EDRi

EUROPEAN DIGITAL RIGHTS

How online ads discriminate

—

Unequal harms of online advertising in Europe

Booklet written by Frederike Kaltheuner
Reviewed by Sarah Chander and Jan Penfrat
Edited by Gail Rego

"From widespread data exploitation that is virtually impossible to avoid, to a lack of accountability in the data supply chain, targeted ads raise fundamental rights concerns, issues around consumer protection, as well as broader societal harms."

[1] See for instance: Kingaby, H., & Kaltheuner, F. (2020). Ad Break for Europe: The Race to Regulate Digital Advertising and Fix Online Spaces. Retrieved from https://assets.mofoprod.net/network/documents/Ad_Break_ for_Europe_FINAL_online.pdf

[2] The Digital Freedom Fund and its partner European Digital Rights (EDRi) are in the initial phases of a new initiative to begin a decolonising process for the digital rights field. See: https://digitalfreedomfund.org/ decolonising/

[3] Kelly, N. (2020, May 2). Coronavirus: 'I'm Being Bombarded by Gambling Ads'. Retrieved from https://www.bbc.com/news/stories-52506113

Introduction

The first online banner ad appeared in 1994, and worked similarly to billboards that appear next to highways, or advertising pages in print magazines: AT&T paid HotWired $30,000 to place a banner ad on their site for three months so that every visitor to that site would see it right on top.

Much has changed since then. Today, hyper targeted online ads have become ubiquitous. They appear in social media stories, in social media feeds, in video content, on apps, next to news stories and on a significant share of the world's websites, blogs and publishers' sites.

The risks and harms that are associated with hyper targeted online ads have been widely documented.[1] From widespread data exploitation that is virtually impossible to avoid, to a lack of accountability in the data supply chain, targeted ads raise fundamental rights concerns, issues around consumer protection, as well as broader societal harms.

On top of all of this, there is little evidence that the amount of tracking and the invasiveness with which most ads are targeted today actually makes them more relevant to those who see them.

One issue, however, that has not received the same amount of attention is the many ways in which harms and risks of online advertising are unequally distributed, and how targeted online advertising can have discriminatory effects. This is the focus of this report.

Discrimination in online advertising is a topic that is both timely and urgent. Unequal treatment and discrimination remain a reality in Europe. There is also an ongoing need to decolonise the digital rights field to ensure that the field reflects the society that it works to safeguard.[2]

Part of this process is also an acknowledgement that digital rights violations often disproportionately affect those who are already marginalised.

The focus on discrimination in online advertising is timely, because the European Commission is embarking on an ambitious plan to regulate tech companies and shape the direction of Europe's digital transformation. New or strengthened rules for digital advertising could be implemented in the Digital Services Act (DSA), the EU Regulation on Artificial Intelligence, the Democracy Action Plan, the ePrivacy Regulation, and the Digital Markets Act.

Tackling discrimination, specifically in online advertising, has also become more urgent. The ongoing COVID-19 pandemic means that many people's work and private lives have entirely moved online, amplifying the negative effects of targeted ads, especially for marginalised groups and people in vulnerable situations.

Targeted advertising allows advertisers to target people at an increasingly granular level.

As a result, people struggling with gambling addictions in the UK have reported that they are being bombarded with gambling ads[3], while YouTube announced in December 2020 that they would allow users to mute gambling and alcohol ads. [4]

The pandemic has also had a devastating impact on people struggling with eating disorders , and media reports show that those who are in recovery or struggling with an eating disorder[5] are finding diet ads on platforms like TikTok or Instagram distressing.[6]

---

[4] BBC (2020, December 11). YouTube Lets Users Mute Gambling and Alcohol Ads. Retrieved from https://www.bbc.com/news/technology-55273687

[5] Northumbria University (2020, August 23). Research Reveals a Toll of Pandemic on Those with Eating Disorders. Retrieved from https://www.sciencedaily.com/releases/2020/08/200823201524.html

[6] Dawson, B. (2020, September 25). Eating Disorder Sufferers on the Danger of Weight Loss Ads on TikTok. Retrieved from https://www.dazeddigital.com/life-culture/article/50566/1/eating-disorder-sufferers-on-the-danger-of-weight-loss-ads-on-tiktok

"The pandemic has had a devastating impact on people struggling with eating disorders , and media reports show that those who are in recovery or struggling with an eating disorder are finding diet ads on platforms like TikTok or Instagram distressing."

TABLE CONTE

OF NTS

01

Discrimination in online advertising

There are two different ways of thinking about discrimination in online advertising: a narrow sense and a broader sense. In the narrow sense, discrimination can occur as a direct result of targeted online advertising.

A person or a group that is shown a targeted ad has either been discriminated against directly or indirectly, through harmful targeting or exclusion from an ad.

Discrimination can also occur in other areas of the broader online advertising ecosystem, such as in the many ways in which data is collected, processed and shared for advertising purposes, in the ways in which advertising supported platforms recommend content, or in decisions about which content and which content producers can rely on advertising to monetise their content online.

This is discrimination in online advertising in the broader sense.

Discrimination in online advertising can result in a number of harms to individuals.

Targeting that leads to unfair exclusion

Ads that exclude people can lead to unfair exclusion. In the case of online job or housing ads that either exclude, or predominately target a specific demographic or otherwise defined group, discriminatory outcomes in online advertising mean that protected groups are excluded from opportunities.

Harmful targeting

Specifically targeting (protected) groups can also lead to harm and distress. For instance, the fact that an ad seems to be based on knowledge about protected categories alone can be distressing and is an invasion of privacy.

One example is when someone has not disclosed their sexual orientation publicly, but an ad assumes their sexual orientation. Targeting of (protected) groups with ads or content that has a negative connotation can also lead to harm, for instance when Google searches for names are associated with negative ads, such as for criminal background checks.

The fact that advertisers can target people at a granular level, including based on protected categories, means that this ability can be exploited.

Misclassification in profiling

Advertising uses a range of techniques to identify and profile individuals. Behavioural advertising in particular can infer very sensitive information (e.g., ethnicity, gender, sexual orientation, religious beliefs) about individuals. Wachter (2020) calls this "affinity profiling", grouping people according to their assumed interests rather than solely their personal traits.[7]

Since such inferences may be inaccurate, or otherwise systematically biased, profiling may lead to individuals being misidentified, or misclassified and such inaccuracies may result in ad targeting that is discriminatory. Such profiling may also form the basis of discrimination, for instance harmful targeting, or targeting to exclude.

Blacklisting of content for advertising

Advertising vendors and brands can block words associated with certain content from monetisation, for instance on news sites.

As a result, news articles on topics that contain or mention blocked words cannot show certain ads, which means reduced or even zero income for publishers. For instance, the word "Coronavirus" was declared "brand unsafe", which meant that the front pages of major news sites were running without ads at the beginning of the pandemic.

According to Jerry Daykin of Outvertising, 73% of LGBTQ+ content is rendered unmonetisable under current blacklists, and keyword exclusion lists include generic terms like "Lesbian" or "Muslim" more often than terms such as "murder".[8]

Advertising is funding hate speech

Online advertising has created a market for smaller sites to monetise content. That includes diverse and marginalised voices, but also far-right websites and disinformation. Since brands often do not know where their ads are displayed, initiatives like Stop Funding Hate and Sleeping Giants are encouraging advertisers to revisit their supply chains and withdraw their ads from websites that encourage hate speech.

At the same time, advertising funds social media platforms, such as YouTube, Facebook and Twitter, many of whom are financially benefitting from hate speech and disinformation on their platforms.

7 Wachter, S. (2020). Affinity Profiling and Discrimination by Association in Online Behavioural Advertising. Berkeley Technology Law Journal, 35(2), pp. 1-74.

8 Daykin, J. (2019, November 13). Save Digital Advertising, Save the World [LinkedIn post]. Retrieved from https://www.linkedin.com/pulse/save-digital-advertising-world-togetherwecan-jerry-daykin/

02

How discrimination occurs in Ad targeting

The advertising ecosystem is a vast, distributed, and decentralised system with multiple actors: There are publishers who publish content online, platforms that host content, advertisers who seek to place their ads, consumers who consume content online, and ad networks, who connect publishers and advertisers. [9]

As a result of the vast advertising ecosystem, there are multiple ways in which discrimination can occur:

An advertiser explicitly and intentionally targets or excludes a group

Here the advertiser deliberately uses targeting criteria provided by a platform, or uploads their own customer, tracking and purchase data to target or exclude a group of people.

An advertiser indirectly or inadvertently targets or excludes a group

Discrimination can also occur indirectly (sometimes inadvertently). Datta et. al (2018) mention three mechanisms through which discrimination in ad targeting can occur indirectly:

- **Via a proxy, or a known correlate**
- **Via a known correlate, but not because it is a correlate**
- **Via an unknown correlate**

Proxies are targeting criteria that are known to correlate with certain criteria. Targeting people who use menstrual apps, for instance, means that an advertiser is likely targeting women, or people who menstruate.

Advertisers can also inadvertedly target a correlate. In racially segregated cities, targeting by postcode can be a proxy for race and socio-economic status. The same happens when interests are used to target groups. This can either be a deliberate way to target people based on special category data, for instance, when advertisers target people with

an interest in "LGBTQ issues" when trying to reach people who identify as LGBTQ.

Finally, there might be correlates between a category and other targeting criteria that are unknown. Such indirect targeting or exclusion, especially when using multiple targeting criteria, can also happen without the explicit intention of the advertiser.

This form of indirect and sometimes inadvertent discrimination or targeting is also common in automated targeting techniques that use machine learning. Facebook's Lookalike Audience, for instance, automatically finds an audience that is similar to an audience that the advertiser knows already (either because they follow or like their page, or because the advertiser has tracked them on their website or app).

In automated techniques like Lookalike Audience, discrimination based on an unknown correlate is an inherent risk, unless proactive steps are taken to continuously audit and tackle discrimination. That is because these techniques find targeting criteria automatically. If an advertiser for real estate has a known audience or customer base that is male and white, for instance, automated targeting techniques will likely target these audiences, thereby excluding everyone who is not white and male.

- **Protected groups are either more likely or less likely to click on and engage with an ad**

  Even when ads are targeted based on neutral criteria, the way in which an ad is designed could mean that certain groups of people are more or less likely to click and engage with it.

  For instance, the text or image used in an add could make it more likely for people of a certain age to engage with the app. This can also have feedback loops with ad optimisation.

- **Protected groups are less likely to spend time on mediums where an ad is placed**

  Similarly, when certain groups are less likely to spend time wherever an ad is displayed, this means that the group is less likely to view and engage with the ad. Again, this can also have

feedback loops with ad optimisation (see below).

The automated ad delivery leads to discriminatory outcomes

Discrimination can also happen during the ad optimisation process.

As this report explains later, even ads that are not specifically targeted can end up being heavily biased, based on ad optimisation processes that automatically display ads to those who are assumed to be the most likely to engage.

The bidding process: decisions of other advertisers

Since ads are auctioned, the decisions of other advertisers can have an impact on who views an ad.

As Datta et al. (2018) explain with regards to gender discrimination in Google AdWords, "if advertisers in general consider female consumers to be a more valuable demographic, they would set higher bids to advertise to them. As a result, if an advertiser [...] sets equal bids for men and women, it could end up only reaching men if it is out bid by other ads for female users."[10]

[9] Datta, A., Datta, A., Makagon, J., Mulligan, D. K., & Tschantz, M. C. (2018). Discrimination in Online Advertising: A Multidisciplinary Inquiry. In Conference on Fairness, Accountability and Transparency. New York University, New York City, USA. Retrieved from http://proceedings.mlr.press/v81/datta18a/datta18a.pdf

[10] Idem.

03

Evidence of discrimination in online advertising

Discrimination in online advertising is a widely studied phenomenon. When reviewing literature on discrimination in online advertising, it is important to keep in mind that the techniques used to target ads and the platform policies that guide online advertising are constantly changing and evolving.

Online advertising is highly dynamic. As Asplund et al. (2000) argue:

> **Practically every factor in these systems is constantly evolving, from the set of ads currently being served, to the targeting and pricing of an advertising campaign, and even the way user profiles are interpreted.**

This puts researchers in a difficult position: auditors must collect as much data as possible in order to catch any confounding variables and must carefully validate that the system they are measuring did not change substantially during the course of their audit.[11]

The online advertising industry as we know it today, is also incredibly complex. Evidence for discrimination on one particular advertising platform, does not necessarily prove that similar discrimination occurs elsewhere, since platform policies and targeting techniques differ. The following explores discrimination on varying platforms.

[11] Asplund, J., Eslami, M., Sundaram, H., Sandvig, C., & Karahalios, K. (2020, May). Auditing Race and Gender Discrimination in Online Housing Markets. Proceedings of the International AAAI Conference on Web and Social Media 14(1), pp. 24-35.

3.1 Google

The first major study on discrimination in online ad delivery was published by Latanya Sweeney in 2013.[12] Based on searches done in the United States, Sweeney found that Google AdSense ads for public records on a person appeared more often for those with black-associated names than with white-associated names, regardless of company.

Furthermore, a greater percentage of Instant Checkmate ads that were using the word "arrest" appeared for black-identifying first names than for white first names.

The study itself raised a number of issues which would soon become recurring themes in this area of research. First of all, this pioneering study shows how even statistically significant discrimination in automated systems is incredibly difficult to prove for those affected.

Even though frequent spotting of arrest records ads next to black-associated names inspired this study, it took comprehensive research to prove that this is not a coincidence, but rather a systemic problem.

Secondly, the study itself could not conclusively identify the reasons why discrimination occurred, or whether this is the fault of the advertiser, Instant Checkmate, Google, or society at large. In the words of Sweeney, "this study raises more questions than it answers."[13]

One reason for this is the inner workings of Google AdSense, specifically the automated and dynamic nature of ad delivery.

Google places keyword-based advertisement slots for various "firstname lastname" searches. Advertisers were able to provide multiple templates for the same search string and Google optimised which search string to display, based on which people are most likely to click on it.

As a result, it is impossible to establish from the outside, whether the advertiser created ad templates suggestive of arrest disproportionately to black-identifying names, or whether the system was providing roughly the same templates evenly across racially associated names, but people who search online were more likely to click on ads suggestive of arrest more often for black-identifying names.

Future research, both by Sweeney (2013) and others, has sought to replicate evidence of discrimination for different types of advertising, while also trying to establish likely causes for discriminatory ads. In 2015, Datta, Tschantz and Datta found that males were shown ads encouraging the seeking of coaching services for high paying jobs more than females.[14]

The study was focused on Google's Ad Settings, a feature introduced at the time, that shows, and allows users to control inferences Google had made about a user's demographics and interest based on their browsing behaviour.

A follow up study from 2018 discusses the causes behind discrimination in the specific case raised in the 2015 study on discrimination of Google AdWords ads.[15]

12 Sweeney, L. (2013). Discrimination in Online Ad Delivery. Communications of the ACM, 56(5), pp. 44-54.

13 Idem.

14 Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. Proceedings on Privacy Enhancing Technologies, 2015(1), pp. 92-112.

The study provides a very useful classification about how discriminatory outcomes come about and who creates inputs that might contribute to a discriminatory outcome in the case of Google AdWords ads:

Factor I: (Who) Possible mechanisms leading to males seeing the ads more often include:

Google alone
Explicitly programming the system to show the ad less often to females, e.g., based on independent evaluation of demographic appeal of product (explicit and intentional discrimination).

The advertiser
The advertiser targeting the ad through explicit use of demographic categories (explicit and intentional discrimination), the pretextual selection of demographic categories and/or keywords that encode gender (hidden and intentional), or through those choices without intent (unconscious selection bias), and Google respecting these targeting criteria.

Other advertisers
Other advertisers' choice of demographic and keyword targeting and bidding rates, particularly those that are gender specific or divergent, that compete with the ad under question in Google's auction, influencing its presentation.

Other consumers
Male and female consumers behaving differently to ads because:

a. Google learned that males are more likely to click on this ad than females
b. Google learned that females are more likely to click on other ads than this ad, or
c. Google learned that there exist ads that females are more likely to click on than males are; and

Multiple parties
Some combination of the above.

Factor II: (How) The mechanisms can come in multiple forms based on how the targeting was conducted:

1. on gender directly
2. on a proxy for gender, i.e., on a known correlate of gender because it is a correlate
3. on a known correlate of gender, but not because it is a correlate, or
4. on an unknown correlate of gender

In 2020, the U.S. Department of Housing and Urban Development (HUD), which has filed a lawsuit against Facebook (see below) announced that it had "worked with Google to improve Google's online advertising policies to better align them with requirements of the Fair Housing Act."

As a result of this, Google banned job, housing, and credit advertisers from excluding either men or women from their ads, along with similar rules for age and other protected groups. [16]

In 2021, research by The Markup showed that Google allowed advertisers to exclude nonbinary people from seeing job ads.[17]

[15] Datta, A., Datta, A., Makagon, J., Mulligan, D. K., & Tschantz, M. C. (2018). Discrimination in Online Advertising: A Multidisciplinary Inquiry. In Conference on Fairness, Accountability and Transparency. New York University, New York City, USA. Retrieved from http://proceedings.mlr.press/v81/datta18a/datta18a.pdf

[16] Merrill, J. B. (2021, February 21). Google Has Been Allowing Advertisers to Exclude Nonbinary People from Seeing Job Ads. Retrieved from https://themarkup.org/google-the-giant/2021/02/11/google-has-been-allowing-advertisers-to-exclude-nonbinary-people-from-seeing-job-ads

[17] Idem.

3.2 Facebook

There is also clear evidence of discrimination in various forms of advertising used by Facebook, even though the platform bans discriminatory advertising in its ads policy.[18]

Facebook has the highest ad volume amongst social media platforms. It also offers numerous ways in which advertisers can target ads on Facebook.

Figure I – Facebook advertising: Targeting techniques offered by Facebook:

Core Audiences

Advertisers can define an audience based on targeting criteria offered by Facebook, such as age, interests, geography and more.

These include over 200,000 attributes which can result in complex targeting formulas when combined.[19]

These attributes can reveal protected categories and special categories of personal data, especially when combined. A few of these targeting attributes are:

a. Location
b. Demographics
c. Interests (including pages liked and engaged with)
d. Behaviour (i.e., prior purchases and device usage)
e. Connections
f. Life events (away from family, away from hometown, long distance relationship, new job, new relationship, recently moved, upcoming birthday)
g. Parents

h. Job title, education
i. Relationship status
j. Languages

Custom Audiences

Advertisers can also upload their own data to Facebook
a. Contact lists (emails and phone numbers)
b. Site visitors (tracking data)
c. App users (tracking data)

Lookalike Audiences

Here Facebook automatically identifies audiences that are similar to an audience that the advertiser already knows. Facebook will then reach people with common interests and traits.

Optimisation for Ad Delivery (optional)

In addition to the targeting options above, advertisers can choose to automatically optimise ad delivery based on a chosen outcome (i.e., number of people who click on the link, or visit the advertiser's website).[20]

Facebook also allows advertisers to automatically A/B test different ads and ad targeting options to help advertisers decide which version works best for their defined goals.[21]

Placement of ads on Facebook

Feeds

a. Facebook News Feed: Ads appear in the desktop News Feed when people access the Facebook website on their computers. Ads appear in the mobile News Feed when people use the Facebook app on mobile devices or access the Facebook website through a mobile browser.

b. Instagram Feed: Ads appear in the mobile feed when people use the Instagram app on mobile devices. Instagram Feed ads only appear to people browsing the Instagram app.

---

18 Facebook (n.d.). Restricted Content [Facebook page]. Retrieved from https://www.facebook.com/policies/ads/ restricted_content

19 Havlak, H., & Abelson, B (2016, February 1). The Definitive List of What Everyone Likes on Facebook. Retrieved from https://www.theverge.com/2016/2/1/10872792/facebook-interests-ranked-preferred-audience-size

20 Facebook (n.d.). Business Help Center [Facebook page]. Retrieved from https://www.facebook.com/business/ help/355670007911605

21 Facebook (n.d.). Facebook Measurement [Facebook page]. Retrieved from https://www.facebook.com/business/measurement

c. Facebook Marketplace: Ads appear in the Marketplace home page or when someone browses Marketplace in the Facebook app.

d. Facebook Video Feeds: Video ads appear between organic videos in video-only environments on Facebook Watch and Facebook News Feed.

e. Facebook Right Column: Ads appear in the right column on Facebook. Right column ads only appear to people browsing Facebook on their computers.

f. Instagram Explore: Ads appear in the browsing experience when someone clicks on a photo or a video.

g. Messenger Inbox: Ads appear in the Home tab of Messenger.

Stories

a. Facebook Stories: Ads appear in people's Stories on Facebook.

b. Instagram Stories: Ads appear in people's Stories on Instagram.

c. Messenger Stories: Ads appear in people's Stories on Messenger.

In-stream

Facebook In-Stream Videos: Ads appear in Video on Demand and in a select group of approved partner live streams on Facebook.

Search

Facebook Search Results: Ads appear next to relevant Facebook and Marketplace search results.

Messages

Messenger Sponsored Messages: Ads appear as messages to people who have an existing conversation with the advertiser in Messenger.

In-Article

Facebook Instant Articles: Ads appear in Instant Articles within the Facebook mobile app.

Apps

a. Audience Network Native, Banner and Interstitial: Ads appear on apps on Audience Network.

b. Audience Network Rewarded Videos: Ads appear as videos people can watch in exchange for a reward in an app (such as in-app currency or items).

Numerous studies have looked at discrimination in various aspects of Facebook advertising to determine whether discrimination has occurred. These can be broadly placed in two categories, before the March 2019 US settlement between civil rights advocates and after.[22]

Between 2016 and 2018, five discrimination lawsuits and charges were filed in the US against Facebook by civil rights groups, a national labour organisation, workers, and consumers.[23]

Each of these cases refers to different audience selection and targeting tools that are available on the Facebook ad platform, such as the targeting criteria provided by Facebook that allow advertisers to directly or indirectly target or exclude audiences based on sex, age, race, national origin, or family status; the ability of advertises to create narrow location-based targeting that could have an adverse effect based on race or national origin; and the impact of the Facebook Lookalike Audience tool to impact various groups, including based on gender, race and age.[24]

Prior to the settlement, various papers and reports had identified discrimination in online recruiting on Facebook.[25] ProPublica[26] also found that Facebook enabled advertisers to not only discriminate but also specifically target audiences with racist views, for instance by targeting "Jew haters."[27]

Under the settlement, Facebook agreed to a number of changes to its advertising platform that were designed to prevent advertisers for housing, employment or credit from discriminating based on race, national origin, ethnicity, age, sex, sexual orientation, disability, family status, or other characteristics covered by federal, state, and local civil rights laws in the US.

22 ACLU (2019, March 19). Summary of Settlements Between Civil Rights Advocates and Facebook. Retrieved from https://www.aclu.org/other/summary-settlements-between-civil-rights-advocates-and-facebook

23 Idem.

24 Idem.

25 Kim, P. T., & Scott, S. (2018). Discrimination in Online Employment Recruiting. St. Louis University Law Journal, 63(1), pp. 1-28.

Various papers and
have identified dis
online employmen
Facebook. ProPubl
that Facebook ena
to not only discrim
specifically target
with racist views,
targeting "Jew hat

d reports
crimination in
nt recruiting on
lica also found
abled advertisers
minate but also
t audiences
for instance by
ers."

These changes have not eliminated discrimination on the platform. A study by Ali et al. (2018)[28] in the US shows that ad optimisation can, still today, lead to discriminatory ads on Facebook. The paper demonstrates that ad delivery is often skewed along racial and gender lines for ads on employment and housing opportunities.

These discriminatory outcomes happened despite neutral ad targeting parameters. Reasons for this included market and financial optimisation effects as well as the platform's own predictions about the "relevance" of ads to different groups of users. Another contributing factor is the advertiser's budget and the content of the ad.

Research by Sapiezynski et al. (2019) looked into Facebook's modified Lookalike Audience tool, called SpecialAd Audiences.[29]

The researchers found that "relative to Lookalike Audiences, SpecialAd Audiences do little to reduce demographic biases in target audiences."

---

26 Speicher, T., Ali, M., Venkatadri, G., Ribeiro, F. N., Arvanitakis, G., Benevenuto, F., ... & Mislove, A. (2018). Potential for Discrimination in Online Targeted Advertising. Proceedings of Machine Learning Research, 81, pp. 5–19.

27 Angwin, J., Varner, M., & Tobin, A. (2017, September 14). Facebook Enabled Advertisers to Reach Jew Haters. Retrieved from https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters

28 Ali, M., Sapiezynski, P., Bogen, & Korolova, A. (2019). Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. Proceedings of the ACM on Human-Computer Interaction, 3, pp. 1-30.

29 Sapiezynski, P., Ghosh, A., Kaplan, L., Mislove, A., & Rieke, A. (2019). Algorithms that "Don't See Color": Comparing Biases in Lookalike and Special Ad Audiences. arXiv preprint arXiv:1912.07579. Retrieved from https://arxiv.org/pdf/1912.07579.pdf

30 Andreou, A., Silva, M., Benevenuto, F., Goga, O., Loiseau, P., & Mislove, A. (2019). Measuring the Facebook Advertising Ecosystem. NDSS 2019 - Proceedings of the Network and Distributed System Security Symposium. San Diego, California, United States. Retrieved from https://hal.archives-ouvertes.fr/hal-01959145/document

31 Kingsley, S., Wang, C., Mikhalenko, A., Sinha, P., & Kulkarni, C. (2020). Auditing Digital Platforms for Discrimination in Economic Opportunity Advertising. 4th Workshop on Mechanism Design for Social Good. Retrieved from https://arxiv.org/abs/2008.09656

The study also found that simply removing demographic features from a real-world algorithmic system like Lookalike audiences alone does not prevent biased or discriminatory outcomes.

This study highlights the challenges of eliminating bias in AI systems and recommends that advertisers that do not want biased outcomes should refrain from using targeting tools that rely on algorithmic systems like Lookalike Audiences.

Another 2019 study looked at ads and advertisers on Facebook at a global scale, based on a browser extension and data from 622 real-world Facebook users.[30]

The study found that a significant fraction of targeting strategies (20%) are either potentially invasive (e.g., make use of Personally Identifiable Information (PII) or attributes from third-party data brokers to target users), or are opaque (e.g., use the Lookalike audiences feature that lets Facebook decide whom to send the ad to based on a proprietary algorithm).

79% of ads were targeted using personal data that can directly identify an individual, such as their phone number or other identifiable information.

The study also confirmed that Lookalike audiences are vulnerable to discriminatory practices by advertisers. Almost one in ten ads used potentially sensitive categories such as politics, finance, health, legal and religion.

In 2020, researchers at Carnegie Mellon University analysed ads for employment, housing and credit that were included in Facebook's archive for political ads (sometimes by mistake).[31]

These were posed before and after the policy change in the US as a result of the settlement. The findings suggest widespread gender bias in credit ads, while housing and jobs were disproportionately shown to women.

04

Evidence of discrimination in Europe

Generally speaking, studies that find evidence for discrimination in online advertising in the US and other parts of the world suggest that similar discrimination also occurs in Europe. For instance, studies that found evidence for bias in ad optimisation on Facebook strongly suggest that similar bias is present in Europe.

A major difference between the US and Europe in particular is the different legal environment surrounding privacy, data protection and non-discrimination laws.

Facebook, for instance, has not implemented all changes the company was forced to make as a result of the March 2019 US settlement with civil rights advocates in Europe.

Only advertisers based in the United States or targeting the United States or Canada and running credit, housing or employment ads, must self-identify as a Special Ad category.[32] European users are not afforded the same safeguards by platforms when it comes to credit, housing or employment ads.

At the same time, the existence of the General Data Protection Regulation (GDPR) in Europe, particularly the definition and additional safeguards around special category data, mean that ad targeting in Europe looks very different than it does in the United States.

In the context of online marketing, advertisers typically need to rely on explicit consent as a legal basis for processing. This applies to special category data that has been collected from the data subject directly, as well as special category data that has been derived and inferred.

As a result, data brokers and social media platforms generally do not provide targeting criteria that allow advertisers to explicitly target people based on protected categories, such as ethnicity. In practice, however, advertisers often rely on known proxies such as interests to target ads based on special category data.

In 2017, for instance, the Dutch Data Protection Authority found that Facebook enabled advertisers to target people based on sensitive characteristics, such as "data relating to sexual preferences" without the explicit consent from users. [33]

In 2018, Facebook changed its data policy as a result – users are now given more extensive information about the ways in which their data is processed, but data processing is still taking place.[34]

A 2020 study by Cabañas et al. (2020). showed that 67% of global Facebook users are labelled with some potentially sensitive ad preferences, which may suggest political opinions, sexual orientation, personal health issues and other potentially sensitive attributes, including EU users.[35]

32 Facebook (n.d.). Discriminatory Processes [Facebook page]. Retrieved from https://www.facebook.com/ policies/ads/prohibited_content/discriminatory_practices

33 Autoriteit Persoonsgegevens (2017, May 16). Dutch Data Protection Authority: Facebook Violates Privacy Law. Retrieved from https://autoriteitpersoonsgegevens.nl/en/news/dutch-data-protection-authority-facebook-violates-privacy-law

34 Autoriteit Persoonsgegevens (2018, July 12). Facebook Changes Policy After Investigation by Dutch Data Protection Authority. Retrieved from https://autoriteitpersoonsgegevens.nl/en/news/facebook-changes-policy-after-investigation-dutch-data-protection-authority

35 Cabañas, J. G., Cuevas, Á., Arrate, A., & Cuevas, R. (2020). Does Facebook Use Sensitive Data for Advertising Purposes? Communications of the ACM, 64(1), pp. 62-69.

36 Stokel-Walker, C. (2019, August 24). Facebook's Ad Data May Put Millions of Gay People at Risk. Retrieved from https://www.newscientist.com/article/2214309-facebooks-ad-data-may-put-millions-of-gay-people-at-risk/#ixzz6o8MqifAk

37 Privacy International (2019, September 3). Privacy International Study Shows Your Mental Health is for Sale. Retrieved from https://privacyinternational.org/long-read/3194/privacy-international-investigation-your-mental-health-sale

> **Sensitive data about health is widely available to advertisers in Europe despite its theoretically stronger protection by GDPR.**

The authors therefore conclude that the GDPR has had "a negligible impact on Facebook regarding the use of sensitive ad preferences for commercial purposes."

Facebook has defended the policy of allowing advertisers to target people based on interests that may reveal special categories of personal data as follows: "the interest targeting options we allow in ads reflect people's interest in topics, not personal attributes [...] people can't discriminate by excluding interests such as homosexuality when they build an ad."[36]

Research by Privacy International (2019) into websites about health in France, Germany and the UK revealed that tracking for advertising is rampant, and often difficult, if not impossible, to reject.

This alone does not prove that users are discriminated based on health-related information in online advertising, but it means that sensitive data about health is widely available to advertisers in Europe despite its theoretically stronger protection by GDPR.[37]

Recently, discussion on discrimination in online advertising has focused on the role of special categories of personal data in Real Time Bidding (RTB). RTB is an auctioning process used to display programmatic advertising.

In its report on RTB, the UK Information Commissioner's Office (ICO) concludes that there is a widespread failure to protect personal data, including special categories of personal data, in a system that leaks the interest and online behaviour of Internet users, "millions of times a second".[38]

The ICO has argued that RTB participants need to rely on explicit consent, which does not correspond to the way in which consent is typically obtained in RTB processes.[39]

A 2020 study by AlgorithmWatch found evidence of discrimination through ad optimisation on both Google and Facebook for employment ads that were displayed in Germany, Poland, France, Spain and Switzerland.[40]

AlgorithmWatch bought job ads linking to real job offers on the portal Indeed for the following positions: machine learning developers, truck drivers, hairdressers, childcare workers, legal counsels and nurses.

A key finding of the report is that Facebook, and to a lesser extent Google, targeted the ads without asking for permission. For example, in Germany, an ad for truck drivers was shown on Facebook to 4,864 men but only to 386 women. An ad for childcare workers, which was running at exactly the same time, was shown to 6,456 women but only to 258 men.

---

38 Fix AdTech (2019, June 29). A Summary of the ICO Report on RTB – and What Happens Next. Retrieved from https://fixad.tech/a-summary-of-the-ico-report-on-rtb-and-what-happens-next/

39 ICO (n.d.). Special Category Data. Retrieved from https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/

40 Kayser-Bril, N. (2020, October 18). Automatisierte Diskriminierung: Facebook verwendet grobe Stereotypen, um die Anzeigenschaltung zu optimieren. Retrieved from https://algorithmwatch.org/story/automatisierte-diskriminierung-facebook-verwendet-grobe-stereotypen-um-die-anzeigenschaltung-zu-optimieren/

"A 2020 study by AlgorithmWatch found evidence of discrimination through ad optimisation on both Google and Facebook for employment ads that were displayed in Germany, Poland, France, Spain and Switzerland."

05

Protections against discrimination in online advertising

As evidence of discrimination through online advertising grows, the gaps in the protection in European legal frameworks become wider. Due to the often indirect and opaque nature of discriminatory advertising, it's likely that redress will be inaccessible under current laws.

Direct and indirect discrimination is already prohibited in many treaties and constitutions, including Article 14 of the European Convention on Human Rights, which states:

> **The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.**[41]

Similarity, EU non-discrimination law, in particular through the concept of indirect discrimination, prohibits many discriminatory effects of automated decision-making[42], including in online advertising.

In practice, however, enforcement is difficult, as those affected need to know that they have in fact been discriminated against.

As the Council of Europe has furthermore argued in their report on discrimination, artificial intelligence, and algorithmic decision-making, non-discrimination law has gaps that leave people unprotected from automated discrimination. One reason is that in practice, discrimination law places a high burden of proof on claimants.

Proving indirect discrimination requires an individual to provide evidence that, as a group, those sharing their protected characteristics are subject to different outcomes or impacts compared to those without this characteristic.

In the case of indirect discrimination, differential outcomes may be justified if the measure is necessary in pursuit of a legitimate aim.

Another reason is the concept of protected characteristics, which non-discrimination laws typically focus on. These gaps leave those who are affected by discrimination unprotected, for instance when individuals are unfairly subjected to differential treatment based on criteria that do not directly match prohibited discriminations under EU law (sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation).

The General Data Protection Regulation also offers a number of protections against automated discrimination in online advertising, specifically though the definition of profiling in Article 4(4), the definition of sensitive data under Article 9, and the principle of fairness in data processing.

Under the GDPR, stricter rules apply to the processing of special categories of personal data, which includes genetic and biometric data as well as information about a person's health, sex life, sexual orientation, racial or ethnic origin, political opinions, religious or philosophical beliefs, and trade union membership.

Guidance on special category data by the UK ICO reiterates a preference for obtaining explicit consent for the processing of special category data.

A popular loophole to avoid safeguards for special category data is to target people based on interests that reveal information about them that are special category data.

For instance, advertisers on Facebook cannot directly target LGBTQ-identifying people using targeting criteria that are provided by the platform, but they can target people with interests in LGBTQ issues, such as pride.

The use of this kind of proxy information for targeting ads at people allows advertisers to effectively circumvent the protections that GDPR is supposed to provide for special categories of personal data.

Profiling refers to the automated processing of data (personal and not) to derive, infer, predict or evaluate information about an individual (or group), in particular to analyse or predict an individual's identity, their attributes, interests or behaviour.[43]

We are yet to see complaints and legal cases that clarify how exactly rules on profiling and automated decision-making will be interpreted by regulators and the courts. On top of this, these provisions have always been narrowly defined.

They do not capture all forms of profiling or automated decision-making but are limited to decisions that are "based purely on automated decision-making", and those with "legal of similarly significant effects".

41 Council of Europe (1952). The European Convention on Human Rights. Strasbourg: Directorate of Information.

42 Zuiderveen Borgesius, F. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. Strasbourg: Directorate General of Democracy.

43 Kaltheuner, F., & Bietti, E. (2018). Data is Power: Towards Additional Guidance on Profiling and Automated Decision-Making in the GDPR. Journal of Information Rights, Policy and Practice, 2(2), pp. 1-17. http://doi. org/10.21039/irpandp.v2i2.45

06

Why discrimination in online advertising persists

There are a number of reasons why discrimination in online advertising persists, even though it is already prohibited under many European laws.

Individuals rarely know if discrimination has occurred

Online advertising is characterised by an overall lack of transparency. This is partially due to the number of companies involved, but also due to the fact that ad delivery is often highly automated.

The ways in which platforms explain how individuals are targeted are often incomplete or overly simplistic. The way Facebook's ad explanations appear to be built, for instance, "may allow malicious advertisers to easily obfuscate ad explanations from ad campaigns that are discriminatory or that target privacy-sensitive attributes".[44]

A 2018 study by Upturn showed that Facebook's ad transparency interface does not include an effective way for the public to make sense of the millions of ads running on its platform at any given time, and does not allow users to understand how an ad is targeted as well as the size and nature of the audience it reaches.[45] This is echoed by research conducted by Privacy International (2020).[46]

Challenges in exercising data rights

As a direct consequence of the overall lack of transparency in online advertising, it is incredibly challenging for individuals to exercise their data rights.

A study by Ausloos, Mahieu, and Veale (2019), for instance, showed that the information which is typically provided by platforms and ad networks to explain how ads are targeted are insufficient for individuals to understand whether they have been profiled in ways that are discriminatory. This would require information about the alternatives that the individual could have been categorised as.[47]

Machine learning and AI is transforming online advertising

The evolution of techniques used in online advertising is another reason why discrimination in online advertising persists – and is likely going to increase in the future.

As Kingaby (2020) argues, "advertising stands at the brink of widespread adoption of AI, which risks ingraining excessive data collection habits, inadvertent discrimination, and decision making based around metrics which consider only advertising 'performance' in its narrowest sense."[48]

---

44 Andreou, A., Venkatadri, G., Goga, O., Gummadi, K., Loiseau, P., & Mislove, A. (2018). Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations. In NDSS 2018-Network and Distributed System Security Symposium. San Diego, California, USA. Retrieved from https://lig-membres. imag.fr/gogao/papers/fb_ad_transparency_NDSS2018.pdf

45 Rieke, A., & Bogen, M. (2018). Leveling the Platform: Real Transparency for Paid Messages on Facebook. Retrieved from https://www.teamupturn.org/reports/2018/facebook-ads/

46 Privacy International (2020, September 24). Facebook Response on Advertising: A Failure to Acknowledge Responsibility. Retrieved from https://privacyinternational.org/news-analysis/4171/facebook-response-advertising-failure-acknowledge-responsibility

47 Ausloos, J., Mahieu, R., & Veale, M. (2019). Getting Data Subject Rights Right. Journal of Intellectual Property, Information Technology and Electronic Commerce Law, 10(3), pp. 283-309.

48 Kingaby, H. (2020). AI and Advertising: A Consumer Perspective. Retrieved from https://789468a2-16c4-4e12-9cd3-063113f8ed96.filesusr.com/ugd/435e8c_3f6555abb25641be8b764f5093f1dd4f.pdf

"As a direct consequence of the overall lack of transparency in online advertising, it is incredibly challenging for individuals to exercise their data rights."

07

Conclusion and recommendations

As this report has shown, discrimination in online advertising is rampant, and there are no easy solutions. Simply banning platforms or ad networks from allowing advertisers to target groups based on protected categories does not eliminate discrimination, as this can be circumvented, and discrimination is not always caused by the deliberate targeting of a protected group.

A main challenge in tackling discrimination is that the online advertising system is complex, opaque and highly automated.

As a result, individuals who are targeted by ads, as well as advertisers who run ads, do not necessarily know how or why an ad has been targeted in any specific way. This makes it extraordinarily difficult for individuals to know that they have been discriminated against, while it is challenging for researchers or regulatory authorities to prove if and how discrimination has occurred.

This combined with the wide range of risks and harms associated with online advertising as we know it today mean that the entire online advertising system is in dire need for regulatory reform.

Recommendation 1: Strengthen regulatory authorities

In order to do their jobs, not merely Data Protection Authorities (DPAs), but also other regulatory bodies, such as consumer protection authorities, equality bodies and human rights monitoring bodies need systematic funding. Those actors need to be able to recruit and maintain staff with the necessary technical expertise.

Recommendation 2: Full investigation into discrimination in online advertising in Europe

There is evidence to suggest that discrimination in online advertising is widespread in Europe.

In order to back up that evidence with additional data, authorities should collaborate on an urgent investigation of discrimination in online advertising in Europe, specifically around the use of "interests" as proxies for sensitive categories.

Regulatory authorities in Europe should also collaborate to enforce and investigate how special category data are used without the explicit consent of individuals throughout the online advertising ecosystem, specifically in RTB, but also in other forms of online advertising.

Recommendation 3: Update discrimination law

Discrimination laws need to be fit for purpose to protect people from new and changing forms of discrimination.

This applies to automated discrimination more broadly, but also to discrimination in relation to targeted online advertising. As the Council of Europe has explained in a report on Discrimination, artificial intelligence, and algorithmic decision-making:

> **AI also opens the way for new types of unfair differentiation (some might say discrimination) that escape current laws. Most non-discrimination statutes apply only to discrimination on the basis of protected characteristics, such as skin colour.**

Such statutes do not apply if an AI system invents new classes, which do not correlate with protected characteristics, to differentiate between people. Such differentiation could still be unfair, however, for instance when it reinforces social inequality.[49]

[49] Council of Europe (2018). Discrimination, Artificial Intelligence, and Algorithmic Decision-Making. Retrieved from https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73

"Authorities should collaborate on an urgent investigation of discrimination in online advertising in Europe, specifically around the use of "interests" as proxies for sensitive categories."

Recommendation 4: Update data protection law and ensure effective enforcement

Protections for automated decision making under the GDPR are currently limited to decisions that have a legal or similarly significant effects, and that are based on solely automated processing.

While additional guidance has clarified that human intervention must be meaningful and cannot be a "token gesture", this still leaves much room for interpretation.

A strengthening of these provisions would give individuals more rights over automated decision making, including profiling, which has implications for online advertising more broadly.

Likewise, enforcement of data protection laws should clarify the status of data that is inferred, derived and predicted.

While not all inferences are personal data, the moment such inferred data allow for the direct or indirect identification of an individual, they clearly fall under the definition of personal data.

This needs to be reflected in enforcement decisions, specifically with regards to the ways in which data brokers, AdTech companies and platforms use profiling for advertising purposes.

Further guidance should clarify that advertisers cannot rely on people's disclosed or inferred interests to target people based on special category data indirectly.

Recommendation 5: Adopt a strong e-Privacy Regulation

The EDRi network has been advocating for a strong e-Privacy legislation since before it was proposed.

---

50 European Digital Rights (2017). EDRi's Position on the Proposal of an e-Privacy Regulation. Retrieved from https://edri.org/files/epd-revision/ePR_EDRi_position_20170309.pdf

The Regulation is aimed at ensuring privacy and confidentiality of our electronic communications, by complementing and particularising the rules introduced by the GDPR.

Specifically, as EDRi has argued on numerous occasions, the legislation needs to ensure that bulk data retention remains banned in law and practice, that privacy by design and by default remains at the core of the Regulation, and that it must allow people to "use a service without being tracked by third parties, especially if the user depends on, and has no real alternative to, this service."[50]

A strong e-Privacy reform would put users back in control of their communication data. This has indirect consequences for discrimination in online advertising as well as increasing the overall transparency of the online advertising system.

Recommendation 6: A sweeping reform of online advertising

The above steps will help to tackle some of the harms and risks to individuals, markets and societies that are associated with online advertising as we know it today.

However, in order to truly tame a surveillance-driven advertising business model, a sweeping reform of the industry is needed. Regardless of the specifics of the reform, any new or updated regulation will need to work towards accomplishing the following goals:

Force greater transparency and accountability on the online advertising system

Greater transparency and accountability are a precondition for tackling discrimination in online advertising.

It is currently virtually impossible for users to understand why and how they are targeted by an ad, and which data, or targeting criteria were used to target them.

This makes it difficult to even realise or notice that discrimination has occurred.

> **From a fundamental rights perspective, a key goal of any reform of the online advertising system needs to limit and reduce the overall amount of data in the system.**

The overall lack of accountability and transparency in the online advertising ecosystem means that researchers who study discrimination, as well as regulatory authorities that want to take action against discrimination in online advertising need to go to extraordinary lengths to find evidence.

Limit and reduce the overall amount of data in the system

A key concern of online advertising in its current form is the amount of personal data that is collected and shared.

From a fundamental rights perspective, a key goal of any reform of the online advertising system needs to limit and reduce the overall amount of data in the system. This also has indirect consequences for discrimination in online advertising.

Tackle market dominance

The online advertising market is dominated by Google's parent company Alphabet Inc. and Facebook.

Tackling market dominance would prevent those companies from de facto imposing their terms and conditions in a take-it-or-leave-it approach.

Ban targeting techniques that are inherently opaque

As this report has shown, some targeting techniques are inherently opaque, meaning that it is often impossible for advertisers to avoid discrimination, even if they deliberately decide to target their ads based on neutral criteria.

Ad optimisation falls into this category, so do targeting tools like Lookalike Audiences.

Recommendation 7: Regulation on AI needs to cover discrimination in advertising

In order to effectively protect people from discrimination in online advertising, European regulation on AI needs to include advertising.

The Commission's draft White Paper on AI, for instance, relied on a particularity narrow definition of risk.

From AI-driven consumer products, data brokers, and the online marketing and Ad-Tech industry, to the personalisation and recommendation systems that fuel social media platforms, this definition left individuals and society at large unprotected from fundamental rights violations in the very sectors that have seen some of the earliest and most widespread adoption of AI.

It is also important to note that risk is unevenly distributed within society. For certain groups of people any application of AI, not just those considered "high-risk", comes with an inherent risk of discrimination and exclusion.

Furthermore, mandatory legal requirements cannot be limited to prohibited discrimination.

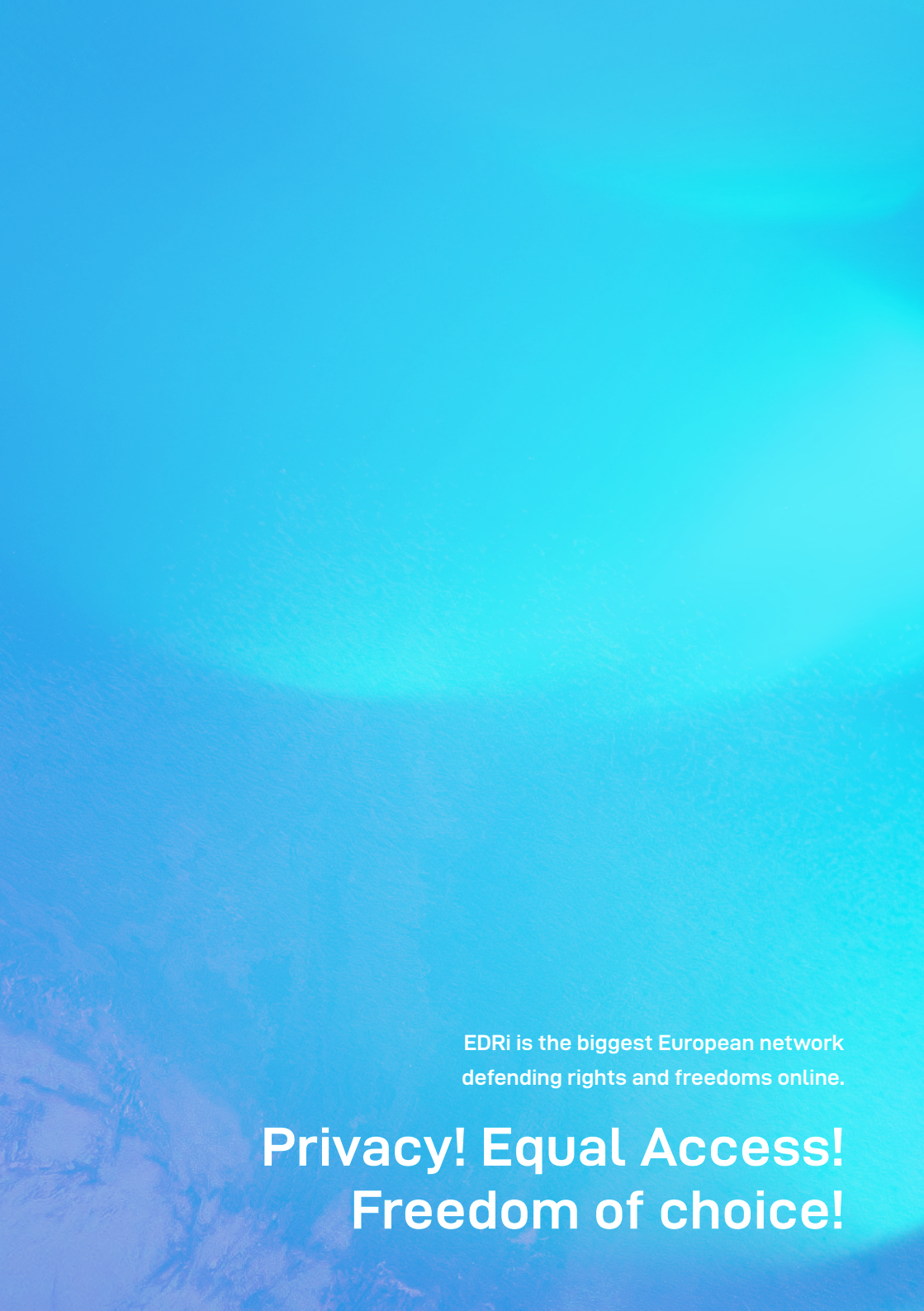
As this report has shown, existing definitions of prohibited discrimination fail to cover all instances of harmful automated discrimination by AI systems, for instance in advertising.

Mass surveillance. Discriminatory Algorithms. Profit Over Communities.

Companies and governments increasingly restrict our freedoms.

—

DONATE NOW:
https://edri.org/take-action/donate


EDRi is the biggest European network
defending rights and freedoms online.
Privacy! Equal Access!
Freedom of choice!

Press enquiries
press@edri.org

Brussels office
brussels@edri.org

Phone number
+32 2 274 25 70

Visit us
Rue Belliard 12
1040 Brussels
Belgium

Follow us
Twitter
Facebook
LinkedIn
Youtube

European Digital Rights (EDRi) is the biggest European network defending rights and freedoms online. We promote, protect and uphold human rights and the rule of law in the digital environment, including the right to privacy, data protection, freedom of expression and information.

www.edri.org