., Peter, S., 2021. Algorithmic audiencing: why we need to rethink free speech on social media. J. Inf. Technol. 36 (4), 409–426.

Sandvig, C., Hamilton, K., Karahalios, K., Langbort, C., 2014. Auditing algorithms: research methods for detecting discrimination on internet platforms. In: Data and Discrimination: Converting Critical Concerns Into Productive Inquiry. 64th Annual Meeting of the International Communication Association, vol. 22(05), pp. 4349–4357.

Sweeney, L., 2013. Discrimination in online ad delivery. Queue 11 (3), 10:10–10:29.

Tucker, C.E., 2014. Social networks, personalized advertising, and privacy controls. J. Mark. Res. 51 (5), 546–562.

Tucker, C., 2019. Privacy, algorithms, and artificial intelligence. In: Agrawal, A., Gans, J., Goldfarb, A. (Eds.), The Economics of Artificial Intelligence: An Agenda. University of Chicago Press, pp. 423–437.

Tucker, C., 2023. Algorithmic exclusion: the fragility of algorithms to sparse and missing data. In: Brookings. Center on Regulation and Markets at Brookings, pp. 1–26.

Vecchione, B., Levy, K., Barocas, S., 2021. Algorithmic auditing and social justice: lessons from the history of audit studies. In: Equity and Access in Algorithms, Mechanisms, and Optimization. Association for Computing Machinery, pp. 1–9.

Vlačić, B., Corbo, L., Costa e Silva, S., Dabić, M., 2021. The evolving role of artificial intelligence in marketing: a review and research agenda. J. Bus. Res. 128, 187–203.

Xenidis, R., Gerards, J., et al., 2021. Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-discrimination Law. European Commission.

Yang, S., Lin, S., Carlson, J.R., Ross Jr., W.T., 2016. Brand engagement on social media: will firms' social media efforts influence search engine advertising effectiveness? J. Mark. Manag. 32 (5–6), 526–557.

Zou, J., Schiebinger, L., 2018. AI can be sexist and racist—it's time to make it fair. Nature 559 (7714), 324–326.

Cecere, G. is Professor of Economics at Institut Mines Telecom, Business School, LITEM. She completed her PhD in Economics at the University of Paris Saclay (France) and the University of Turin (Italy). She holds her master's degree from the College of Europe, Natolin in Poland. She was a visiting researcher at SPRU, University of Sussex and ZEW, Mannheim. Her main research interests are digital economy and more particularly the economics of privacy, algorithms and machine learning, economics of mobile applications, and digital marketing. In 2019, she was awarded the 'Marie-Dominique Hagelsteen' prize for responsible advertising by the French professional regulation of advertising authority. She is scientific expert for the French Audiovisual Regulator, Arcom (2020–2024) and she is scientific evaluator for the INRAE (2020–2024). She obtained funding from numerous institutions, including grants and prizes from Institut DataIA (France), ANR (French National Research Agency), DAWEX (a French company) and the Social Science Foundation. She has several collaborations with researchers in the USA and in major European Institutions. She is on the board of the "Association Francophone de Recherche en Economie Numérique".

Jean, C. is currently an assistant professor in Information Systems at Grenoble Ecole de Management. She obtained her PhD in Economics from Université Paris-Saclay. During her PhD, she was also a research engineer at Epitech. In 2019, she was a visiting researcher in the department of economics at Clemson University, Clemson (USA). Clara holds double master's degrees from Paris Dauphine University with a specialization in digital economics and from Aix-Marseille School of Economics with a specialization in economic engineering. Her research focuses on algorithmic bias, algorithmic decision-making in the context of advertising and personal data. In 2019, she was awarded the Marie-Dominique Hagelsteen prize for responsible advertising from the French professional regulation of advertising authority (ARPP), and received a grant from University Paris-Saclay to organize a workshop on Natural Language Processing (NLP) and gender bias. Her current research focuses on understanding the drivers of algorithmic decision-making in online advertising for STEM fields.

Le Guel, F. is an associate professor (HDR) at the University Paris-Saclay, in the RITM research center in economics. He has published in international journals such as Applied Economics, Technological Forecasting and Social Change or Review for Economic Research on Copyright Issues. He has already participated in several research projects financed by the French ANR related to digital economics such as ESPRI (economics of privacy), MOBITICS (mobility and ICT use), EXPERTIC (business model of digital economy) and co-leads an interdisciplinary research project named DAPCODS (2017–2021).

Manant, M. is an associate professor (HDR) at the University Paris-Saclay, in the RITM research center in economics. He has been a researcher at the RITM since 2009. He defended his thesis in 2009 on the theme of cooperation in innovation as a competitive strategy. He works on the Economy of Innovation by addressing in particular issues of inter-company cooperation, and more broadly collaborative strategies. More recently, he has also been interested in the digital economy, particularly currently in privacy issues on smartphones and on social networks.