Overprofiling Analysis on Major Internet Players

Francisco Caravaca
Universidad Carlos III de Madrid
fcaravac@pa.uc3m.es

José González-Cabañas
UC3M-Santander Big Data Institute
jose.gonzalez.cabanas@uc3m.es

Ángel Cuevas
Universidad Carlos III de Madrid
UC3M-Santander Big Data Institute
acrumin@it.uc3m.es

Rubén Cuevas
Universidad Carlos III de Madrid
UC3M-Santander Big Data Institute
rcuevas@it.uc3m.es

ABSTRACT

Many Internet services obtain their revenue through the delivery of online advertisements based on the commercial exploitation of users' profiles. The accuracy and size of these profiles have important implications in terms of advertisers' campaign performance and users' privacy. Despite the importance of auditing the profiling accuracy, very little effort has been devoted both in industry and academia. This paper presents the most comprehensive auditing effort to understand the profiling accuracy of four major online advertising platforms: Google, Facebook, Twitter, and LinkedIn. Our work unveils that less than 50% of the assigned interests are relevant. Moreover, platforms can distinguish what interests within the assigned ones are more relevant but hide this information from users and advertisers. Finally, we have proposed a very simple solution that only uses 25 general interests per user. This proposal outperforms all the analyzed platforms in terms of profile accuracy while improving users' privacy.

KEYWORDS

privacy, social networks, advertisement, interests

1 INTRODUCTION

Online advertising is a major financial backbone of the current Internet and it is the core business model of a large portion of Internet-based companies such as websites, mobile apps, social media platforms, search engines, etc. This includes some of the most important technological companies based on their market capitalization such as Google or Meta. Compared to traditional marketing channels such as TV, radio, or traditional newspapers, online advertising allows advertisers to define much more fine-grained audiences to deliver their ads. This means users may receive much more relevant ads to their interests that will potentially increase the likelihood they get engaged with the service, product or good being advertised. To implement the described business model, the online advertising ecosystem requires the creation of user profiles (as accurately as possible). To achieve this, users are continuously tracked by the online advertising ecosystem as they browse websites, use mobile apps, interact with smartwatches, etc. The collected data is used by dozens of companies to create user profiles that may include, among other things: (i) demographic information such as age, gender, location (e.g., country, region, city, zip code, etc.), civil status (e.g., single, married); (ii) interests (e.g., football, rock music, Italian cuisine, etc.); (iii) technology use (e.g., mobile device used, browser used, etc.), etc. Therefore, when an advertiser aims to target *"users between 30 and 45 years old, interested in Italian cuisine and located in New York City"* the online advertising platforms deliver the ad to those users whose profiles match the target audience.

This business model is sustained by the collection of users' data, which includes the collection of personal information. The massive collection of personal data from digital platforms was one of the reasons why some countries decided to define modern data protection regulations to protect users' privacy from abusive and risky practices. The most popular data protection regulations are the General Data Protection Regulation (GDPR) [15] and the California Consumer Privacy Act (CCPA) [7] which are enforced in the European Union (EU) and California, respectively. These regulations place user consent at the core of users' privacy. Roughly speaking, these regulations require companies (including Internet businesses) to obtain explicit consent from the users to collect and process personal data for commercial purposes. Even more, the GDPR introduces an advanced principle referred to as *Data Minimization* that states that any entity processing personal data should collect the minimum required personal data for their business. If we adapt the data minimization principle to the online advertising ecosystem, online advertising companies should build profiles with the minimum number of attributes that allow them to deliver relevant ads to users.

In this context, it is reasonable to think that the goal of online advertising platforms (e.g., Google, Facebook, Instagram, Twitter or X, TikTok, LinkedIn, etc.) should be building very accurate profiles that increase the likelihood of delivering relevant ads to the users to engage them and achieve some type of conversion (e.g., good sell, service subscription, app installation, etc.), and at the same time being compliant with regulations such as the GDPR. The opposite behavior, i.e., creating very wide profiles assigning users many attributes may be a counterproductive strategy because (i) many of the assigned/inferred attributes may be irrelevant to the users and dramatically reduce the chances of the user to engage with the ads delivered based on them; (ii) it will not be aligned to the data minimization principle and may lead to potentially break regulations like the GDPR.

We believe auditing the profile accuracy of online advertising platforms is a very important task that should be in place as a de

Proceedings on Privacy Enhancing Technologies 2024(4), 929–946

https://doi.org/10.56553/popets-2024-0149

facto feature included within the online advertising ecosystem. We envision three players that would benefit from such an auditing feature: (i) Advertisers will ensure they are not wasting their budget and be sure their ads are delivered to users that match (with high probability) the attributes used in the targeted audience; (ii) Data Protection Authorities (DPAs) will ensure online advertising platforms are not abusing the GDPR and collecting much more personal information than required arguing legitimate business interest; (iii) Users will be labeled with few attributes reducing the privacy risks associated with very large profiles. For instance, large profiles simplify re-identification attacks [22, 29]. The more attributes assigned to a user profile the more likely is that the user is unique in the platform.

While there are metrics to study the performance of an ad, e.g., cost per click (CPC), clickthrough rate (CTR), we do not find this for profiling accuracy. As to the best of our knowledge, there are no standard auditing solutions in the online advertising ecosystem to evaluate profiling accuracy. Similarly, DPAs in the EU do not have any specific procedure to monitor whether platforms are creating accurate (and narrow) profiles fulfilling the data minimization principle. Finally, we could only find a handful [4, 34] of research work addressing this issue in the literature despite its relevance in terms of advertising performance and user privacy.

Unfortunately, our intuition, derived from a few years of researching the area of online advertising, is that online advertising platforms may be assigning users many attributes that: (i) may downgrade the performance of advertising campaigns, (ii) break the data minimization principle included in the GDPR, (iii) increase the privacy risks of users. We refer to this practice as *Overprofiling*. To the best of our knowledge, this work introduces the most comprehensive *overprofiling* analysis since it covers four major online advertising platforms such as Google, Facebook, Twitter, and LinkedIn. In particular, we audit the accuracy of the interests assigned to users by those platforms based on their online activity. In the case of Google, we also audit a few other inferred demographic parameters. The reason why we look into interests is that they are attributes very meaningful to create targeted ad campaigns with a clear context, which in addition can be combined with demographic attributes such as age, gender, or geographical location.

Auditing the profile accuracy of ad platforms is not an easy task, since the only partners getting access to the overall performance of the interests used in the ad campaigns are the ad platforms themselves. This work defines a methodology that allows a third-party to audit the performance of the profiles. In particular, users directly evaluate, on a scale of 1 to 5, the relevance of interests assigned by the analyzed platforms. Our methodology requires: (i) collecting the interests of the users across the four analyzed platforms; (ii) developing a friendly User Interface (UI) enabling users to efficiently score a large number of interests. We have implemented our auditing methodology within a web browser extension that allows us to complete both tasks. Therefore, users are only required to install our browser extension to carry out the interest scoring. Additionally, the extension captures the interests associated with ads delivered to users on all platforms except Google, along with user interactions (i.e., ad clicks).

Performing a comprehensive analysis requires recruiting dozens of users that classify thousands of interests altogether. We have recruited over 150 participants via the widely-utilized research platform, Prolific [33]. Overall, these participants scored 45310 interests. Moreover, we have captured 1477 ads targeting users based on their interests. To the best of our knowledge, this is the largest database to audit the profile accuracy of major online advertising platforms to date.

Our work will analyze the data within our dataset to answer the following research questions:

-*(i) Do online advertising platforms implement overprofiling?* We anticipate that platforms with accurate profiling will yield a majority of user-assigned interests with high scores of 4 or 5. This would mean the interests are relevant to users, consequently allowing advertisers to utilize meaningful interests in their campaigns.

-*(ii) Would a simple algorithm using a few tens of interest outperform online advertising platforms in terms of interest scoring?* We propose a simple algorithm inspired by the recent proposal of Topics API [10] within the Privacy Sandbox strategy [32] in which Google aims to improve users' privacy by eliminating the use of third-party cookies. Our solution assigns users 25 interests extracted from the topics (i.e., interests) Google assigns to the websites the user has visited more frequently in the last 2 months. This represents a simple and privacy-friendly strategy since the only required information is the topics associated with the websites, without the necessity of identifying the specific websites visited by users. Initially, we expect this simple strategy to present worse performance compared to the inference algorithms used by the analyzed platforms.

-*(iii) Do online advertising platforms distinguish relevant from irrelevant interests among the assigned ones?* If online advertising platforms consider all the interests they infer are relevant for the users, we would expect the interests embedded in the delivered ads within our dataset to be homogeneously distributed between relevant and irrelevant interests according to the score provided by the users. Contrarily, if real ads delivered to users are using more frequently relevant interests, we may conclude that platforms can weigh the relevance of the content they assign. This implies that platforms may have an internal classification of the relevance of the attributes assigned to users, which are hidden from the users. In other words, they may not be honest with their users and clients (i.e., advertisers).

- *(iv) Can we provide general guidelines to significantly reduce the number of assigned interests with little impact on the online advertising platform business model?* The goal is to provide some guidelines to only select interests that are more likely to be relevant. This would have three clear benefits: (i) advertisers will increase the probability of using relevant interests in their ad campaigns; (ii) users privacy will improve since their profiles will contain less information; (iii) online advertising platforms will align to the GDPR data minimization principle.

2 BACKGROUND

This section briefly introduces how the platforms considered in this work profile users to deliver them tailored ads. Most of these platforms retrieve some demographic information, such as age, gender, or location, directly from the user during the registration phase. In contrast, these platforms infer users' interests, which is the major focus of this paper since they allow advertisers to define

fine-grained audiences in their advertising campaigns, from their activity inside and, in some cases, outside the platform. Finally, we briefly discuss what are the privacy implications associated with overprofiling practices.

2.1 User profiling

Online Ad platforms rely on gathering, inferring, and recording user information to construct user profiles. Advertisers can define the profiles they want to target within their ad campaigns, referred to as audiences, and ad platforms will deliver those ads to users matching the targeted audience. Therefore, the efficiency of ad campaigns very much depends on the ability of ad platforms to build accurate profiles to maximize the relevance of the ads received by their users. Each platform has its own proprietary algorithms to infer users' interests based on the online activity of the user. The profiles are composed (along with demographic information and some other attributes) of lists of interests or categories that are assigned to the user. These lists are in constant evolution and some interests may appear or disappear over time [42].

2.1.1 Google. Google uses some of the basic information that is provided in the accounts' sign-ups, as users may want to fill in their birth date or gender. Moreover, Google may take advantage of the prevalence of its search engine (as it is currently being used by more than 93% of users [35]) to gather data. Furthermore, its ecosystem is much bigger, as the company has other services that may be used to leverage advertising, such as Google Maps, YouTube, Google Chrome, or even information generated by Android devices [30].

Furthermore, Google is present on external websites with third-party cookies, which may help this platform to make more accurate profiles of the users. However, this kind of cookies may be phased out in several years, as Google is pushing to remove them to improve users' privacy, providing an initiative known as Privacy Sandbox [32]. Even though this change seems like a positive step for protecting users, it will not stop any other kind of tracking done in Google's services. These cookies are not as necessary for Google as they are for other companies, as its position of their searcher, OS, and Internet Browsers will benefit Google against smaller competitors [19].

2.1.2 Facebook. This company claims that it mainly uses data from its social media platforms to profile its users. Facebook's (or Meta's) ecosystem is mainly composed of Facebook, Instagram, Messenger, and WhatsApp. Nevertheless, Meta states that neither of the last two services uses private message content for advertisement [41], but user registration information (e.g., phone number) or user behavior are tracked and shared with other Meta companies. Facebook tracks the activity of its users in terms of activity inside their platforms, such as what pages or profiles are followed, which posts are liked, and what the user comments or shares. Facebook can also gather information about the users' location, as its mobile app may be used to get this data [31]. With all the collected data, Facebook builds up the profiles of its users, providing a list of Interests. It is worth noting that Meta also owns its ad network to deliver ads outside their platforms (e.g., websites, mobile apps, etc.) and uses cookies to track users outside their platforms.

2.1.3 LinkedIn. As in the previous two companies, users can also fill out information during the sign-up process on this platform. In addition to this, LinkedIn allows its users to fill up *skills*, which refer to abilities, competencies, and expertise that they can showcase on their profiles. As Facebook and Google, this platform is present in many websites in the shape of third-party cookies [21]. With the information they collect from the users' behavior on their platform and pages visited, LinkedIn also compiles a list of interests.

2.1.4 Twitter. It is also similar to the previous two social media platforms, analyzing the users' behavior and their presence on the web via third-party cookies. They assign a list of generic interests, but they also include keyword targeting, which is similar to interests, but they appear when someone has interacted with a tweet that contains a certain keyword [38]. These keywords could be used for both ads and content recommendations inside the platform.

2.2 Ad targeting

These ad platforms allow advertisers to manage ad campaigns, which allows them to configure the audience to be targeted by a series of attributes (demographic information, interests, skills, etc.). There are a few differences between each of the services, mainly between Google and social media platforms.

2.2.1 Google Ad Targeting. This company has a predominant role in online advertisement, as it allows advertisers to show their ads in different channels: from Google Search results and ads in Google's services (e.g., YouTube), to regular websites and mobile apps. Google will be in charge of selecting the most relevant ads for a user visiting a webpage/mobile app, taking into consideration different aspects such as the available ad spaces, and information regarding the user's profile, such as interests and demographic data. Ads are sorted by an Ad Rank, which is calculated every time an ad is eligible to be shown [23].

2.2.2 Facebook, LinkedIn, and Twitter. These platforms offer advertisers to deliver ads inside of their social media platforms, nevertheless, some of them (e.g., Facebook) may still offer ads on third-party webpages or mobile applications but their prevalence is much lower than that of Google. Inside these platforms, the ads are mainly shown as if they were regular posts (in the users' feeds), but they may also appear in other spaces, such as in sidebar columns.

Advertisers are provided different options to target their ads, but they handle interests differently. In Facebook and LinkedIn, advertisers can narrow down the targeted audience by aggregating interests (and skills, in the case of LinkedIn) that have to be assigned to the profile of the users who may receive the ads. Therefore, advertisers can target very precise profiles. On Twitter, however, when an advertiser uses multiple interests in the audience definition the ads may reach any user with at least one of those interests. Therefore, using multiple interests means enlarging the targeted audience in practice.

2.3 Ads Managers and Transparency Tools

Online advertising platforms offer advertisers their Ads Managers to define ad campaigns: define the targeted audience, the campaign duration, the campaign budget, etc. At the same time, these platforms provide users with some transparency tools divided into Ad

Preferences Managers and Ad Transparency tools. Following, we briefly describe Ads Managers, Ads Preferences Managers, and Ad Transparency tools.

2.3.1 Ads Manager. Ads Managers are tools that are available to advertisers to manage advertising campaigns on these online platforms. Advertisers can select the targeted audience (e.g., selecting a range of interests or demographic data) for a particular ad campaign. In most cases, these tools also report an estimation of how many people have a particular interest worldwide or in a given geographical region. In the context of this research, the ads managers of Facebook, LinkedIn, and Twitter report the size of a pre-defined audience. However, Google does not report that information.

2.3.2 Ad Preferences Managers (APMs). These tools inform users of what interests they have been assigned. Furthermore, they also allow users to modify whether they want to be targeted by a specific interest. Each platform has its page and the amount of data available to see and change differs between them [16, 24, 28, 37]. Unfortunately, based on our experience, the amount of users accessing and managing their ad preferences is negligible, even though this kind of tools can effectively reduce concerns about data collection [17].

2.3.3 Ad Transparency Tools. These are tools that report to users the reasons why a particular ad has been displayed. We have observed and checked that some of these tools offer richer information than others:

- **Facebook**: This platform currently provides the most detailed explanation of why each ad is shown, as it shows the complete list of interests that were used to target an ad.
- **Google**: Although they give reasons in a clear manner, this transparency tool falls short of giving details about each reason, e.g., it tells the users that they were targeted by an estimation of their interests, but it does not specify which one.
- **LinkedIn**: This transparency tool only provides one reason for each category, e.g., it may report that an ad was shown due to one skill and one interest, but in reality, the advertiser may have used multiple interests and skills in their audience definition. This behavior was checked with Ads that were targeted by using a list of skills and interests.
- **Twitter**: It only reports one interest despite the advertiser may have used multiple interests to build the targeted audience. Keywords are not reported.

2.4 Privacy and GDPR considerations

The platforms analyzed in this study create wide user profiles based on: (i) the information users provide in the registration phase, (ii) information they infer out of the activity of the users within their applications, (iii) information they obtain by heavily tracking users in third-party applications (e.g., websites, mobile apps, etc.) using third-party cookies and other techniques such as fingerprinting [26]. Therefore, online platforms may store very large profiles from their users.

Although we acknowledge these platforms have to create users' profiles for their regular business operation (i.e., deliver relevant ads to users), we believe they should apply the data minimization principle exposed in Article 5(1)(c) of the European General Data Protection Regulation (GDPR) [15]. This principle states the data controller (i.e., ads platforms), should only collect personal data *"adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed"* (in this case, deliver relevant ads using personal data). In other words, overprofiling users by assigning them many interests does not align with the GDPR data minimization principle, since many of those interests may be inaccurate and, therefore, will be useless to deliver relevant ads to the user.

Someone could argue that users' interests are not personal data since they cannot identify individuals. For instance, assigning *Alice* an interest in *chocolate* does not imply *chocolate* is personal data since many other users on the platform will have that same interest. However, previous research has demonstrated that the combination of 4 rare interests or 22 random interests uniquely identifies a user on Facebook with a probability of 90% [22]. In addition, the authors in that work proved that a combination of interests could be activated to launch a *nano-targeting attack*. This is delivering an ad exclusively to a user based on a set of interests they have been uniquely assigned within a user base of $3b$ users. Nanotargeting attacks could be used for spear-phishing attacks, user manipulation, blackmailing, malvertising, etc. Recently, another paper demonstrated that something similar could be done on LinkedIn [29]. Therefore, it is necessary to consider user profiling a critical privacy issue, which means that platforms should be audited to check whether they comply with privacy-related regulations.

Overprofiling users with spurious interests may increase the chances for a third-party to create narrow profiles of the user to implement nanotargeting attacks (or micro-targeting attacks that would include a few hundred other users in addition to the victim, which may still be good enough for an attacker). Note that interests could be further combined with other attributes such as location, gender, age, etc. In a nutshell, overprofiling is a practice that: (i) increases the privacy risks for users, (ii) does not comply with the data minimization principle included in the GDPR.

3 METHODOLOGY

The goal of this paper is to assess the accuracy of users' profiles on Google, Facebook, LinkedIn and Twitter. To achieve this, we aim to recruit dozens of users who score the real interests they have been assigned in these platforms to later analyze the users' scoring and compute the volume of overprofiling in each platform.

Our methodology to being able to collect the data to carry out the overprofiling analysis is twofold: (i) we need to collect a large number of interests from users across the four analyzed platforms, and (ii) we need to provide users with a friendly interface to score hundreds of interests assigned to them by each platform in a reasonable amount of time. To achieve this, we have implemented a browser add-on that solves both tasks.

3.1 Browser extension implementation

Our methodology is implemented by means of a browser extension. Such a tool can execute JavaScript code in any tab of the browser (with the corresponding permissions). This allows us to collect the required information from users' platforms: (i) the list of interests of each social media platform, (ii) the ads that are delivered to users along with the reasons why they are receiving them.

Furthermore, the browser extension allows us to implement a friendly user interface through dedicated tabs. In particular, we have implemented a web page-like User Interface (UI) where the user will be able to rank the interests.

3.1.1 Data collection. The extension collects the following data:

- **Interests**: We leverage the four Ad Preference Managers to retrieve the list of interests assigned to the user. These are the interests the users will score using the UI.
- **Google Demographic information** Google infers some demographic information of the users (e.g., age or parental status). This information is retrieved from Google's Ads Preference Managers. Users will also evaluate whether Google's inference is correct.
- **Browsing history**: We retrieve the domain name of the websites visited by the user in the last two months. This is needed for the browsing approach detailed in Section 3.2.
- **Ads** The extension collects information related to the ads that were delivered via Facebook, Twitter, LinkedIn, as well as the ads delivered by Google on websites visited by the user while the browser was active. This includes IDs of the ads, posts, images, text, author of the ad, and the landing page of the ad (URL the user will visit when they click on the ad). The extension also collects the reasons why a particular ad has been shown: this may include a demographic factor (e.g., age), interests, etc.

Data was obtained with different methods: scraping, using Chrome API (for browsing history) or doing HTTP/S requests to different services. Annex C specifies the collection methods, with Table 3 showing a summary and specifying in which sections each kind of data is being used.

3.1.2 User Interface (UI). The UI includes multiple features. Some of them go beyond the scope of this paper, such as the classification of the relevance of the ads received by the user. The extension is divided into different tabs, being the relevant ones for this paper:

- **Interests Tab:** Users can rank assigned interests from 1 to 5 stars, with 1 being irrelevant and 5 being very relevant. The platform assigning the interest is undisclosed to prevent bias. Interests from all platforms are shown randomly, ensuring users classify interests from all platforms. Priority is given to interests found in delivered ads. If an interest appears on multiple platforms, users classify it only once.
- **Ads Tab:** In this tab, the users can see which ads were targeted to them and what the reasons were (if they were available).

In figure 1 there are two screenshots of the extension showing the Interests Tab and the Ads Tab.

3.1.3 Browser availability. The Overprofiling plugin has been submitted to the Chrome Store to facilitate its installation. We have passed all the required filters established in the Chrome Store to

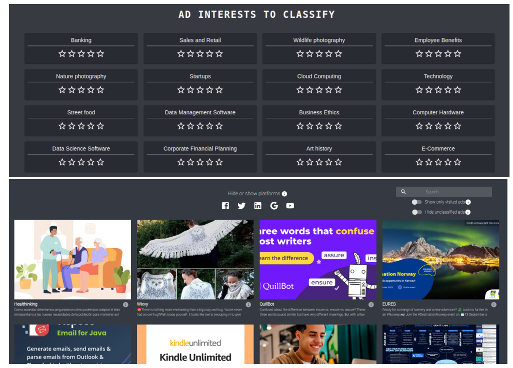

Figure 1: Browser extension screenshots

upload and update the extension. The plugin is publicly available in the Chrome Web Store[1].

Although the extension is only available in the Chrome Store, it could be downloaded using other Chromium browsers, these browsers, on the whole, take 90.47% of the market share [40].

3.2 Browsing history topics approach

Google is working in the so-called Privacy Sandbox [32] that aims to improve the privacy of end users. In the context of our research, each website is assigned several *Topics* that capture the content and context of the website. The proposal of Google, known as *Topics API* [10], assigns 5 interests to each user per week corresponding to the most frequent topics the user is exposed to according to their browsing history.

The list of topics for each URL is done by the Machine Model that is provided in the Google Chrome browser when the Topics API is active, which is available at `chrome://topics-internals` [9]. This model uses the domain name of a page, and determines from zero to three topics for that specific page. It is important to notice that this is still an experimental feature, and the model needs to be improved as some topics may be incorrect.

Topics are similar to interests, but they lean towards a broader and less specific spectrum to eliminate exceedingly niche interests. This approach aims to replace the need for third-party cookies, as they will no longer be needed as ad platforms could potentially access certain topics associated with each visitor. There are some safeguards in this system to avoid users being tracked through their list of topics; for instance, various websites will receive distinct topics, and there is a possibility of receiving a random topic as well.

In this paper, inspired by this proposal, we have designed a similar approach that is an approximation to the Topics API. The goal of this approach is twofold: (i) understand how a very simple approach that does not require complex inference algorithms works based on users' scores of the potentially assigned interests; (ii) have a benchmark to compare the profiling of the selected platforms with. We remind that these platforms use proprietary, and, very likely, complex inference algorithms or machine learning models to extract users' interests.

[1]https://chrome.google.com/webstore/detail/overprofiling/mnmnepgfknlklcegefknonnpdaafmgcb

Our approach uses the last two months browsing history of the user and computes the 25 most popular topics across the websites they have visited. Compared to the Topics API, we are extending the time window from three weeks to two months. The reason for this design is ensuring to have a few more interests samples to get more resolution. Nevertheless, in essence, our proposal is a very close proxy to Topics API.

4 EXPERIMENT AND USER RECRUITMENT

The next step of our research was recruiting users to rank their interests. Therefore, we relied on a professional recruiting service: Prolific [33] which is widely used by academic researchers to recruit users to complete specific tasks. In exchange, the recruited users receive economic compensation for the time they spend completing the requested task. Following, we describe the task users were requested, the final number of users participating in our experiment along with some data analysis of their participation.

4.1 Experiment

The participants were asked to fulfil a list of requirements to be eligible to participate in the experiment. These requirements were: (i) use Chrome as desktop browser; (ii) complete all sections (iii) be logged in to Google; (iv) have an active account and be logged in to at least two of the following platforms: Facebook, Twitter, and LinkedIn; (v) verify that Ad Settings is turned on.

Once the user has verified all the requirements, they can proceed to install the extension via the Chrome Web Store. Afterward, they can start the experiment, which will take roughly between 40 minutes and one hour. Note that although the study can be completed in that time-frame, the extension will query the browsing history of the user of two months prior to the moment the terms are accepted. The experiment is divided into the following parts:

- Complete the installation of the browser extension: reading and accepting the Terms of Use. If the user does not accept them the extension will not work, and we will not collect any information from the user.
- Filling a demographic form for age, gender, and country.
- Browsing the Social Media Feeds (Facebook, LinkedIn, and Twitter) for a few minutes each, to allow the extension to collect all the interests and ads delivered on every platform. Our objective is to collect a total of 60 Ads per user, with a minimum of 10 ads per platform. We also ask the user to browse regular websites to collect ads from Google ads.
- Once the user has collected the 60 required Ads, they can move to the Interests Tab. The first action they have to do in that tab is to answer how often they use each social media platform and classify as correct or incorrect the demographic parameters inferred by Google.
- Next, the user is asked to complete the core task for our research. They are instructed to classify at least 200 Interests (in case they have been assigned those many interests).
- After the user has classified the interests, they have to complete a final task where we ask them to classify the relevance (1 to 5 stars) of at least 50 ads. However, this part of the experiment is out of the scope of this research because it is not relevant to analyze the problem of overprofiling. Nevertheless, this information may be useful for future research analyzing the relevance of ads received by users on different platforms.

As the objective of the study was to do an experiment that took the user one hour, we decided to do all the tasks described above in that time frame. After several tries, we decided which were the most adequate thresholds for evaluated interests (200) and obtained Ads (60) to limit the study duration and obtain as much data as possible. These thresholds are enough to do the data analysis of this study, as we will have a representative sample of interests with valuations, and we will be able to obtain many ads from each user.

Although this experiment is conducted in an experimental setting and not in a more extended scenario, it is relevant to consider: (i) the collected data will not be very different; (ii) it would mean fewer participants due to budget constraints. The first point is that user profiles are built over time so that the user profile will stay mostly the same within a week or a month. Also, ads can be delivered by this profiling. As this does not change as much, the data obtained in this experiment will be similar whether the user browses the feed for some minutes or whether they have the extension installed during a week. Therefore, we have designed an experiment that only takes one hour to complete. However, we acknowledge that it is possible that more ads per user would have been obtained by doing a more extended study.

It is important to highlight that we have implemented controls to ensure the quality of the reply and avoid low-effort work from users. In Annex A we describe the implemented controls.

4.2 User base analysis

We have recruited 154 valid users, who received 6.8 pounds on average to complete the whole experiment. Here, we analyze the user base and some associated statistics to assess the quality of the dataset we will use to analyze the profiling accuracy. In Annex E, we carry out an analysis of the diversity of collected ads.

4.2.1 Demographics. Figure 2 shows the distribution of ages, genders, and countries of the participants.

The participants in our study range between 18 and 55 years old. Three out of four participants belong to the two younger groups, being the group 26-35 the most populous one (38.6%). The remaining quarter of participants is divided into 15.7% for the group 36-45 and 9.1% for the group 46-55. Although our dataset is imbalanced towards men (65%), we count with a relevant number of women participating in our study (i.e., 48 participants). Finally, we have participants from 25 different countries from Europe, North America, South America, Africa, and Asia. Although Prolific allows to set up studies with balanced samples, these are only available for the United States and the United Kingdom, and as we wanted to consider a wide range of countries, we did not use this feature.

In summary, we believe the recruited user base achieves a relevant heterogeneity degree that we have not been able to obtain without a professional recruiting service.

4.2.2 Interests. The number of interests assigned to the users substantially diverges across platforms and among users within the

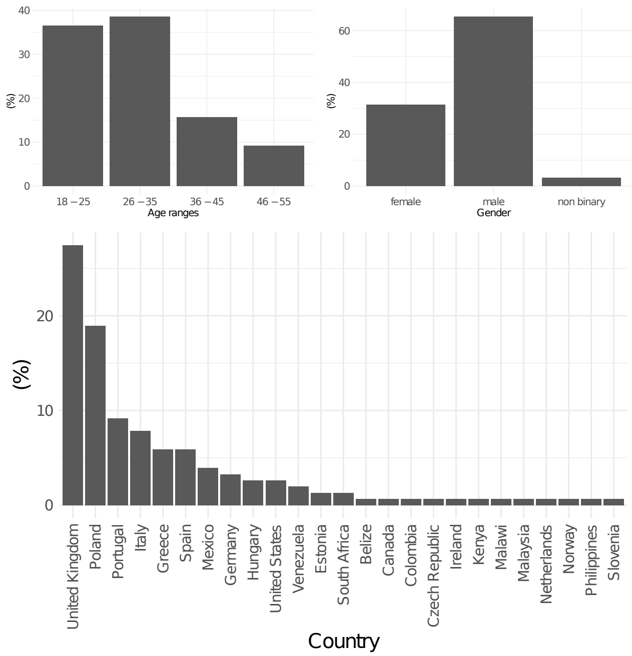

Figure 2: Demographic diversity of the participants. Each participant was asked to fill their age, gender, and country.

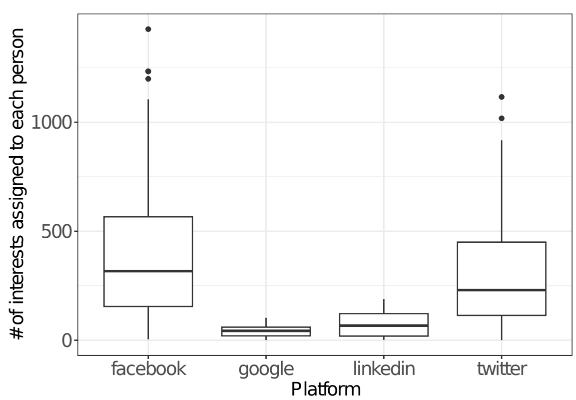

Figure 3: Boxplot with the number of interests assigned to each user per platform.

same platform. Figure 3 shows the boxplot distribution of the number of interests per user for Facebook, Google, LinkedIn and Twitter.

Facebook and Twitter are the platforms that assign more interests to their users. In median, they assign 317 and 230 interests, respectively, in contrast, LinkedIn assigns 67 and Google 43.

In addition, Facebook and Twitter, present a larger divergence among the amount of interests assigned to the users ranging from few interests for some users up to thousands of interests for other users. The user with the most interests in our database (across the four platforms) had 2146 interests. In contrast, the participant receiving less interest had only 7 interest. In median, a user in our dataset has been assigned 543 interests.

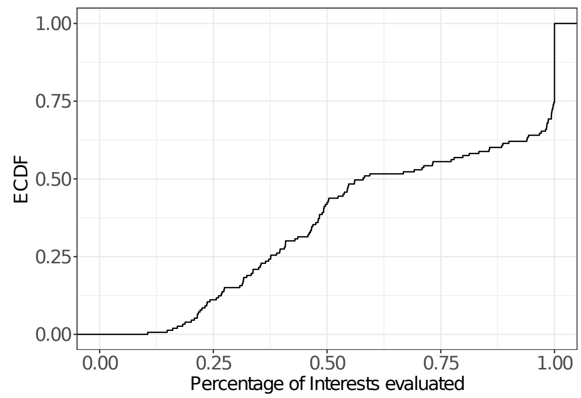

Figure 4: ECDF with the percentage of interests evaluated from the total of interests assigned.

We were aware that many participants with hundreds or thousands of interests might give up the classification task if we had forced them to rank all the assigned interests. That is why we defined a 200 interest threshold that we thought was a reasonable trade-off to obtain sufficient data for our profile accuracy analysis while keeping a good experiment completion ratio. Surprisingly, the median number of classified interests was 276, which is considerably higher (38%) than the established minimum threshold.

Figure 4 shows an ECDF (Empirical Cumulative Distribution Function) of the percentage of Interests evaluated from the total of assigned interests for each user. Almost 60% of users evaluated more than 50% of their interests, and more than 25% of the participants evaluated all of their interests.

We have performed an analysis of the data to understand whether users show some relevant fatigue as they progress in their task of classifying interests. Our results, discussed in Annex B, suggest users do not experience such fatigue.

4.3 Ethical considerations

Our study involves users' participation in completing the interests classification task. Some of the information we collect from them may lay under the category of personal data as defined by the GDPR. All the users participating in our study provided an informed consent that granted us permission to use the collected data for our research. We note that without providing such consent they would not be allowed to participate in our experiment.

In addition, the browser extension installed by the users to participate in the experiment includes a privacy policy including all the elements required by the GDPR. This privacy policy is publicly available at https://overprofiling.github.io/#/privacy_policy and includes information on how to exercise various data rights, such as access, rectification, erasure, restriction of processing, data portability, and objection. The policy clearly outlines the purpose of the extension, which specific data will be collected, and how it will be treated to ensure the protection of user data. Furthermore, to ensure compliance with regulations and to protect minors, the extension will not start working immediately after installation. Instead, users will be prompted to confirm whether they are adults and accept the

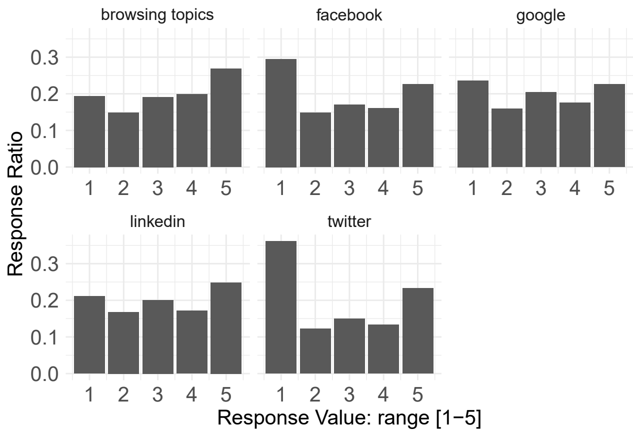

Figure 5: Interests users' response distribution.

terms and conditions outlined in the extension before it can begin collecting data.

Finally, it is important to highlight that our research has been reviewed and approved by the Data Protection Officer (DPO) of our institution, that certified it meets all the elements required by the GDPR, which applies to our country. The DPO is also a member of the Data Protection subcommittee of the Ethics Committee (EC). Our DPO considered the way we designed the experiment did not involve any ethical concern because users were providing informed consent to participate in the experiment.

5 PROFILING ACCURACY

In this section, we use the interest score provided by the participants in our experiment to analyze the profiling accuracy of the four platforms as well as the performance of our browsing history approach. We will compare the extracted results using two different angles. First, we will use the score distribution for each platform. Second, we will compute the average user's score. These two analyses will allow us to quantify up to what extent each platform is labelling users with irrelevant interests, i.e., overprofiling them. Finally, we also analyze the Google's accuracy regarding the inference of demographic attributes. It is relevant to remark that, although we are only analyzing a snapshot of each participant's interests and attributes, these platforms are constantly updating and removing interests that they deem no longer relevant to their users.

5.1 Interests score distribution

Figure 5 includes five bar plots that show the distribution of scores (from 1 to 5) in each of the five proposals analyzed: browsing history, Facebook, Google, LinkedIn, and Twitter, respectively.

All five proposals present an uneven distribution of scores. At first glance, the browsing history solution and LinkedIn are the best-performing approaches since they have the lower portion of interests classified with 1 star and the larger portion of interests classified with 5 stars. Facebook presents a dual pattern where 1 star (almost 30%) and 5 stars (more than 20%) are the most common answers, while the remaining interests are evenly distributed between 2 and 4 stars. Google presents the most homogeneous distribution where the portion interests ranked with 1, 3, and 5 stars is very similar (20%-23%). Finally, Twitter is the worst-performing platform, as over 35% of its interests receive only 1 star.

Table 1: Interests users' valuation distribution, grouping different valuations together.

Platform	1 star	2 or fewer	3 or fewer	**4 or more**
Browsing history	0.194	0.342	0.533	**0.467**
Facebook	0.295	0.444	0.614	**0.386**
Google	0.236	0.395	0.599	**0.401**
LinkedIn	0.211	0.379	0.580	**0.420**
Twitter	0.361	0.484	0.633	**0.367**

To carry out a more comprehensive comparative analysis, Table 1 aggregates the portion of interests for each approach receiving: only 1 star, 2 or fewer stars, 3 or fewer stars, and 4 or more stars.

We believe that in a good-performing platform, most of the assigned interests should be relevant for the users. That means, most of the interests should obtain a score equal to 4 or 5. If we use this reference to evaluate the profiling accuracy of the five analyzed solutions, none of them achieves 50% of relevant interests assigned to their users. Interestingly, the browsing history proposal is the best-performing approach (46.7% of relevant interests). The best real platform is LinkedIn, where 42% of the assigned interests are relevant. Google and Facebook show a very similar performance with roughly 4 out of 10 users's interests being relevant. Again, the worst performing platform is Twitter, for which only 36.7% of the interests are considered relevant by the users.

If we look at the results from a different angle and state that any content ranked with 1 and 2 stars may be considered irrelevant, we will obtain similar comparative results. Roughly 1/3 of the interests assigned by LinkedIn and the browsing history approach are considered irrelevant by the users. They are followed by Google and Facebook, with 39.5% and 44.4%, respectively. Once more, Twitter shows the worst performance since almost 50% of the interests are irrelevant.

Even though the effect on the platform is clear, we want to confirm these results and perform a Pearson's Chi-squared test with users' valuations and platform, and we encountered a very significant association (p-value < 0.001) between the two variables, thus the null hypothesis can be rejected.

Overall, if we had to answer the question: are platforms overprofiling their users with irrelevant interests? The clear answer is yes, they are. Even in the best performing platform, i.e. LinkedIn, more than 1/3 of the assigned interests are irrelevant and less than half are relevant.

It is also interesting that a solution using the browsing history, which assigns interests as the most frequent topics among the websites visited by the users, is the best-performing approach. This result suggests that the introduction of Topics API may have a double positive impact: (i) it will reduce the number of interests to only 15, which will improve users privacy, (ii) this huge reduction in the number of interests (3× w.r.t. Google, 4× w.r.t LinkedIn, 15× w.r.t. Twitter, and 21× w.r.t. Facebook) does not impact the relevance of the assigned interest, but instead it improves most of the current

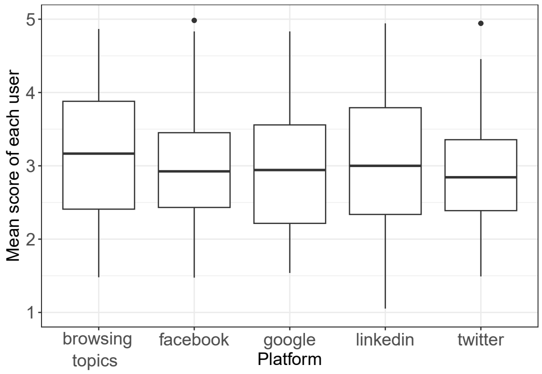

Figure 6: Mean user score distribution. Each sample is a the mean score of a user in a given platform. Only users that have evaluated at least 10 interests in said platform are considered.

existing inference algorithms. We have run an experiment that discards the impact of self-reporting biases in our results in Annex D.

Finally, we want to highlight that the reported results are especially worrying in the case of platforms that assign a large number of interests to users, such as Facebook and Twitter. These two platforms assign a very large number of irrelevant interests that may negatively impact the privacy of their users. Those irrelevant interests may also negatively impact advertisers since the impressions delivered to wrongly labeled users will have a very small likelihood to lead a conversion.

5.2 Average score per user

The previous section aggregates all the interests classified by users in each platform. In the current section, we obtain the average score provided by each user on each platform. Figure 6 shows a boxplot per platform depicting the distribution of the average user's score. We note that we have only considered those users that have ranked at least 10 interests in a given platform.

The results deliver a similar picture to the one obtained in the previous section. Although in terms of the median, all the platforms but Twitter (2.84) show a similar a rather similar value, if we look at the distribution, the 75th percentile of LinkedIn (3.79) and the browsing topics (3.88), are substantially higher than in any other solution (Google (3.56), Facebook (3.45), and Twitter (3.36)). The percentage of users providing an average score $\geq$ 4 is 22.58%, 17.82%, 12.22%, 10.0%, and 8.51% for LinkedIn, browsing topics, Google, Facebook, and Twitter.

We have also checked whether we can find a significant influence from the platform and the users into the valuations using an analysis of variance (ANOVA), in this case the null hypothesis can also be rejected (p-value < 0.001).

These results backup our analysis from the previous section. If the analyzed platforms were doing a good interests inference we should have many more users with an average score $\geq$ 4. The reported results reveal that Facebook, Twitter, Google, and LinkedIn are assigning their users a large portion of irrelevant interests. In other words, they are overprofiling them with useless interests (since they should provide not relevant ads) that may impact users' privacy.

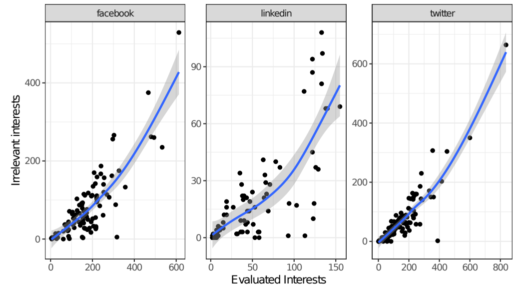

Figure 7: Profile size and irrelevant interests relationship.

5.3 Irrelevant interests and profile sizes

The previous sections did not address the impact of user profiles on profile quality. Figure 3 illustrates a significant difference between platforms with diverse and high-interest profiles (Twitter and Facebook) compared to those with fewer interests and lower variability (Google and LinkedIn). Platforms with excessive interests tend to have poorer user perception, as indicated in Table 1.

Intuitively, when a platform assigns a large number of interests to a user, the volume of information managed by the platform increases substantially. In a hypothetical scenario where user A has 1000 interests and user B has 10, the potential for spurious (irrelevant) interests is much higher for user A. This implies that managing irrelevant information for user A is considerably larger than for user B, aligning with the data minimization principle.

Figure 7 depicts a positive correlation between the number of interests classified by the user (X axis) and the number of irrelevant interests (scored 1 or 2 by the user, Y axis). Using generalized additive models, there is an exponential increase in the number of irrelevant interests with the profile size. Essentially, a broader profile not only results in more irrelevant interests but also increases the proportion of irrelevant interests relative to the total assigned interests.

5.4 Google demographics

To conclude this section, we evaluate the performance of Google inferring demographic attributes. We remind the demographic attributes Google reports within its Ad Preference Manager are: Age, Parental Status, Gender, Home Ownership, Relationships, Company Size, Education, Job Industry, Language, Education Status, and Household income. We retrieve all the cases in which Google assigned a value to any of these parameters. We ask the users to perform a binary evaluation indicating whether Google's inference is correct or incorrect. Table 2 shows the accuracy of Google for each of the listed parameters as the portion of times Google correctly infers an attribute within our experiment database.

Although accuracy seems quite high in most cases, some estimations are not as good as one could expect from a company that manages a lot of data from its users. Google performs very well for gender (94.4% accuracy) and Education Status (~90%). However, Google is showing a rather bad performance in categories such as

Table 2: Demographic data assigned by Google and its accuracy.

Category	Total	Correct	Wrong	Accuracy
Age	135	114	21	0.844
Gender	126	119	7	0.944
Parental Status	123	90	33	0.732
Relationships	117	72	45	0.615
Home ownership	116	77	39	0.664
Company Size	96	56	40	0.583
Education	92	63	29	0.685
Job Industry	76	45	31	0.592
Language	12	11	1	0.917
Household income	11	7	4	0.636
Education Status	8	7	1	0.875

Parental Status (~75%) or Homeownership (~65%). Even for the age, they fail in almost 15% of the cases. This may have an impact on the advertisers' performance, e.g., an advertiser targeting Homeowners (the Homeownership category has only two options: Homeowners or Renters) will be showing their ad to many users (roughly 30%) that do not own a house.

Overall, the main lesson of the demographic analysis regarding the inference of interests, is that if a large company like Google, with a huge amount of resources, is not able to properly infer simpler attributes than interests, the inference of interest may still be a complex exercise where online advertising platforms have still much room to improve.

6 INTERESTS USED IN REAL DELIVERED ADS

The previous section has depicted how platforms overprofile users with a large number of irrelevant interests. There are two important questions following the obtained result: (i) Do platforms select homogeneously campaigns targeting high-score and low-score interests? (ii) Do advertisers target homogeneously high-score and low-score interests?

To tackle these questions, we analyze the interests included in the ads collected in our methodology by leveraging the Ads Transparency Tools described in Section 2.3.3. This means a portion of the ads delivered to the users in our experiment were targeting them based on some interest(s) assigned to the user, and we retrieved those interests. Overall, our dataset includes 1477 ads targeting users based on their interests: 572 on Facebook, 589 on Twitter, and 316 on LinkedIn[2]. We can compute the score distribution associated with interests used in real ads and compare it with the overall score distribution, including all the interests assigned to the user (see Figure 5). Such comparison will denote whether platforms and advertisers tend to select higher-score interests when it comes to real ad campaigns.

If the interests used in actual ad campaigns are randomly selected from the interests assigned to the user we would expect to obtain a similar distribution as the ones reported in Figure 5.

Figure 8 shows the interests' score distribution for Facebook, LinkedIn and Twitter. The left side (red) shows the score distribution

Figure 8: Distribution of interests valuations used as a reason in ads (left column), and general interest valuation (right column).

for the interests used in real ads available in our dataset. The right side (blue) shows the overall interest distribution (i.e., the same as in Figure 5).

We observe that the distribution extracted from real ads is biased towards higher score values. If we take as reference the portion of relevant ads (i.e., 4 or 5 stars) LinkedIn grows from 42.0% in the overall interests distribution to 49.68% in the real ads distribution, Facebook from 38.6% to 61.54%, and Twitter from 45.7% to 62.1%. Similarly, if we consider the portion of irrelevant ads (i.e., 1 or 2 stars) LinkedIn decreases from 37.9% in the overall interests distribution to 22.5% in the real ads distribution, Facebook from 44.4% to 22.9% and Twitter from 48.4% to 18.7%. Overall, the portion of relevant ads notably increases and the portion of irrelevant ads substantially decreases on all the platforms, especially on Twitter and Facebook.

If the interests used in actual ad campaigns were randomly selected, both distributions should be similar. We carried out an ANOVA test to confirm this statement, and we found a significant effect (p-value < 0.0001). The obtained results suggest two important conclusions: (i) platforms are somehow weighing the relevance of the interests they assign to users. Even if they assign many interests to users, when they have to deliver an ad (from thousands of campaigns using interests), they tend to choose ads campaigns targeting more relevant interests (i.e., high score); (ii) advertisers targeting tend to use interests better scored within the whole pool of interests.

The most important implication of our results is that online platforms can make a distinction between more relevant and less relevant interests among those they assign to users. Therefore, they could create more accurate profiles and substantially reduce the number of interests they assign to the users, focusing on those that the platform knows are more relevant. This will reduce the privacy risks platforms expose to users by assigning more interests than needed and will align with the data minimization principle stated in the GDPR. Similarly, it seems removing most of the irrelevant interests may not significantly affect their business since our results suggest that advertisers tend to also use more relevant interests in their targeting audience definition.

[2]Note that we are excluding Google from this analysis since, as we described in section 2.3.3 Google does not provide the interests why an ad was delivered to a user.

7 POPULARITY VS. RELEVANCE

To conclude our research, we aimed at exploring some simple recommendations that could be widely used by online platforms to decide what type of interests are better suited to label users if they were to reduce the current overprofiling approach they are adopting.

Our intuition was that popular interest (i.e., those assigned to a very large number of users) may be a good compromise for two reasons: (i) popular content is less invasive in terms of privacy since you need to aggregate many of them to uniquely identify an individual. We proved that 4 rare (or niche) interests are enough to uniquely identify a user with 90% probability [22]. In contrast, a third-party would need 22 random interests to achieve the same reidentification likelihood. Random interests include both niche and popular interests, so it is very likely that if a third-party could only combine popular interests, the required number to uniquely identify an individual would grow way beyond 22; (ii) Although there could be some advertisers running ad campaigns targeting very reduced audiences based on some niche interest(s), most advertising campaigns do not use very niche interests on their ad campaigns. However, we acknowledge that advertisers could still fine-tune their targeting audience using other factors such as age, location, gender, and other targeting options.

We collected the audience size of each interest from each platform[3] scraping their Ad Manager platform. The median and lowest (worldwide) audience sizes associated with the interests used in the real ads stored in our dataset are 148.8M and 3440, respectively. This supports our claim that, in most cases, advertisers are using very common interests.

In this section, we aim to evaluate whether we can find a positive correlation between interests' popularity (i.e., audience size) and interests' score. If such a correlation exists our recommendation for online advertising companies would be to generally skip niche interests and focus on popular ones.

Next, we first briefly present the interests' audience size distribution on Facebook, Twitter, and LinkedIn. Then, we analyze the correlation between popularity and relevance in each platform. Finally, we verify the correlation when using only the interests associated with real ads.

7.1 Audience size distribution per platform

Figure 9 shows the distribution of the audience size from all interests that were collected for each of the platforms, excluding Google, which does not allow the retrieval of this information.

Each of the three platforms has unique distributions. Twitter has a very spread audience range of audiences, whereas LinkedIn takes the opposite approach and offers a very narrow audience range, which means that they offer rather generic interests. Facebook provides a wider range (in terms of audience) of interests, which are generally bigger than those from Twitter.

The case of Twitter may be explained as they offer two different types of interests: (i) regular interests, which are usually generic, as there are only 374 of them; (ii) keywords, which can be anything.

[3]We exclude Google, as it does not provide this information.

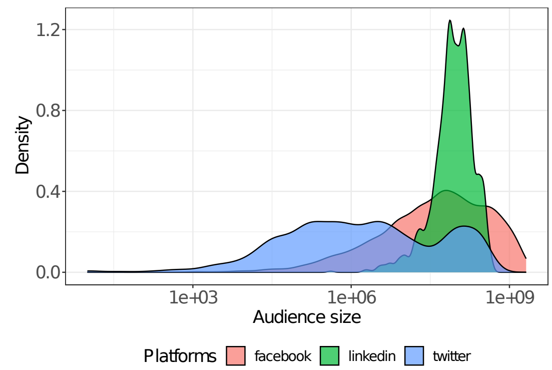

Figure 9: Audience sizes distribution of the interested assigned in Facebook, LinkedIn, and Twitter.

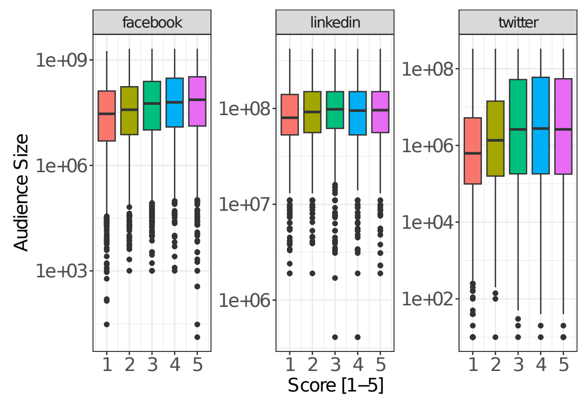

Figure 10: Audience sizes and valuations of interests

7.2 Audience size Vs. scores

Figure 10 shows for all the interest in a platform receiving the same score, the distribution of the audience size using a boxplot. The results show a positive correlation between the audience size and interest score. If we focus on the median audience size in the case of Facebook, it grows exponentially (note the y-axis is on a logarithmic scale and we see a linear growth on this scale). In the case of Twitter, irrelevant interests (1-2 stars) present a median audience size between 1 and 2 orders of magnitude lower than relevant interests (4-5 stars). In contrast, on LinkedIn we observe very similar distributions for all the scores, even so, the smaller medians still correspond to irrelevant categories. This result is explained by the narrow audience size distribution, within large audience values, compared to Facebook and Twitter. For instance, 0.08%, 8.59%, and 3.03% of the interests on LinkedIn, Facebook, and Twitter, respectively, report an audience size $\leq$1m.

To further validate the results, an ANOVA test was performed, and both user valuations and the platform have a significant effect on the audience size of the interests (p-value < 0.001).

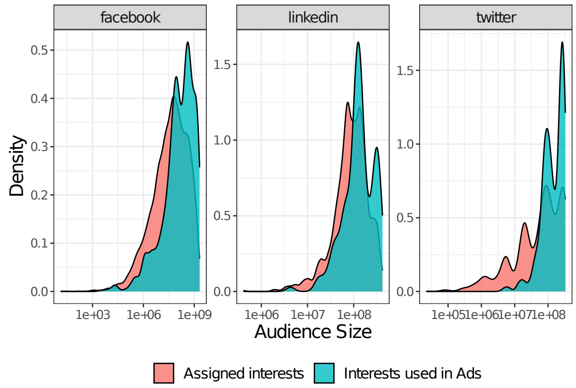

Figure 11: Distribution of the audience sizes of the assigned interests against interests used in ad targeting

7.3 Distribution of audience size in real ad campaigns

Given that higher audience size leads to higher interest scores, we expect that the audience size distribution of the interests used in real ad campaigns is biased toward higher audience interests on Facebook, LinkedIn, and Twitter.

Figure 11 shows for each platform the distribution of audience size for (i) all the interests assigned to users in our dataset[4] (red) and (ii) the interests used in the ads stored in our dataset (blue). As we expected, in all three platforms, the distribution associated with real ads is biased to higher audiences.

For instance, on Facebook, we find that the median size across all assigned interests is 50M, whereas the median audience size of interests used on targeted ads is 180M. But the most relevant insight is that if we get the 5th percentile, we found that interests used in real ad campaigns have 1.3M of population, but in particular, only 2.84% of interests selected by advertisers have less than 1M population. In the case of LinkedIn, these values are 95M (all interests) and 130M (interests used on ads), and the 5th percentile in the interests used in ads has an audience of 34.8M. Finally, in the case of Twitter, the median values are 76M and 151M, respectively, with a 5th percentile of used interests of 40.1M.

The results obtained in this section reinforce our conclusion that interests associated with higher audiences are more likely to be relevant for the users. Therefore, our recommendation for online advertising platforms is to avoid assigning niche interests to users and focus as much as possible on popular interests.

8 DISCUSSION

This research has demonstrated that advertising online platforms systematically overprofile users with interests that are irrelevant to them. However, it appears that platforms such as LinkedIn, particularly Twitter and Facebook, demonstrate a discerning ability to identify which among the assigned interests are more likely to be relevant to users. This finding correlates with the observed phenomenon where interests used in actual ad campaigns exhibit a more favourable score distribution than when considering all assigned interests. Following, we discuss privacy considerations from this study and describe a few recommendations for interest-based advertisement.

8.1 Re-identification attacks

Overprofiling users increases the probability that a third party may re-identify the user and even nanotarget them with exclusive messages. As shown in one of our previous research works [22], selecting 27 random interests is enough to uniquely identify a user on Facebook (with a 95% confidence), thus it will be possible to nano-target a specific ad to a user (e.g., to do spear-phishing attacks) if the attacker knows a proportion of the interests of the user. But, this attack is way easier using the least popular interests, as only 6 interests will be necessary. Considering this, and also our results in Section 7, it is clear that niche interests, which are more likely to be wrongly assigned, are the ones that provide the most privacy risks, allowing easier re-identification attacks.

8.2 Data minimization considerations

Setting unnecessary interests is also a concern regarding the adherence to the data minimization principle. The principle stipulates that personal information collection should be limited to what is directly relevant and necessary for a specified purpose. In the context of ad platforms, our results suggest that the collection of numerous user interests, which may not directly contribute to delivering relevant ads, contradicts this principle and poses unnecessary privacy risks by overloading users with more interests than essential for effective ad targeting.

8.3 Recommendations

Considering the privacy risks of interest-based advertisement, we propose a few recommendations: (i) reduce the number of interests assigned to users to 25 or fewer. Each platform can implement its algorithms to decide what are the more relevant interests to be included in the user profile; (ii) select popular interests and remove niche interests; (iii) demand platforms to increase their transparency and not only report what interests they have assigned to a user, but also what is the relevance level the platform attaches to each of them. The only drawback of removing niche categories is that advertisers lose some of the potential targeting capabilities.

9 RELATED WORK

The literature has widely audited major technological players in the last years to assess whether their implemented practices where ethical and subject to different data protection regulations.

There is a body of works that aims to unveil how users are tracked in the web ecosystem. We present some examples next. Englehardt and Narayanan made an analysis crawling 1 million webpages, unveiling and measuring different tracking strategies [12]. Acar et al. [1] unveil that some tracking methods may circumvent users' tracking preferences using browser features. Englehardt et al. [13] demonstrate that third-party tracking cookies allow better identification of a user rather than their IP address. Some works suggest that in some cases, fingerprinting could be better than cookies for tracking users [11] such as the case of fingerprinting

[4]For Twitter, we only show the assigned interests as keywords are not reported in their Ad Transparency Tool as we detail in Section 2.3.3

users based on the fonts they use [18]. Since online advertising is constantly evolving, there is a very interesting study by Lerner et al. that used the Internet Archive's Wayback Machine [27] to study the evolution of tracking strategies from 1996 to 2016.

A different body of literature analyzes online advertising platforms to unveil unethical and, in some cases, non-compliant practices according to the most advanced data protection regulations. For instance, Ali et al. show that ad delivery on Facebook is different depending on factors such as race or gender [2]. The same authors demonstrated that the prices for reaching a user in political ads depend on the political alignment of the audience [3]. Whereas we showed that Facebook was tagging 43% of EU Citizens [5] and 22% of worldwide citizens [6] with sensitive ad preferences (i.e., interest).

The literature has revealed several potential attacks that exploit the possibility of running targeted ads in only advertising platforms. Korolova [25] shows how it was possible to send a targeted ad to infer private information from a victim. In a more recent work [22] we show how it is feasible to deliver an ad exclusively to an individual on Facebook based on their interests, which in isolation are not considered as personal data.

Also, the academic literature has analyzed in several works the quality of the information delivered to users in Ads Transparency Tools on platforms like Google or Facebook. Researchers have found that the reasons provided both in Google [8] and Facebook ads [39] may be incomplete. The authors of these studies recommend that Google and Facebook should provide more concrete reasons regarding why a particular ad is delivered to a user. In another study, Eslami et al. concluded that users are more satisfied when the ad reasons are linked with their identity [14]. Gkiouzepi et al. defined a framework to infer/deliver more precise ad targeting reasons [20] to overcome some of the limitations currently present in Ads Transparency Tools.

Despite there is a large body of literature analyzing the practices and behaviors of major Internet players, there are very few ways to audit their profiling performance, which is the main focus of our work. For instance, Tschantz et al. [36] analyzed the demographic inferences by Google in 2018, although since then, Google could have changed their inferring methods. Furthermore, there are more categories that were not present such as parental status, relationships or home ownership. The most similar work to the one presented is the one from Bashir et al. [4] in which they study user profiling through the APMs of the following platforms: Google, Facebook, BlueKai, and eXelate. They recruit participants only from US and Pakistan, surveying 20 interests from each user. They focus on whether assigned interests are related to recent online behavioral data and they also analyze whether users recall seeing an ad related to an interest. Their study captures that the majority of users say that less than 50% of interests were relevant to the user. However, they do not analyze why some interests may be wrongfully assigned. Similarly, Sabir et al. [34] focused on the inferring process of Facebook interests. They also noted that interests used to target ads are usually the ones that are relevant to the user. However, their dataset is rather small to derive significant results. In their study, from the 73 users that received ads, only 37 of them received interest-based ads, accounting for 102 ads with that kind of ads. In contrast our study has recruited 154 users, from which 128 of them obtained ads with interests as targeting reasons, with a total of 1477 interest-based ads.

In summary, our work improves preceding studies as follows: (i) it is the only one addressing simultaneously 4 of the most popular Internet services. Other studies either focus on a single platform or include data management platforms in their analysis, which are barely known by regular Internet users. (ii) Our work is the only one that proposes and evaluates an alternative solution that outperforms major Internet platforms in terms of profiling accuracy. The proposed solution may be used as a general rule of thumb guideline with the goal of improving users' privacy and advertiser campaign performance, and, at the same time, having a low impact on online advertising platform revenue.

10 CONCLUSION

Our results show that less than 50% of the interests are relevant for the users, independently of the platform. However, we can observe differences among them, as LinkedIn provides the best profiling out of the four platforms, with Twitter being the worst. This practice increases users' vulnerability in terms of privacy and negatively impacts the performance of ad campaigns.

We have unveiled an even more worrisome issue that suggests platforms can distinguish what interests within the assigned ones are more likely to be relevant and use them more frequently in the actual ads delivered to the users.

Our results also suggest that the likelihood of an interest being more relevant increases with the popularity of the interest. This means participants in our study tend to rank better popular interests than niche interests. We have proposed a browsing history approach, inspired by Topics API, that supports our claim. Our proposal only uses 25 general interests derived from the topics assigned to the websites the user visits more frequently. Our solution outperforms all platforms. Overall, this result allows applying a very simple rule, i.e., only label users with popular interests. This improves users' privacy, without heavily impacting the business model since most ad campaigns target popular interests.

We acknowledge that inferring a large number of accurate attributes, such as interests, to build the online profile of a user is not a simple task. As we have found that building smaller profiles is usually easier for the platforms in terms of performance. But what we usually see is that the platforms are *overprofiling* users with unnecessary interests, which has negative privacy implications. Not only does the profiling itself require a lot of personal information, and simpler methods such as Topics API may be a better solution, but having so many interests makes it easier for *nano-targeting* attacks to exist. This means that data protection agencies and other privacy stakeholders should request online advertising platforms to reduce the number of attributes they assign. In addition, those attributes should be popular content that makes it complex to re-identify the user through a combination of them. This approach not only benefits users' privacy and meets advanced data protection principles such as the GDPR data minimization principle but will also reduce the budget wasted by advertisers delivering targeted ads to users based on certain attributes irrelevant for a large portion of users receiving them.

ACKNOWLEDGMENTS

This research has received funding from: the project ENTRUDIT (Grant TED2021-130118B-I00) funded by the MCIN/AEI/10.13039/501100011033 and the NextGeneration EU/PRTR funds; the European Union's Horizon 2020 innovation action programme under the TESTABLE project (Grant 101019206); and the Ministerio de Asuntos Económicos y Transformacion Digital and the European Union-NextGenerationEU through the project PRTIACLOUD; the project PRTR-INCIBE ANTICIPA

REFERENCES

[1] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. 2014. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (Scottsdale, Arizona, USA) *(CCS '14)*. Association for Computing Machinery, New York, NY, USA, 674–689. https://doi.org/10.1145/2660267.2660347

[2] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. Discrimination through optimization: How Facebook's Ad delivery can lead to biased outcomes. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–30.

[3] Muhammad Ali, Piotr Sapiezynski, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2021. Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (Virtual Event, Israel) *(WSDM '21)*. Association for Computing Machinery, New York, NY, USA, 13–21. https://doi.org/10.1145/3437963.3441801

[4] Muhammad Ahmad Bashir, Umar Farooq, Maryam Shahid, Muhammad Fareed Zaffar, and Christo Wilson. 2019. Quantity vs. Quality: Evaluating User Interest Profiles Using Ad Preference Managers. In *Proceedings 2019 Network and Distributed System Security Symposium*. Internet Society, San Diego, CA, 15 pages. https://doi.org/10.14722/ndss.2019.23392

[5] José González Cabañas, Ángel Cuevas, and Rubén Cuevas. 2018. Unveiling and Quantifying Facebook Exploitation of Sensitive Personal Data for Advertising Purposes. In *Proceedings of the 27th USENIX Conference on Security Symposium* (Baltimore, MD, USA) *(SEC'18)*. USENIX Association, USA, 479–495.

[6] José González Cabañas, Ángel Cuevas, Aritz Arrate, and Rubén Cuevas. 2020. Does Facebook use sensitive data for advertising purposes? *Commun. ACM* 64, 1 (2020), 62–69.

[7] California State Legislature. 2018. California Consumer Privacy Act. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=CIV&division=3.&title=1.81.5.&part=4.&chapter=&article= Cal. Civ. Code §§ 1798.100–1798.199.

[8] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2014. Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. *CoRR* abs/1408.6491 (2014), . arXiv:1408.6491 http://arxiv.org/abs/1408.6491

[9] Sam Dutton. 2022. Topics API Developer Guide. https://developer.chrome.com/docs/privacy-sandbox/topics/.

[10] Sam Dutton. 2023. The Topics API. Private Advertising Technology Community Group Individual Draft Space.

[11] Peter Eckersley. 2010. How Unique Is Your Web Browser? In *Privacy Enhancing Technologies*, Mikhail J. Atallah and Nicholas J. Hopper (Eds.). Vol. 6205. Springer Berlin Heidelberg, Berlin, Heidelberg, 1–18. https://doi.org/10.1007/978-3-642-14527-8_1

[12] Steven Englehardt and Arvind Narayanan. 2016. Online Tracking: A 1-Million-Site Measurement and Analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Vienna Austria, 1388–1401. https://doi.org/10.1145/2976749.2978313

[13] Steven Englehardt, Dillon Reisman, Christian Eubank, Peter Zimmerman, Jonathan Mayer, Arvind Narayanan, and Edward W. Felten. 2015. Cookies That Give You Away: The Surveillance Implications of Web Tracking. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 289–299. https://doi.org/10.1145/2736277.2741679

[14] Motahhare Eslami, Sneha R. Krishna Kumaran, Christian Sandvig, and Karrie Karahalios. 2018. Communicating Algorithmic Process in Online Behavioral Advertising. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174006

[15] European Parliament and Council of the European Union. 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council.* European Parliament and Council of the European Union. https://data.europa.eu/eli/reg/2016/679/oj

[16] Facebook. 2023. Ad Preferences | Facebook. https://www.facebook.com/adpreferences/ad_settings.

[17] Florian M Farke, David G Balash, Maximilian Golla, Markus Dürmuth, and Adam J Aviv. 2021. Are privacy dashboards good for end users? Evaluating user perceptions and reactions to Google's My Activity. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, Vancouver, B.C., Canada, 483–500.

[18] David Fifield and Serge Egelman. 2015. Fingerprinting Web Users Through Font Metrics. In *Financial Cryptography and Data Security*, Rainer Böhme and Tatsuaki Okamoto (Eds.). Vol. 8975. Springer Berlin Heidelberg, Berlin, Heidelberg, 107–124. https://doi.org/10.1007/978-3-662-47854-7_7

[19] Damien Geradin, Dimitrios Katsifis, and Theano Karanikioti. 2020. Google as a de Facto Privacy Regulator: Analyzing Chrome's Removal of Third-party Cookies from an Antitrust Perspective. https://doi.org/10.2139/ssrn.3738107

[20] E. Gkiouzepi, A. Andreou, O. Goga, and P. Loiseau. 2023. Collaborative Ad Transparency: Promises and Limitations. In *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, Los Alamitos, CA, USA, 2639–2657. https://doi.org/10.1109/SP46215.2023.10179448

[21] Richard Gomer, Eduarda Mendes Rodrigues, Natasa Milic-Frayling, and M.C. Schraefel. 2013. Network Analysis of Third Party Tracking: User Exposure to Tracking Cookies through Search. In *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. IEEE, Atlanta, GA, USA, 549–556. https://doi.org/10.1109/WI-IAT.2013.77

[22] José González-Cabañas, Ángel Cuevas, Rubén Cuevas, Juan López-Fernández, and David García. 2021. Unique on Facebook: Formulation and Evidence of (Nano)Targeting Individual Users with Non-PII Data. In *Proceedings of the 21st ACM Internet Measurement Conference (IMC '21)*. Association for Computing Machinery, New York, NY, USA, 464–479. https://doi.org/10.1145/3487552.3487861

[23] Google. 2023. About Ad Rank - Google Ads Help. https://support.google.com/google-ads/answer/1722122?hl=en.

[24] Google. 2023. My Ad Centre. https://myadcenter.google.com.

[25] Aleksandra Korolova. 2010. Privacy Violations Using Microtargeted Ads: A Case Study. In *Proceedings of the 2010 IEEE International Conference on Data Mining Workshops (ICDMW '10)*. IEEE Computer Society, USA, 474–482. https://doi.org/10.1109/ICDMW.2010.137

[26] Pierre Laperdrix, Nataliia Bielova, Benoit Baudry, and Gildas Avoine. 2020. Browser Fingerprinting: A Survey. *ACM Transactions on the Web* 14, 2 (May 2020), 1–33. https://doi.org/10.1145/3386040

[27] Adam Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. 2016. Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. In *Proceedings of the 25th USENIX Conference on Security Symposium* (Austin, TX, USA) *(SEC'16)*. USENIX Association, USA, 997–1013.

[28] LinkedIn. 2023. Advertising Data. https://www.linkedin.com/mypreferences/d/categories/ads.

[29] Ángel Merino, José González-Cabañas, Ángel Cuevas, and Rubén Cuevas. 2024. Analysis and Implementation of Nanotargeting on LinkedIn Based on Publicly Available Non-PII. , 22 pages. https://doi.org/10.1145/3613904.3642107

[30] David Nield. 2019. All the Ways Google Tracks You—And How to Stop It. *Wired* , (2019), .

[31] David Nield. 2020. All the Ways Facebook Tracks You—and How to Limit It. *Wired* , (2020), . https://www.wired.com/story/ways-facebook-tracks-you-limit-it/

[32] Privacy Sandbox. 2023. The Privacy Sandbox: Technology for a More Private Web. https://privacysandbox.com/intl/en_us/.

[33] Prolific. 2023. *Quickly Find Research Participants You Can Trust.* Prolific. https://www.prolific.co/

[34] Aafaq Sabir, Evan Lafontaine, and Anupam Das. 2022. Analyzing the Impact and Accuracy of Facebook Activity on Facebook's Ad-Interest Inference Process. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (March 2022), 1–34. https://doi.org/10.1145/3512923

[35] StatCounter. 2023. *Search Engine Market Share Worldwide.* StatCounter Global Stats. https://gs.statcounter.com/search-engine-market-share

[36] Michael Carl Tschantz, Serge Egelman, Jaeyoung Choi, Nicholas Weaver, and Gerald Friedland. 2018. The Accuracy of the Demographic Inferences Shown on Google's Ad Settings. In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society* (Toronto, Canada) *(WPES'18)*. Association for Computing Machinery, New York, NY, USA, 33–41. https://doi.org/10.1145/3267323.3268962

[37] Twitter. 2023. Ads Preferences / Twitter. https://twitter.com/settings/ads_preferences.

[38] Twitter. 2023. Keyword Targeting. https://business.twitter.com/en/help/campaign-setup/campaign-targeting/keyword-targeting.html.

[39] Giridhari Venkatadri, Elena Lucherini, Piotr Sapiezynski, and Alan Mislove. 2019. Investigating Sources of PII Used in Facebook's Targeted Advertising. *Proceedings on Privacy Enhancing Technologies* 2019, 1 (2019), 227–244. https://doi.org/10.2478/popets-2019-0013

[40] W3Schools. 2023. Browser Statistics. https://www.w3schools.com/browsers/.

[41] WhatsApp Help Center. 2023. What Information Does WhatsApp Share with the Meta Companies? https://faq.whatsapp.com/1303762270462331.

[42] Craig E. Wills and Can Tatar. 2012. Understanding What They Do with What They Know. In *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic*

Society (Raleigh, North Carolina, USA) *(WPES '12)*. Association for Computing Machinery, New York, NY, USA, 13–18. https://doi.org/10.1145/2381966.2381969

A INTERFACE CONTROLS TO ENSURE QUALITY SURVEY RESPONSES

Our implementation of the extension may allow users to quickly and randomly score interests to complete the task as soon as possible. As we have described, we recruit paid users to complete the classification task, so we wanted to be sure they properly accomplished the task. This analysis complements the experiment Section 4.1. To this end, the extension incorporates two mechanisms to reject low-effort tasks.

- **Control interests:** the extension adds pseudo-random strings which are clearly not real interests. Some of the interests which are shown are *Dmxpchtn qm 14* or *Nr mgnqpyq 70*. We expect these interests to be evaluated with 1 star, otherwise, this may be a signal the user is not reading the interests they are classifying. For every 200 real interests, we include around 5 control interests. To accept a valid submission, the mean valuation of these interests should be 2 or lower. The reason why we do not request a mean equal to 1 since there may be some miss-clicks and data is still valuable even though not all of the control interests are one. The participants were not warned about this control.
- **Interest inter-clicking time:** We run some experiments ourselves doing proper interests classification and obtain the distribution of the inter-clicking time (i.e., the time between the valuation of two consecutive interests). It seems reasonable to think that the authors will have an average inter-clicking time lower than standard users. We have implemented a conservative approach where we compute the average inter-clicking time considering the last 12 classified interests. If over those interests valuations, the mean time between two ratings is less than 590ms, we understand it is suspicious behavior. In that case, we block the extension interface for a few seconds. Once the task is resumed, if the user again shows suspicious behavior the blocking time will be increased. Our goal is to incentivize low-performing users to give up and not complete the task. At the same time, if a legit user is going too fast, we ensure it reduces a bit the speed to be sure they take sufficient time to read the interests and classify them more accurately.

B USER FATIGUE ANALYSIS

We understand that the evaluation of interests may become a repetitive task for the participants of the study, as they are requested to evaluate at least 200. Therefore, as an additional check, we measure the evolution of the interest inter-click time for all the accepted submissions, to validate the results shown in Section 4.2.2.

We obtained the inter-clicking time of each user in order. We divided these values into 10 different intervals for each user, calculating the median value. The results are shown in Figure 12. As can be seen, the first interests take a bit more time in comparison with the rest of the interests. That decrease in the times could be understood as participants became more familiar with the extension and the task. But after the 30% mark, the evaluation times became very similar.

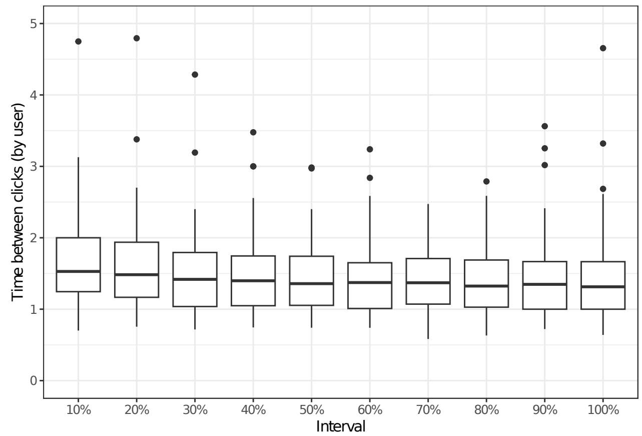

Figure 12: Median interest inter-click time

C DATA COLLECTION AND PROCESSING DETAILS

In this annex, we specify how certain data is collected, and also processed, as it was introduced in Section 3.1.1. Table 3 shows a summary of data collected and where it is used in the paper.

The extension is based on Manifest V3, which uses a service worker and several content scripts which are executed at specific pages, which allows to listen to HTTPS requests. These content scripts cannot communicate with the service worker directly, so in some occasions is necessary to inject hidden HTML elements for the service worker to be able to collect. Then, the service worker will save the data into a IndexedDB data base (local database for browsers), and finally, this data is sent to our server. The IndexedDB is necessary to avoid sending duplicate information, and to be able provide information in the UI, for the user to rank interests and see the collected information.

The list of collected items is the following:

- **Facebook Interests:** HTTPS Request towards the Interests page: parsing the response
- **Twitter Interests:** Opening the interest page in a new tab and scraping the interests.
- **LinkedIn Interests:** Parsing the JSON from https://linkedin.com/psettings/advertising/li-enterprise-product?asJson=true.
- **Google Interests:** HTTPS Request: parsing the response.
- **Browsing History:** First, using the chrome.history API: `chrome.history.search` to see the visited pages, and then `chrome.history.getVisits` to obtain the number of visits towards each URL.
- **Detecting Facebook Ads:** Listening Facebook requests with 'api/graphql' in their URL, check whether it has a SPONSORED category. In that case, extract data from HTTPS request, and inject in the Facebook Tab as hidden divs (this is because this part cannot directly communicate with the extension).

Table 3: Data Collection summary

Platform	Data	Method	Sections
Facebook	Interests	HTTPS Requests and parse response	4.2, 5, 6, 7
	Ads	Extract Ads from HTML	6, 7.3
	Ad Reasons	Open a new tab and parse HTML	6, 7.3
Twitter	Interests	Open a new tab and parse HTML	4.2, 5, 6, 7
	Ads	Listening to HTTPS responses to detect ads.	6, 7.3
	Ad Reasons	HTTPS Requests and parse response	6, 7.3
LinkedIn	Interests	HTTS Requests and parse a JSON	4.2, 5, 6, 7
	Ads	Listening to HTTPS responses to detect ads.	6, 7.3
	Ad Reasons	HTTPS Requests and parse response	6, 7.3
Google	Interests	HTTPS Requests and parse response	4.2, 5.1, 5.2
TopicsAPI	Interests	Reading the browsing history with chrome.history API	5.1, 5.2

- **Collecting Facebook Ads:** Reading the Injected Ads from the hidden divs.
- **Detecting Twitter Ads:** Listening Twitter requests until we detect a timeline object. These objects have an instructions property which defines the tweets, from there selecting the Ads and Inject them into the current tab in hidden divs.
- **Collecting Twitter Ads:** Reading the Injected Ads from the hidden divs.
- **Triggering LinkedIn Ads HTTPS requests:** To detect ads, we need to click the options button for each post; this triggers HTTPS requests which are needed to get the ID of the ads and know which posts are Ads.
- **Detecting LinkedIn Ads:** Listening to `/voyager/api/feed updates`, and inject the result
- **Collecting LinkedIn Ad Details:** Read the injected divs and then scrape the page to obtain the post details of each of the Ads detected.
- **Collecting Twitter Ad Reasons:** HTTPS Request to the `about-ads` page of each ad (using the ad ID)
- **Collecting Facebook Ad Reasons:** Opening the Ad Reasons in a new tab. If there is a button where there are more reasons why, click over it. Then obtain all reasons why the ad was targeted. Then close the tab.
- **Collecting LinkedIn Ad Reasons:** HTTPS Requests, using the CSFR LinkedIn token in the header and with AD ID.
- **Detecting Google Ads:** Reading the contents of each site to detect whether an Iframe contains a Google Ad, which is usually the case. Also trying to detect if an ad is running outside of an Iframe.
- **Collecting Google Ad Reasons:** HTTPS Request to the `why I am seeing this ad` for each Google ad.

D SELF-REPORTING BIASES

It is important to consider that there could self-reporting biases from the participants of the study: that users in the study inaccurate report information about themselves, which could lead to wrong assumptions. Therefore we need to analyze whether the responses of the users are related to the actual behaviour of the user, in order to validate the results from Section 5.1. We think that the proposed browsing history approach in this study, already sheds some light on this issue as it provides the best responses, as this method only uses behavioural activity (web browsing activity).

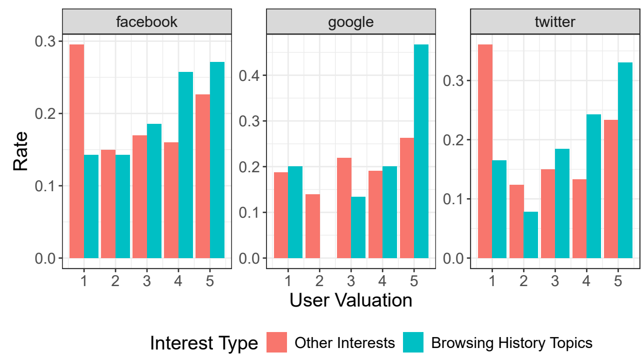

Figure 13: Browsing history interests found in other platforms against the rest of interests

It is to consider that there are browsing topics that could also be found in any of the platforms, e.g., a user may have the same interest in Facebook than in the browsing history approach. Although the method to assign the interest may be quite different, with the browsing history it only depends on user activity, whereas on a platform it may depend on several factors which we are not able to guess, as platforms do not include this information in their Ad Preferences Managers (APMs). Therefore, in the ad platforms we could separate the interests that we are sure they depend on user behaviour (as they are found in the browsing history approach), against those which could have been assigned by the platforms for any reason.

In particular, we have found 188 interests of users that were both in the users browsing history approach and in other platforms, which were evaluated. The results are shown in Figure 13. What we can observe here is how interests which we are use are from the behaviour of a user perform much better than the other ones. If we were to find any generalized self-reporting bias, we will find a much more similar distribution.

Table 4: Category table using the interests that were used to target ads in Facebook, LinkedIn and Twitter

Category	Facebook	LinkedIn	Twitter	Total
Technology/Computing	449	163	246	858
Business/Finance	130	102	69	301
Travel/Leisure	231	3	19	253
Entertainment	165	2	59	226
Sports	118	0	43	161
Social/Cultural	94	10	15	119
Science/Nature	22	19	57	98
Fashion/Apparel	68	1	21	90
Food/Dining	75	0	5	80
Health/Wellness	26	11	8	45

E AD DATASET DIVERSITY

In this annex we show the diversity of the ads that were collected by the 154 participants of the study as was described in Section 4.2.

E.1 Categories from Interests

Considering the wide range of interests that are used to target ads, we have grouped them in different categories, which are the same for all platforms. The results can be seen in Table 4. For instance, on Facebook and Twitter, we found that the category Entertainment is very common whereas not on LinkedIn, in which Business/Finance is more widely found, as it could be expected. There are also other categories such as Sports or Food/Dining which were not found with the mapping of interests on LinkedIn.

E.2 Topics from Landing Pages

From every ad we extracted the landing page, and from these we got the domain name and put it through the Topics API classifier, this classifier may give from 0 to 3 topics, and we then aggregate these topics. Table 5 shows the 20 most common topics from each platform. In particular, our ads cover 313 different topics, being Facebook the most diverse platform in that sense, with 280 different topics; while LinkedIn is the platform with least amount of topics with 84. The most common topics across all platforms are: Arts & Entertainment (789), News (445) and Business & Industrial (364). LinkedIn is the only platform where Arts & Entertainment is not the most common category, and some categories are more relevant on LinkedIn than in the rest of the platforms (for instance Colleges & Universities which is the third most common topic on this platform).

This analysis uses all ads that were collected, not just the ones that used interests as targeting options.

E.3 Top Advertisers

Next, we display which are the top advertisers in our dataset, this information is available from Twitter, Facebook, and LinkedIn, as we retrieve who "posted" the ad, i.e., their profile name. The results can be checked in Table 6. This table shows the diversity of the most common advertisers, which could be grouped in different categories such as Streaming Services (*Disney+*, *HBO Max*, *Bloomberg*), mobile phone manufacturers (*Samsung*, *Huawei*, *Apple*), Food industry (*Coca-cola*, *McDonald's*), clothing (*SHEIN*, *Nike)*, video games (*DITOGAMES*, *Hero Wars Web*), learning companies (*Udemy*) or even financial services (*Interactive Brokers*, *Western Union*). These results include all ads that were collected (not only those with interest based targeting), we encounter over 8k different advertisers, but from the majority of them (72.93%) we only collected a single ad.

F CODE AVAILABILITY

Code for the browser extension and the analyses can be found in the following GitHub repository: https://github.com/fcaravaca/OverprofilingPETS.

Table 5: Table with the 20 most common topics using the ads' landing pages.

Twitter	Ads	Google	Ads	Facebook	Ads	LinkedIn	Ads
Arts & Entertainment	165	Arts & Entertainment	162	Arts & Entertainment	441	Business & Industrial	27
News	97	Business & Industrial	84	News	267	Arts & Entertainment	21
Computer & Video Games	72	News	74	Business & Industrial	204	Colleges & Universities	12
Online Communities	54	Shopping	69	Shopping	176	Internet & Telecom	10
Business & Industrial	49	Autos & Vehicles	62	Computer & Video Games	156	News	7
Internet & Telecom	48	Travel & Transportation	48	Online Communities	140	Online Communities	7
Shopping	38	TV & Video	48	Education	106	Business & Productivity Software	6
Music & Audio	29	Internet & Telecom	44	People & Society	103	Advertising & Marketing	5
Law & Government	29	Hotels & Accommodations	44	Travel & Transportation	102	Software	5
Computers & Electronics	27	Education	42	Internet & Telecom	88	Autos & Vehicles	5
Software	27	Online Communities	42	Music & Audio	87	Computer & Video Games	4
Investing	24	Computer & Video Games	39	Colleges & Universities	85	Music & Audio	4
Colleges & Universities	24	Local News	31	Local News	82	Job Listings	4
TV & Video	24	Music & Audio	30	Cooking & Recipes	80	Shopping	4
Local News	23	Banking	28	Apparel	77	Industrial Materials & Equipment	4
People & Society	23	Vehicle Shopping	28	Sports	69	Banking	4
Travel & Transportation	23	Colleges & Universities	26	Computers & Electronics	68	Movies	3
Education	21	People & Society	26	TV & Video	64	Computers & Electronics	3
Sports	21	Apparel	26	Hobbies & Leisure	63	Investing	3
Banking	19	Software	24	Autos & Vehicles	59	Education	3

Table 6: Top 50 advertisers in the dataset

Platform	Author	Ads	Platform	Author	Ads
Twitter	Nordace	114	Twitter	Goodergear	23
LinkedIn	LinkedIn Ads	85	Twitter	Google UK	23
Twitter	DITOGAMES	77	LinkedIn	Microsoft	22
Facebook	Hero Wars Web	65	Twitter	Kia UK	22
Twitter	Interactive Brokers	61	Facebook	Coca-Cola	21
Facebook	Disney+	51	Twitter	HBO Max Latinoamérica	21
Facebook	SHEIN	47	Facebook	McDonald's	20
Facebook	Samsung	46	Twitter	Kairosoft	20
Twitter	HPE	44	Twitter	Apple	19
Facebook	Google Ads	39	Twitter	Strayed Lights	19
Twitter	bridgecrew	38	Facebook	Games4all	18
Facebook	Adobe Photoshop Lightroom	37	Facebook	Nike	18
Twitter	Ducon.space	37	Twitter	Bloomberg	18
Facebook	Adobe Photoshop	34	Twitter	Luzido	18
Facebook	HBO Max	34	Facebook	Adobe Creative Cloud	17
Facebook	Zapier	34	Facebook	Displate	17
Facebook	OKO.press	31	Facebook	Żabka Polska	17
LinkedIn	LinkedIn Learning	30	LinkedIn	Grammarly	17
Twitter	Huawei	29	Twitter	Aspose	17
Twitter	Miahcombat	29	Twitter	Omaze UK	17
Twitter	Survey Compare UK	29	Facebook	Domestika	16
LinkedIn	LinkedIn Talent Solutions	27	Facebook	Udemy	16
Facebook	Media Expert	25	Facebook	Western Union	16
Twitter	Raid: Shadow Legends	25	Twitter	Altroconsumo	16
Facebook	Pipedrive	23	Twitter	Diesel Experts	16