Increasing fairness in targeted advertising

The risk of gender stereotyping by job ad algorithms

Outcomes of the winter 2021 Clinic of the *Ethics of Digitalisation* project
hosted by the Alexander von Humboldt Institute for Internet and Society

EDITED BY

Nadine Birner, Shlomi Hod, Matthias C. Kettemann, Alexander Pirang, and Friederike Stock

WITH CONTRIBUTIONS FROM

Ezgi Eren, Lukas Hondrich, Linus Huang, Basileal Imana, Matthias C. Kettemann, Joanne Kuai, Marcela Mattiuzzo, Alexander Pirang, Ana Pop Stefanija, Sylvi Rzepka, Marie-Therese Sekwenz, Zora Siebert, Sarah Stapel, and Franka Weckner

Increasing fairness in targeted advertising: The risk of gender stereotyping by job ad algorithms

Outcomes of the winter 2021 Clinic of the *Ethics of Digitalisation* project

Who gets to see what on the internet? And who decides why? These are among the most crucial questions regarding online communication spaces – and they especially apply to job advertising online. Targeted advertising on online platforms offers advertisers the chance to deliver ads to carefully selected audiences. Yet, optimizing job ads for relevance also carries risks – from problematic gender stereotyping to potential algorithmic discrimination. The winter 2021 Clinic *Increasing Fairness in Targeted Advertising: The Risk of Gender Stereotyping by Job Ad Algorithms* examined the ethical implications of targeted advertising, with a view to developing feasible, fairness-oriented solutions.

The virtual Clinic brought together twelve fellows from six continents and eight disciplines. During two intense weeks in February 2021, they participated in an interdisciplinary solution-oriented process facilitated by a project team at the Alexander von Humboldt Institute for Internet and Society. The fellows also had the chance to learn from and engage with a number of leading experts on targeted advertising, who joined the Clinic for thought-provoking spark sessions.

The virtual Clinic is part of the research project *The Ethics of Digitalisation – From Principles to Practices*, which aims to develop viable answers to challenges at the intersection of ethics and digitalisation. Innovative formats facilitate interdisciplinary scientific work and a global dialogue on application-, and practice-oriented questions and aim to achieve outputs of high social relevance and impact. The project, led by the Global Network of Internet & Society Research Centers, is conducted under the patronage of the German Federal President Frank-Walter Steinmeier and is supported by Stiftung Mercator. In addition to the Alexander von Humboldt Institute for Internet and Society, the main project partners are the Berkman Klein Center at Harvard University, the Digital Asia Hub, and the Leibniz Institute for Media Research I Hans-Bredow-Institut.

The objective of the Clinic was to produce actionable outputs that contribute to improving fairness in targeted job advertising. To this end, the fellows developed three sets of guidelines – this resulting document – that cover the whole targeted advertising spectrum. The guidelines provide concrete recommendations for platform companies and online advertisers, and are also of interest to policymakers.

Table of Contents

Contributors

Fellows

Ezgi Eren – Affiliation: Student (LLM) at The University of Edinburgh, Great Britain

Lukas Hondrich – Affiliation: Researcher at AlgorithmWatch, Germany

Linus Huang – Affiliation: Postdoctoral Fellow at the Society of Fellows in the Humanities, University of Hong Kong, Hong Kong

Basileal Imana – Affiliation: PhD student at the University of Southern California; USA

Joanne Kuai – Affiliation: PhD student at Karlstad University, Sweden

Marcela Mattiuzzo – Affiliation: PhD Candidate at the University of São Paulo, Brazil

Ana Pop Stefanija – Affiliation: PhD researcher at imec-SMIT, Vrije Universiteit Brussel, Belgium

Sylvi Rzepka – Affiliation: Postdoctoral Researcher at the University of Potsdam, Germany

Marie-Therese Sekwenz – Affiliation: Student at University of Vienna, Vienna, Austria

Zora Siebert – Affiliation: Brussels Head of EU Policy Programme at Heinrich Böll Foundation, European Union, Belgium

Sarah Stapel – Affiliation: Student (LLM) at the University of Amsterdam, Netherlands

Franka Weckner – Affiliation: Student at the University of Heidelberg

Project team

Nadine Birner – Coordinator: The Ethics of Digitalisation | NoC

Shlomi Hod – PhD Student at Boston University | Associated Researcher at HIIG

Matthias C. Kettemann – Research Group Leader at HIIG

Alexander Pirang – Researcher at HIIG: AI & Society Lab

Wolfgang Schulz – Research Director at HIIG

Tom Sühr – Student Assistant at HIIG: Platform Governance and Copyright (H2020)

Friederike Stock – Student Assistant at HIIG: The Ethics of Digitalisation | NoC

Experts / Sparks

John Byers – Professor of Computer Science at Boston University, USA

Elisabeth Greif – Associate Professor at the Institute of Legal Gender Studies at the Johannes Kepler University of Linz, Austria

Nicolas Kayser-Bril – Reporter at AlgorithmWatch, Germany

Aleksandra Korolova – Assistant Professor of Computer Science at USC, USA

Nicole Shephard – Freelance Researcher and Consultant

Keywords

job ads, targeted advertising, algorithms, algorithmic discrimination

License

MATTHIAS C. KETTEMANN, ALEXANDER PIRANG

Setting the stage: what we know and what we don't know about targeted advertising

Put yourself in the shoes of a young professional wishing to land a job in the tech industry. Your friend, who has a similar background, recently found a good position after learning about it through a tech company's advert on social media. You keep looking for similar job ads when going through your news feed, but you seem to have no luck. The likely difference between the two of you? Your gender – deemed relevant by an algorithm.

This hypothetical scenario may well happen in real life. A study from 2018 found that younger women were significantly less likely to be exposed to STEM[1] ads on Facebook than men (Lambrecht & Tucker, 2018). Similar cases of discrimination by online platforms' targeted advertising are well-documented (Kofman & Tobin, 2019). This raises urgent questions: How can we best avoid harmful societal consequences of seemingly well-intentioned targeted advertising strategies? Do we need less targeted systems for ad deliveries or ones that grant their target groups a meaningful choice regarding the kinds of job ads they are interested in? Thus far, there is no clear-cut solution on how to ensure that algorithmic models optimized for relevance do not reinforce gender stereotypes in the labor market.

The targeted advertising process

In order to understand the extent of these risks, we need to examine targeted advertising more closely. The underlying process is complex and – despite fueling a multi-billion ad tech industry – highly opaque. In the following, we will take a closer look at job advertising on online platforms.

> **Ads on platforms versus banner ads**. There is a difference between ads that you see on online platforms, and banner ads that you see elsewhere online (e.g. mobile apps or news sites) (Iwanska, 2020). In the Clinic, we focused on the former.
>
> **Behavioral versus contextual ads**. Behavioral ads are personalized and targeted to you based on extensive profiling; on the other hand, contextual ads are placed next to matching pieces of content and do not require knowledge about you or your past browsing behavior (Iwanska, 2020). In the following, we will mainly discuss behavioral advertising.

---

[1] Science, Technology, Engineering and Math.

Broadly speaking, targeted advertising can be subdivided into three main aspects (see illustration):

— **ad targeting by advertisers**, which mainly includes creating the ad, selecting the target audience, and choosing a bidding strategy;

— **ad delivery by platforms**, which mainly includes carrying out the automated ad action and optimizing ads for relevance; and

— **ad display to users**, which includes ads placement and any transparency or control tools afforded to users.
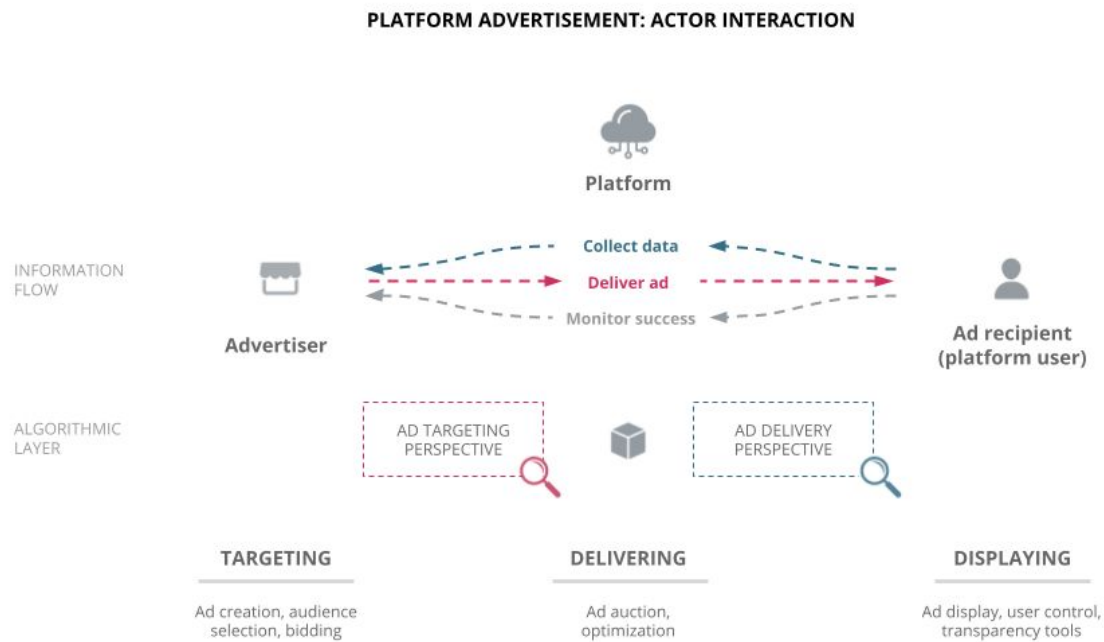
PLATFORM ADVERTISEMENT: ACTOR INTERACTION

Illustration 1. Overview of actor interactions, information flows, and stages of the targeted advertising process

Ad targeting

PLATFORM ADVERTISEMENT: AD TARGETING PERSPECTIVE

Ad created with
targeted audience
and bid

Select all users
targeted by
advertiser

Auction ad
According to
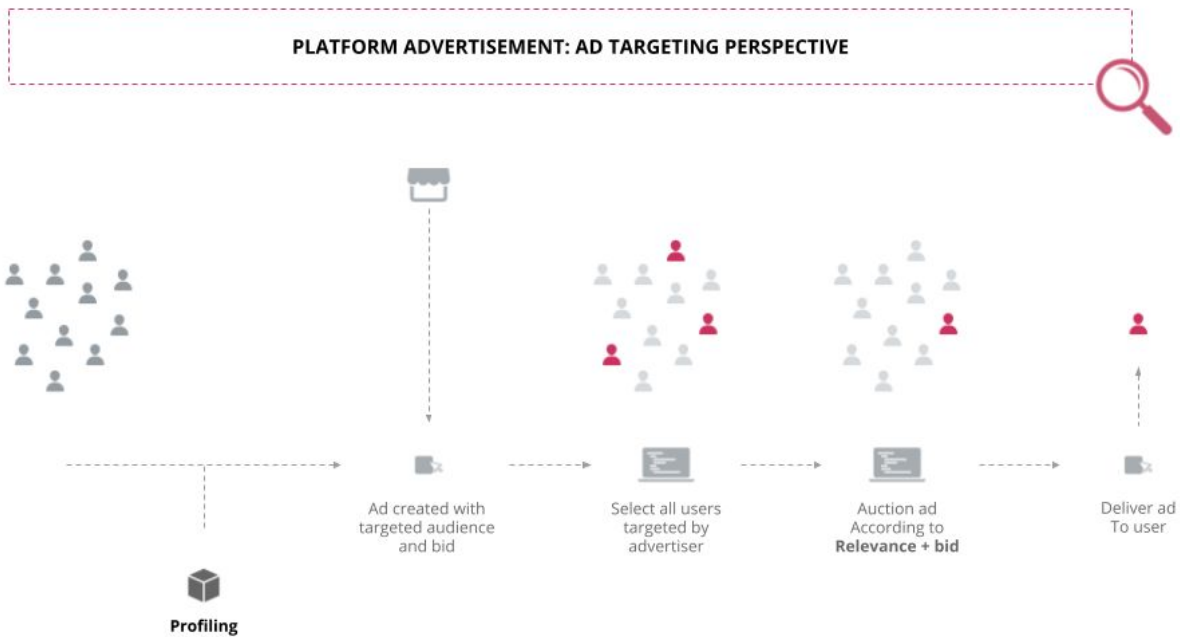**Relevance + bid**

Deliver ad
To user

Profiling

Illustration 2. Overview of the targeted advertising process from the ad targeting perspective

Platforms offer advertisers a range of options that ensure their adverts are delivered to tailormade audiences. In the following, we will take a closer look at job advertising on online platforms.

Audience selection (based on Facebook's audience selection options)

- **Demographics and attributes:** Advertisers can select audiences based on demographics (e.g. age or location), as well as profile information, activity on the platform, and third-party data (Ali et al., 2019).

- **Personal information:** Advertisers may also specify the exact users they wish to target by uploading users' personally identifiable information or by making use of tracking tools on third-party sites (Ali et al., 2019).

- **Similar users:** Advertisers are also allowed to create audiences similar to previously selected users. To this end, algorithmic classifiers identify the common qualities of the users in those previous audiences and create new audiences based on those shared qualities (Ali et al., 2019).

In addition to selecting an audience, advertisers also need to choose an objective and a bidding strategy when placing an ad. Facebook, for instance, allows advertisers to optimize for "awareness" (i.e. views), "considerations" (i.e. clicks and engagements), and "conversion" (i.e. sales generated by clicking through the ad) (Facebook for Business). A bidding strategy may include the start and end time of an ad campaign and budget caps, based on which the platform places bids in ad auctions on the advertisers' behalf (Ali et

al., 2019) Platforms then deliver ads based on these targeting parameters and allow advertisers to monitor how their ad campaign is performing.

In light of the variety of targeting options, researchers have voiced concerns about potentially discriminatory targeting choices, which may exclude marginalized user groups from receiving job or housing ads (Speicher et al., 2018). Although discrimination based on certain protected categories such as gender or race is prohibited in many jurisdictions, and even though platforms such as Google and Facebook restrict sensitive targeting features in sectors like employment and housing (Sandberg, 2019), this problem still seems to persist due to problematic proxy categories (Benner et al., 2019; Kofman & Tobin, 2019).

To conclude, discrimination in targeted job advertising may occur where ad targeting choices are discriminatory. This is hardly surprising; less intuitive, however, are study results showing that ad delivery can still be skewed even where advertisers are careful not to exclude any kind of user group from their ad campaign (Ali et al., 2019). To address this question, we need to take a closer look at how ads are delivered.
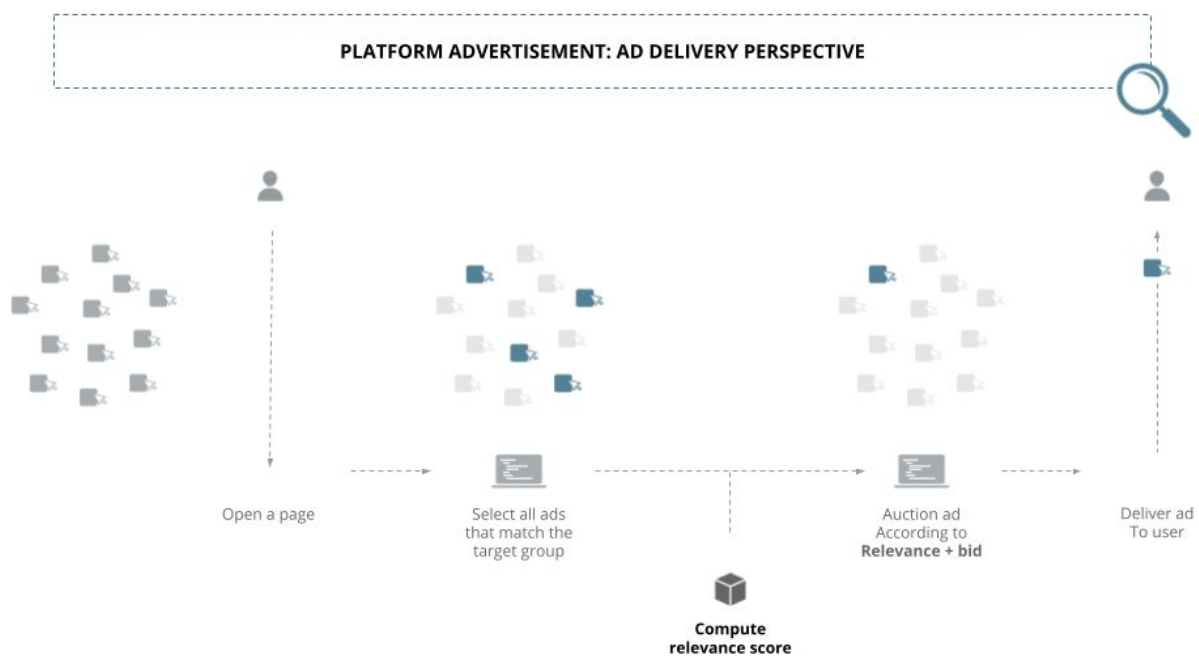
Ad delivery

Illustration 3. Overview of the targeted advertising process from the ad delivery perspective

What happens when you go on a platform? In the milliseconds it takes a site to load, algorithms first select all the live ads whose audience parameters fit your profile. Next, an automated ad auction takes place, much like a traditional auction, which determines which ad you are going to see. The higher the bids, the more competitive the ads. Still, platforms may not always pick the highest bids; in order to

maximize relevance, platforms commonly compute a relevance score for individual ads (i.e. estimating how well an ad will likely perform), thus occasionally allowing ads with lower bids to win over those with higher bids (Ali et al., 2019).

In sum, ad delivery is a second – extremely opaque – round of optimization, which platforms carry out without the advertisers' involvement. This opacity is concerning, as studies reveal that ad delivery may also cause discriminatory outcomes of online advertising. In fact, two separate causes appear to be at play:

First, **budget effects can cause skewed ad delivery**. Researchers from London Business School and MIT, who ran real job ads on Facebook, observed that their ads were predominantly shown to men even though women had consistently higher engagement rates (Lambrecht & Tucker, 2018). The likely reason? Higher engagement rates lead to women being more costly to advertise to, and advertisers with lower budgets are thus more likely to lose auctions for women in comparison to auctions for men (Ali et al., 2019).

Second, **the content of ads can also cause skewed ad delivery**. Researchers from Northeastern University, USC, and Upturn also created and ran actual job ads on Facebook and demonstrated that the company's ad-serving algorithms, when optimizing for clicks, delivered "to vastly different racial and gender audiences depending on the [ad content] alone." (Ali et al., 2019) Their results showed that gender stereotyping in job ad delivery is very likely to occur on Facebook. For instance, their ads for cashier positions in supermarkets reached an 85 % female audience, whereas ads for jobs in the lumber industry – with the same ad targeting options – reached a 72 % male audience (Ali et al., 2019).

Ad display

Users commonly have little means to look behind the scenes of targeted advertising and understand why they are seeing certain ads and why they aren't seeing others. Existing transparency initiatives by platforms still far short of providing users meaningful transparency (Andreou et al., 2018). The proposed Digital Services Act[2] imposes online advertising transparency obligations on online platforms, but these provisions have yet to become law.[3]

In many jurisdictions, anti-discrimination legislation in principle allows users to bring action against discriminatory ad practices. In practice, however, enforcement is difficult, as the burden of proof is on users (Kayser-Bril, 2020). They need to prove discrimination on the grounds of what they are *not* seeing – a seemingly impossible feat.

In sum, users face significant obstacles when trying to obtain information about targeted ads and when seeking legal redress.

---

[2] European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. 2020 (hereinafter DSA Proposal).

[3] Article 24 DSA Proposal and, pertaining to very large online platforms, Article 30 DSA Proposal.

Why guidelines?

The results of this Clinic have been turned into guidelines. Guidelines are important tools to help shape how AI is used, especially in ad targeting, delivery and display. Historically, the notion of guidelines is associated with lines for cutting things into shape, i.e. showing *how tools* (knives) can be used. Later "guidelines" were used to make sure hot-air balloons did not float away, that is, they were used to keep something valuable tethered to earth foundation. The guidelines the Clinic fellows have developed fulfill both purposes: they are meant to inform how the tools of the ad trade should be used and give direction. At the same time, they connect the ad ecosystem to the normative, ethically informed debate on fairness in advertising. Formulated in a way that makes them directly implementable, the guidelines can serve to form and inform the debate on how to make sure job ads are delivered fairly.

SPOTLIGHT: AD TARGETING

LUKAS HONDRICH, MARCELA MATTIUZZO, ANA POP STEFANIJA, ZORA SIEBERT
Guidelines: Improving fairness in ad targeting

To mitigate stereotyping in advertising, leading to discriminating outcomes in behavioural targeted advertising, we are focusing on the process of **audience selection**. Although different from platform to platform, the audience selection process is based on characteristics, which are themselves inferences, a result of datafication of user behaviour, often merging online and offline behavioral data. Audience selection can originate from:

— *Data holders*, including (i) advertisers, (ii) platforms (both from user behaviour on and off the platform — which is collected e.g. via a pixel) and (iii) third parties (such as advertisers themselves, including data brokers, apps, ad networks etc.) and
— *Data subjects* (volunteered data from users, for example when setting up an account).

Audience selection is one of the more opaque processes where criteria selection relies on inputs resulting from machine learning workings of proprietary software. It holds potential risks of stereotyping and discrimination that can be mitigated by introducing technical, infrastructural and legal elements at a system and inference level. By modifying the existing infrastructure of the advertising platforms, the overburdening of the platforms can be avoided and will not force them to build a system from scratch.

Added value

Following these guidelines will avoid unintended, undesirable consequences like stereotyping and discrimination based on gender, age, ethnicity or other sensitive categories in job ads. A higher level of transparency will be achieved for:

— advertisers, who will see which audience they actually reach;
— oversight bodies, who can check if job ads are delivered to the audience based on legal categories;
— independent researchers and platform users, who can access the ad database as well.

With both platforms and advertisers in mind, we propose two solutions: Legality by default and a transparency loop (see illustration below).
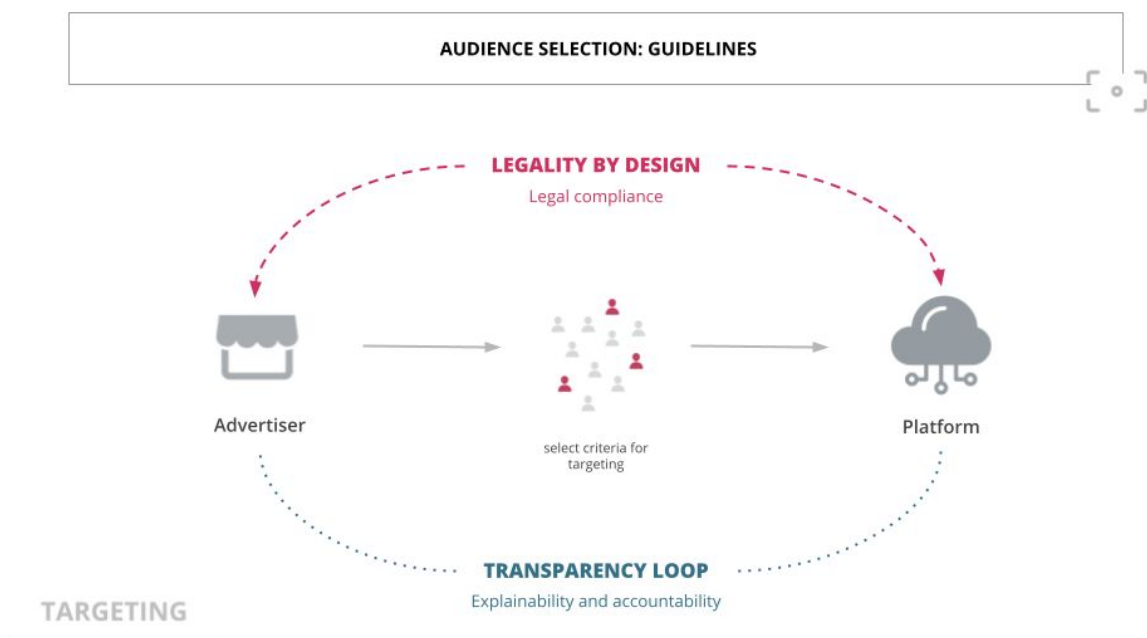
Illustration 4. The two guiding principles, their aims and relation with the main actors in the process of audience selection

With this twofold approach, we aim to enable audience selection that is not discriminatory in an illicit/abusive way. Full transparency will also lead to more legal compliance and to a more trustworthy ad delivery system for users.

Legality by default

Platforms will provide advertisers an explanation while they select audiences for the ad on the platform. Easily understandable icons will tell advertisers what the audience will be selected based on (see illustration below). Advertisers then confirm if they want to go ahead or change the criteria. This feedback loop will educate advertisers and empower them to make better informed decisions. Legality by default is *jurisdiction-specific*. Platforms and advertisers alike should be mindful that each country/jurisdiction has specific rules on what is considered discriminatory.

CRITERIA AVAILABLE FOR AUDIENCE SELECTION	EVALUATION OF CURRENT LEGAL PROTECTION FOR JOB ADS	DESIRED PLATFORM DESIGN
Gender **Ethnic group**	Protected in several jurisdictions	Clear information about legal prohibition
Age	Protected in some jurisdictions	
Interest **Geolocation**	Not protected in most jurisdictions	

Table 1. Criteria for how to select audience based on legal restrictions

Additionally, legality by default aims to contribute towards *awareness-enhancing*. This means that clear and visible information should be provided for advertisers during audience selection regarding whether the characteristic they choose to target for is protected by law (see illustration below).

Illustration 5. Example of interface interaction

Transparency loop

Platforms and/or advertisers can be made responsible if the outcome of the delivered targeted ads is discriminatory. To avoid legal and reputational damage, advertisers and platforms will pay due diligence in trying not to stereotype ads. The publicly accessible ad database where job ad targeting results and targeting input parameters are published will guarantee compliance and is easy to implement as similar ad databases e.g. for political ads already exist. This guideline is envisioned to be achieved via incorporating the audience selection into the dashboard itself (trial audience selection) and the creation of a job ads library.

Running a *trial* audience selection

Here advertisers can estimate what effect choosing specific criteria will have on the targeted audience. The information is visualized as distributions over sensitive criteria like gender, age and ethnicity (see illustration below). Estimating a targeting audience is no trivial task because it involves uncertainty about critical variables, e.g. what other ads will be run at the same time. However, for large platforms with a wealth of historical data and sufficient engineering capacity, reaching a satisfactory accuracy should be feasible. Giving advertisers the possibility to detect unwanted skews, biases and discrimination is a prerequisite for holding them accountable and will help define responsibilities between advertisers and platforms.
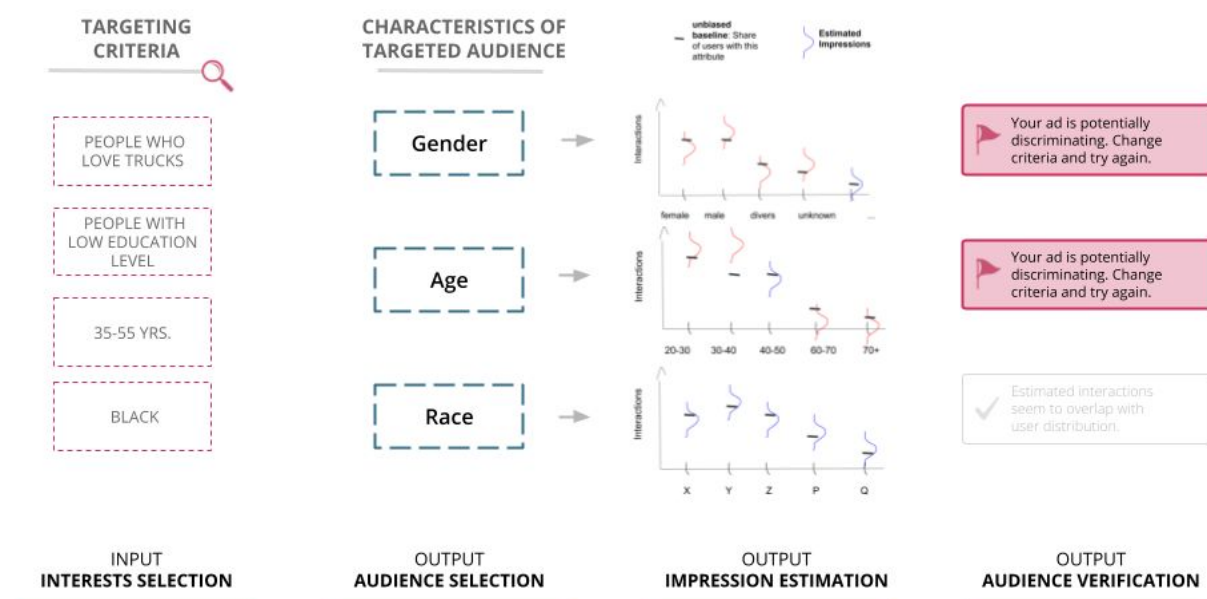

Illustration 6. The Audience selection dashboard, communicating possible bias resulting from the selection of particular attributes of the audience

Job ads library

The biggest platforms (e.g. those with > 1 million monthly users), including search, general and professional job platforms, will have to create and regularly update a "job ads" library (see illustration below). This will mitigate the potentially adverse impact of targeted advertising on individuals. Focusing on the biggest platforms ensures that they have the (financial & technical) means to comply with such

rules, and does not create market barriers for small and medium platform companies. A few platforms (e.g. Facebook) already have similar mechanisms, enabling both API and interface access for interested parties. Therefore, adding a separate library will be somewhat easier to implement.
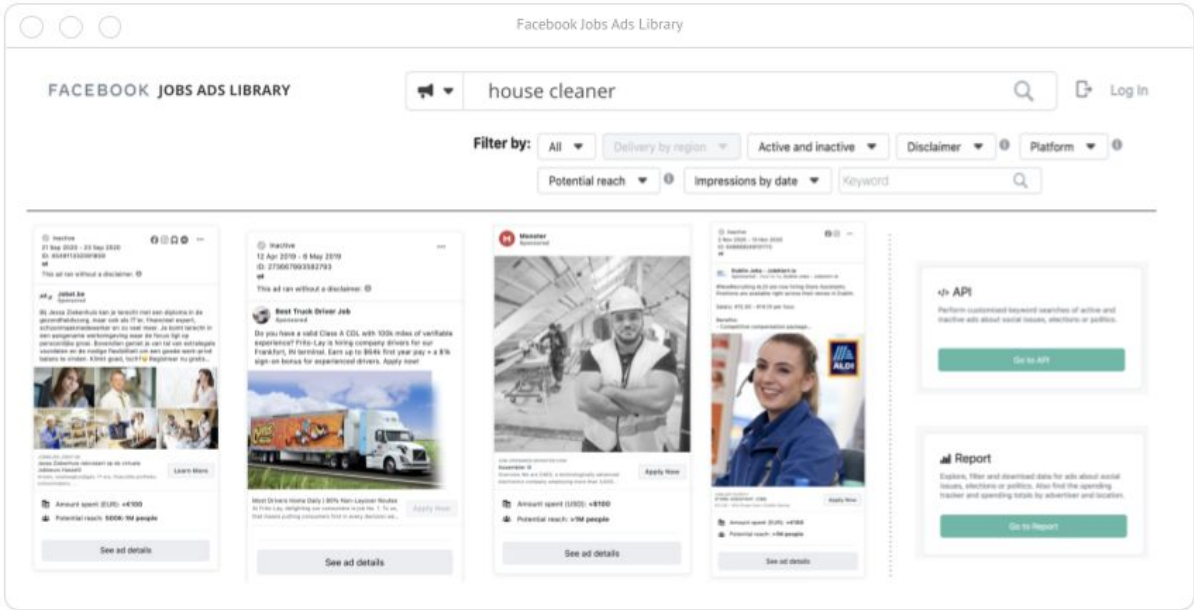

Illustration 7. Landing page of Job Ads library, on the example of existing Facebook Ad library

The job ads library will enable both transparency and accountability. *Transparency* because it will help to investigate potentially biased and discriminatory outputs, the number of people that have been potentially affected, and the concrete inputs that lead to particular outputs (see illustration below).
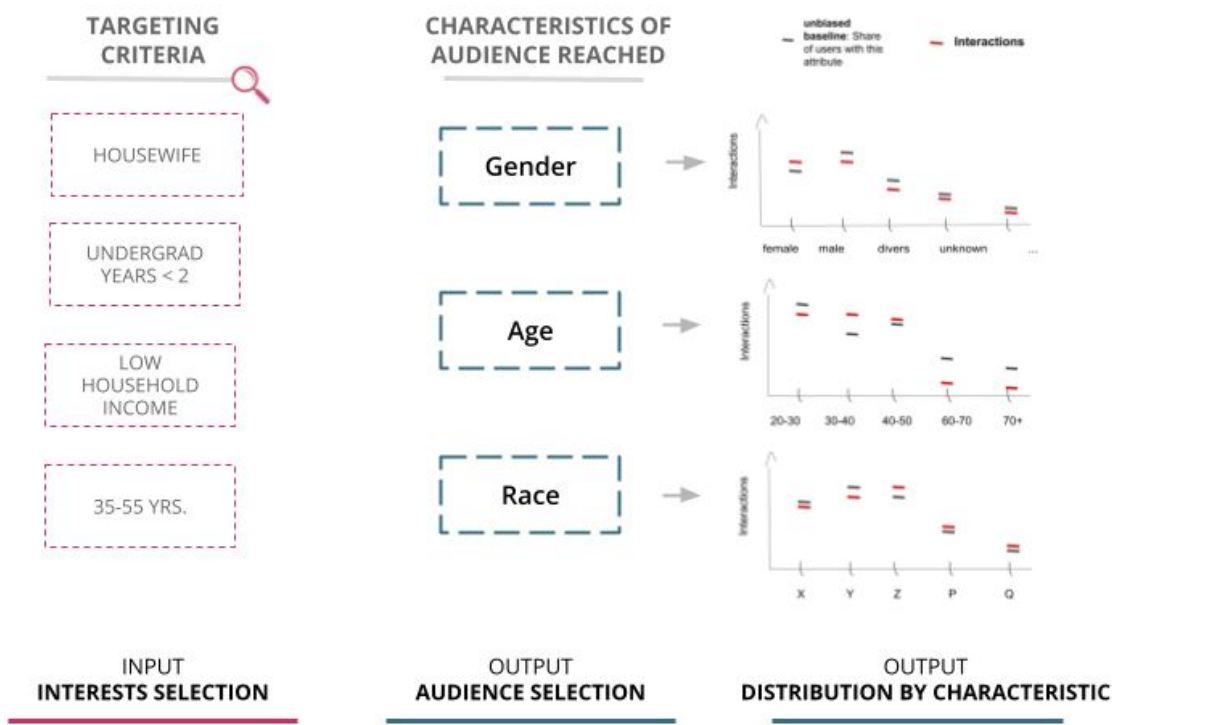

Illustration 8. Visual explanation of the inputs and outputs of a particular ad

The library also contributes towards increased *explainability*. Moreover, this library will enable *accountability*. When independent third parties inspect the system based on API/interface access, potential issues can be detected and mitigated. Therefore, the transparency loop will contribute towards fairness-oriented technology (AI/ML) design and regulatory guidelines.

SPOTLIGHT: AD DELIVERY

BASILEAL IMANA, JOANNE KUAI, SARAH STAPEL, FRANKA WECKNER

Guidelines: Improving fairness in ad delivery

The following guidelines are created for platforms that aim to create a user-centered system of ad delivery. The goal is to establish a system of delivering advertisements in a manner that protects user experience, by promising customizable and relevant advertisements while protecting user privacy and increasing user self-determination. By placing the user on the center stage, these guidelines provide a fair solution to platform advertising that is beneficial to the users, the advertisers, and the platform.

The importance of a user-centered approach

Our proposal, as explored in the following sections, promotes a system of ad delivery that is designed by the user. Users are often unaware of how their data is being collected and analyzed to determine ad delivery standards. By creating their own online advertising profiles, users are granted an online right to self-determination in relation to how they manage and process their data. In the current models of ad delivery, it is often the case that platforms optimize on the basis of what they think is relevant for the user. Our approach reverses this top-down approach by instead allowing the user to decide on the information they are targeted with. Users can hereby choose what information about their profiles should and should not factor into the decision-making process.

Aside from allowing the user to be in charge of her/his/their advertising profiles, such an approach also allows for a more effective avoidance of biased and discriminatory practices in advertising. Targeted advertising can be discriminatory when ads are delivered on the grounds of information that is inferred about the user. A user-centered approach prevents these inferences by limiting the data collection and analysis to the criteria selected by the user. These guidelines propose a delivery framework that avoids unintended and undesirable consequences for the user and creates an environment of trust between the user, the platform, and the advertisers.

Design features and example use-case in job delivery

We next give a practical use-case of how a user-centered approach to ad delivery would work for a platform that delivered job advertisements. The main problem with the existing ad delivery systems is the way they estimate relevance of an ad to users is an inscrutable black-box. By establishing an ad delivery approach that provides users autonomy over the advertising tools and allows users to design the input of data, users, advertisers, and platforms gain agency against the mysteries of the "black box" in automated decision-making. In the context of job ads, the goal of our suggested approach is to clearly define what inputs are used to calculate relevance and what controls the user has over those inputs.

Opt-in policy for users

When it comes to user interface, the users would be offered a choice of whether to create a profile to have her/his/their preferences registered or not when they first start using the platform. They will also be asked about their preferences in terms of which ads they would like to be shown before interacting with other content.

Initially, a user who has just registered on a social platform should be shown non-personalized ads until the user explicitly opts-in. All the choices that the platform provides the user to control what kind of ads they see should follow an opt-in policy and should be initially set to a default setting that shows the user non-personalized ads.

At the ad preference setting page under the general account setting, the user can also choose to which extend they would like to have their preferences to be calculated as a factor in deciding which ads they are shown, meaning when the effect is set at zero, all ads would be randomized, and when set to 100, the ads would be highly customized and targeted to cater to one's preferences. More detailed setting options would offer the users a choice to determine the significance of each of the input data they have registered.

Ad preference settings

The illustration below shows how LinkedIn, a popular professional networking platform, affords settings for ad preferences.

Illustration 9. Ad preference options for LinkedIn users

While this provides a good example that other platforms can follow as a starting point, it can be improved by using an opt-in policy, instead of opt-out, and by not only making these settings available but also reminding users that they have the option to change it.

Reminders to adjust ad preferences

Users should be made aware and regularly reminded of the controls they have over what ads they see, for instance, through pop-ups (Illustration below).
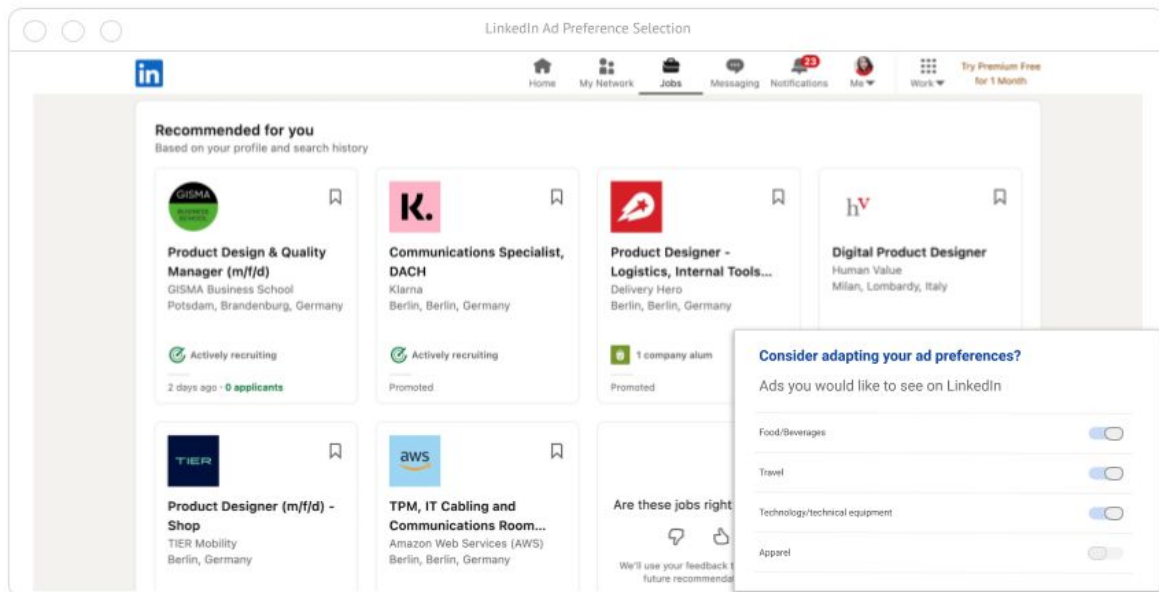

Illustration 10. Possible template of promted ad perferences

Dropdown menu with information about why ad is displayed

Another way a user can opt-in to be shown more relevant ads is through a dropdown menu where the user can see why this ad is shown to them and also choose to change the preference on this particular ad or this category of ads. This feedback would be used to train and improve the algorithms and generate better machine learning models.

Option to delete activity log

The platform can also provide an option to have any activity history log to be deleted.

Third-party data stored in-house instead

On top of this, the guideline also advocates a first-party program rather than the currently commonly used third-party data sharing, meaning the data would be stored in-house, which in turn would provide enhanced privacy and data protections for users.

Defining inputs to relevance estimation

If a user opts in to see personalized ads, only data inputs that are specific to the individual and chosen by them should be used in estimating relevance. Currently, ad platforms estimate relevance based on not just a user's data but also activities of *similar* members on the platform.

From the different data sources shown in Figure 1, a user may not be aware that the companies they follow or who they connect with determines what ads they see. By following an opt-in policy, the platform can ensure the user knows what data they provide is used to determine what job opportunities they see.

For job ads, for example, a user may opt-in to their profile data that shows their education and professional qualification to be used to show them ads that are more relevant to them. In this case, the "relevance score" that the platform calculates would be an estimate of how likely a user is to qualify for a job based on their background.

The platform should not merge the user's data with other "similar" users' data to determine relevance as such grouping has been shown to have unintended consequences that propagate stereotypes and result in unfair or even illegal discriminatory outcomes.

The illustration below gives a summary that shows how one can clearly define the inputs to the black-box used to estimate relevance, and give the user control over those inputs.
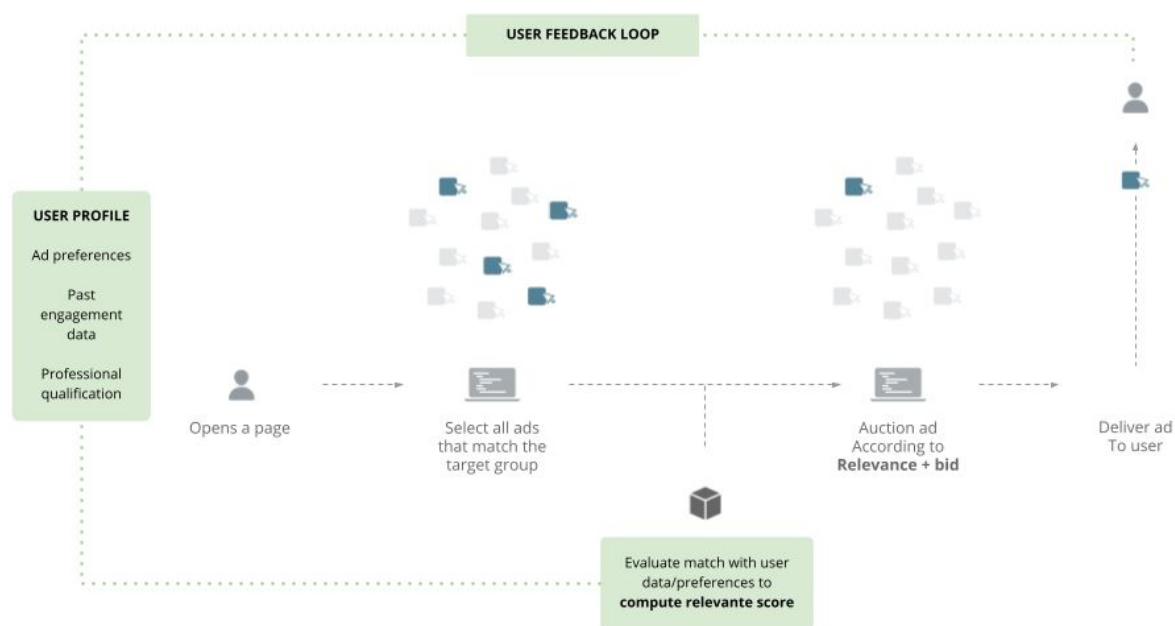

Illustration 11. Diagram showing how relevance estimation in ad delivery can work with user-defined inputs

Added value to platforms and advertisers

A user-centered approach is not only valuable to the user, but also benefits the advertisers and the platform. First, a user-centered approach allows for more accurate advertising profiles and inferences to be made about the users. Advertisers may lose user-engagement due to current automated practices inaccurately matching the ad to the user. In a system where the user can actively say what they are interested in, the advertiser will know what ads will be most effective and can plan accordingly. Second, this approach creates an environment of user-trust, by granting the user autonomy over how their ads are being delivered. In an environment of user-trust, users will more readily engage with platforms and advertising content.

SPOTLIGHT: AD DISPLAY

EZGI EREN, LINUS HUANG, SYLVI RZEPKA, MARIE-THERESE SEKWENZ

Guidelines: Improving fairness in ad display to users

Avatar-centric, educational approach to meaningful transparency

It is vital for the user to understand why they are seeing a specific job ad, what ads they might be missing out on (via Ad Repositories) and to have direct control over the ads they receive. To reach these goals, our guidelines are targeted towards platforms. The main tool we propose is an "Avatar": a user-friendly, "gamified" tool to visually communicate the information collected by the platform and the attributes used to target the user with job ads.

The Avatar will represent how the platform "sees" the user, via visuals that are natural and easy to comprehend.

- simple, layered view
- visual-heavy, video explanations, dashboards etc.
- easy interactions via clicking, sliders, scales etc. and via a chat bot

The Avatar will present any attributes and factors that influence ad targeting and delivery (gender, location, interests etc.). Information relating to such attributes and factors may be provided by the user, observed or inferred by the platform, or provided from third party sources (e.g., "custom audiences" feature on Facebook). It will also outline the respective weights of these attributes and factors in the ad delivery process. Lastly, the Avatar will provide information on any additional targeting carried out by the platform.

In order not to overwhelm the user, the information will be provided in a layered structure. More information will be available through more interaction. Video explanations, dashboards may be used to help them understand how the Avatar works and how to interact with it.

Give control to the user:

- change the Avatar's attributes (location, language, gender, interest etc) and relevant factors → receive ads accordingly
- the option to completely remove some attributes

The Avatar will also provide users with increased control, as users will be able to change their attributes via clicking on it or using sliders or scales. They will have access to a chatbot to have more direct interactions. To simplify the process, users will be able to choose their Avatar from one of the provided templates and fine-tune it to ensure it reflects the attributes on the basis of which users want to be targeted. By engaging with the Avatar, users will also be able to limit targeting that is based on certain attributes such as gender, language, or location, and disable targeting based on observed/inferred attributes or information provided from third party sources (see illustration below).
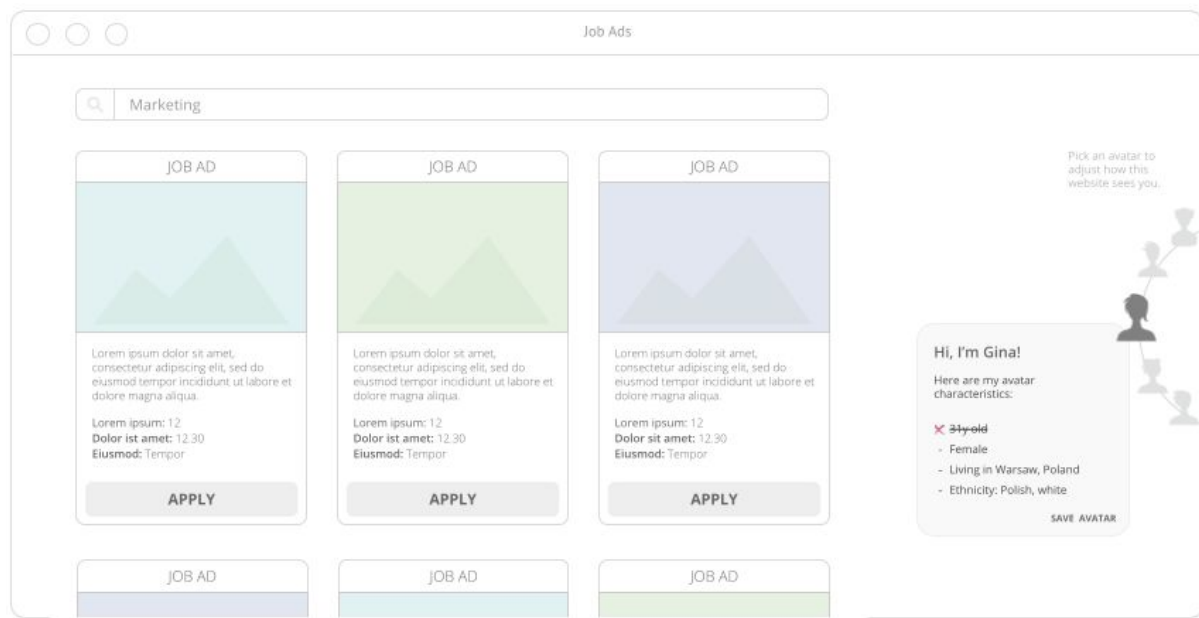
Illustration 12. Illustration for a possible interface with Avatar selection and adaption

Thanks to increased salience and control the Avatar provides, users will be presented with fairer, more relevant and less intrusive ads.

User education as core mission

Status Quo:
  – platforms provide information for ads display (usually hidden in the depth of settings)
  – yet, users are unaware of its significance and how to adjust the settings to their advantage

To combat stereotyping, education is a key missing link:
  – promote reflection on targeted advertising
  – contextualize against broader social issues (e.g. bias and structural injustice)
  – inform users of their rights
  – empower and facilitate further actions

In this section, we step back a bit to talk about the larger goal. The status quo now is that some platforms, such as Facebook, already provide ads-display related information and setting. They are often hidden in the user settings to discourage active usage. But even if the users are aware of such information, they are often unaware of its significance and how to act on it. In particular, they might not fully understand how targeted advertising can affect their opportunities, nor how they can improve their opportunities by adjusting the settings. As a result, education is a key missing link that enables actions.

Our Avatar-centric guideline can educate the users in order to empower them. For example, once the Avatar of a user captures their attention, it can be used to promote some reflection on how the platform

sees them and how it affects the ads they receive. Through additional interactions with the Avatar, the user can learn more about targeted advertising, and broader issues such as gender bias and structural injustice. Also, at the relevant time, a chatbot, e.g., can inform them about their rights, and instruct them to adjust the settings to protect them. Finally, additional features can be provided to empower the user, such as the ability to flag a potentially problematic advertisement or to provide feedback.

Next, we will talk about a more specific use case, and illustrate how a user can learn and be empowered through hands-on control experience.

Ad-repository to soothe user's FOMO

Problem:
Discrimination can happen through NOT seeing ads within an online environment.

How can we help to overcome the problem?
How can the user know about the ads they are not targeted with?

The phenomenon called Fear of Missing Out (FOMO) is usually related to problematic technology use and describes two scenarios. First, the feeling that others have experienced what the subject is missing, and second, the desire to stay in contact with other individuals within a group (Elhai et al., 2020). The first concept of FOMO can easily be adapted to the online world within the context of job seekers. What job advertisement can others see that might not be visible to me? Through this inability to see all possible job ads and the pre-selection of the algorithm legal discrimination might occur (Ali et al., 2019).

This structural problem – i.e. identifying what ads are "not there" – should be addressed through the design elements of an ad-repository as already suggested in the Digital Service Act proposal.[4] This repository should increase transparency for the user and enable individualized search for ads. The repository has to be designed in a user-centric way that makes the search easy to adapt and the tool easy to use. The repository should also be supported by the Avatar, which should help to explore the features but also explain to the user the search results, the parameters or the way results are displayed.

The user also can navigate through the repository by simple keyword search or geographic limitation options.

- — add-repository to collect the ads created and displayed on the platform
- — tool for the user to look for ads he/she/they might be interested in
- — repository can be queried by the avatar to help the user

---

[4] See Article 30 DSA Proposal.

How does the Avatar come to life?

The platforms themselves may lack an incentive initially to present their ad targeting strategies in form of an Avatar. Therefore, we suggest implementing the Avatar in form of an optional wrapper-app. This is an external app that uses the information on data and targeting provided by the platform and transforms it into the Avatar as described above. In time, those interacting with the Avatar will be more educated and empowered and share their enhanced user experience in their networks. This may make it more attractive for the platforms themselves to adopt these principles of salience and user control on their site.

References

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. *Proceedings of the ACM on Human-Computer Interaction, 3(*CSCW), 1-30.

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* (pp. 1-15).

Benner, K., Thrush, G., & Isaac, M. (2019). Facebook engages in housing discrimination with its ad practices, US says. *New York Times*.

Elhai, J. D., Yang, H., & Montag, C. (2020). Fear of missing out (FOMO): Overview, theoretical underpinnings, and literature review on relations with severity of negative affectivity and problematic technology use. *Brazilian Journal of Psychiatry* (AHEAD).

Facebook for Business. About Advertising Objectives. https://www.facebook.com/business/help/517257078367892.

Iwanska, K. (2020). Behavioural Advertising 101. *Medium*. https://medium.com/@ka.iwanska/behavioural-advertising-101-5fee17913b22.

Kayser-Bril, N. (2020). Automated Discrimination: Facebook Uses Gross Stereotypes to Optimize Ad Delivery. *AlgorithmWatch*. https://algorithmwatch.org/en/story/automated-discrimination-facebook-google/.

Kofman, A., & Tobin, A. (2019). Facebook ads can still discriminate against women and older workers, despite a civil rights settlement. *ProPublica*.

Lambrecht, A., & Tucker, C. E. (2018). Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *SSRN* (March 2018).

Sandberg, S. (2019). Doing More to Protect Against Discrimination in Housing, Employment and Credit Advertising. *Facebook Newsroom*.

Speicher, T., Ali, M., Venkatadri, G., Ribeiro, F. N., Arvanitakis, G., Benevenuto, F., ... & Mislove, A. (2018, January). Potential for discrimination in online targeted advertising. In *Conference on Fairness, Accountability and Transparency* (pp. 5-19). PMLR.