

**Document Version**

Final published version

**Licence**

CC BY

**Citation (APA)**

Mousavi, S., Dash, A., Zannettou, S., & Gummadi, K. P. (2026). Does Ad-Free Mean Less Data Collection? An Empirical Study of Platform Data Practices and User Expectations. In *WWW 2026 - Proceedings of the ACM Web Conference 2026* (pp. 3287-3298). (WWW 2026 - Proceedings of the ACM Web Conference 2026). Association for Computing Machinery (ACM). <https://doi.org/10.1145/3774904.3792558>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.  
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

**Sharing and reuse**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.



# Does Ad-Free Mean Less Data Collection? An Empirical Study of Platform Data Practices and User Expectations

Sepehr Mousavi

Max Planck Institute for Software Systems  
Saarbrücken, Germany  
smousavi@mpi-sws.org

Savvas Zannettou

TU Delft  
Delft, Netherlands  
s.zannettou@tudelft.nl

Abhisek Dash

Max Planck Institute for Software Systems  
Saarbrücken, Germany  
adash@mpi-sws.org

Krishna P. Gummadi

Max Planck Institute for Software Systems  
Saarbrücken, Germany  
gummadi@mpi-sws.org

## Abstract

Online platforms increasingly offer "paid" *ad-free subscriptions* as an alternative to the traditional "free" *ad-based model*. The transition to ad-free models ostensibly removes advertising as a key justification for data processing under the GDPR. So, normatively, platforms should collect less user data. However, platforms may justify continued data collection as a means to provide an improved, personalized experience. This tension between privacy principles and platform incentives raises a critical underexplored question: do data collection practices vary between ad-free and ad-based subscription models?

In this paper, we shed light on this important privacy issue by investigating the alignment between platform data collection practices and related user expectations. With respect to data collection process, our analyses of data exports from three major online platforms — Instagram, Facebook, and X — reveal that these platforms continue to retain or collect some ad-related data, even in ad-free subscriptions. With respect to user expectations, our survey among 255 participants on Prolific reveals that 69% of the participants normatively expect data collection to be reduced, indicating their expectation of improved digital privacy in an ad-free model. However, when asked what they think actually happens, 63% of these participants believed that platforms would still collect about the same amount of data, highlighting skepticism about platform practices. Our findings not only indicate a significant disconnect between data practices and normative user expectations, but also raise serious questions about platform compliance with core GDPR principles, such as purpose limitation, data minimization, and transparency.

## CCS Concepts

- Security and privacy → Social network security and privacy;
- Social and professional topics → Privacy policies.



This work is licensed under a Creative Commons Attribution 4.0 International License.  
WWW '26, Dubai, United Arab Emirates  
© 2026 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-2307-0/2026/04  
<https://doi.org/10.1145/3774904.3792558>

## Keywords

Measurement analysis, Policy and legal aspects, GDPR, Privacy policies, Data download packages

### ACM Reference Format:

Sepehr Mousavi, Abhisek Dash, Savvas Zannettou, and Krishna P. Gummadi. 2026. Does Ad-Free Mean Less Data Collection? An Empirical Study of Platform Data Practices and User Expectations. In *Proceedings of the ACM Web Conference 2026 (WWW '26)*, April 13–17, 2026, Dubai, United Arab Emirates. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3774904.3792558>

### Resource Availability:

The source code of this paper has been made publicly available at <https://doi.org/10.5281/zenodo.18368052>.

## 1 Introduction

Online platforms such as Instagram, Facebook, and X (formerly Twitter) are increasingly offering users a choice to access their services in two structurally different models, namely *ad-based* and *ad-free*. In the ad-based model, platforms generate revenue from advertisers by serving ads to their users, who receive free access to their services. In contrast, the ad-free model allows platforms to generate revenue directly from their users in the form of subscription fees, without interrupting their experience with ads. Put differently, in ad-free service models, users are the *customers* of the platform services, while in ad-based service models, users are the *commodities* offered by the platforms to their advertising customers.

In this paper, we focus on *comparing the data gathering practices* of very large online platforms under these two service models. While the privacy research community has extensively examined various aspects of data collection by online platforms, little attention has been given to how data collection practices change when switching between the two models.

Comparing a platform's data collection practices under ad-based and ad-free models is important both from a regulatory as well as an end-user perspective. From a regulatory perspective, the General Data Protection Regulation (GDPR) mandates a set of core principles for data processing, such as purpose limitation and data minimization (Article 5 (1)) [16]. In both models, platforms have an incentive to collect user data to provide personalized services to users. However, since serving targeted ads is no longer a purpose in the ad-free model, platforms should collect less user data than in

the ad-based model to adhere to the purpose limitation and data minimization principles. Furthermore, in practice, the lack of a clear distinction between personal data for targeted advertising and service personalization may allow platforms to justify any data collected based on their legitimate interest (Article 6 (1)(f)) [16].

Beyond investigating platform practices, understanding user expectations is equally important as they impact how users choose between the two service models. Ascertaining users' normative expectations (i.e., what they believe platforms *should* do), with respect to data gathered about them, would reveal whether users expect better privacy in an ad-free model, i.e., *whether their willingness to pay subscription fees is to avoid ads or is based on expectations of protecting their privacy*. Similarly, assessing their descriptive expectations (i.e., what they believe platforms *will* do) may reveal doubts about whether platforms will meet users' normative expectations, which could affect the adoption of ad-free models. Finally, checking user expectations against actual platform practices would help users make more informed service choices, as they can validate their expectations or correct their misperceptions.

**Current work:** Against this background, in this work, we attempt to answer the following crucial questions:

**RQ1:** *How do platforms change their data collection practices in an ad-free model as compared to an ad-based model?*

**RQ2 (a):** *What are end-users' normative and descriptive expectations of personal data collection practices adopted by platforms?*

**RQ2 (b):** *How do these practices affect users' choice of service model?*

To answer the first question, we collected data from three major online platforms that provide their services in ad-based and ad-free subscription models: Instagram, Facebook, and X. First, we created accounts on these platforms in their ad-based subscription model and consumed content online. Then, we switched to an ad-free model for one month and continued to use the services. Next, we exercised our *Right of access* (GDPR Article 15(3)) to personal data by requesting a copy of all the data gathered by the platforms. We compared the collected data exports during ad-based subscription and ad-free subscription to investigate the differences in data collection practices.

To answer the second question (collectively), we surveyed 255 participants on the online crowd-sourcing platform Prolific.<sup>1</sup> We asked these participants about their normative expectation (*should* platforms collect less data in ad-free subscription) and descriptive expectation (*would* platforms collect less data in ad-free subscription) on data collection by platforms across different subscription models. Additionally, we also asked our participants if platforms continuing a similar amount of data collection in the ad-free model (as compared to the ad-based model) would affect their decision to choose an ad-free model.

**Main findings:** Next, we summarize the answers to our RQs.

**Answer to RQ1:** Despite removing some ad-related data under the ad-free model, all platforms continue to retain or collect certain personal data used exclusively for advertising purposes.

- Instagram and Facebook remove some ad-related data once users subscribe to the ad-free model.

- However, under the ad-free model, all three platforms continue to retain or collect some personal data that are exclusively used for advertising purposes.

**Answer to RQ2:** We observe a significant gap between the normative and descriptive expectations of our survey participants.

- Majority (69%) of participants normatively expect less data should be collected in the ad-free model, indicating they expect better digital privacy in the ad-free model.

- However, only 17% of participants descriptively expect less data to be collected in the ad-free model.

- Finally, 34% of participants mentioned that the continued collection of a similar amount of data would deter them from switching to an ad-free subscription model.

**Implications:** In summary, our findings reveal a misalignment between user expectations, platform practices, and regulatory requirements. Users are *not* getting the privacy they expect, even in ad-free models. Platforms often fall short of GDPR obligations, and potential poor coordination between legal, engineering, and product teams may prevent the implementation of compliant data practices. These gaps point to the potential role of clearer regulatory guidance and better coordination across stakeholders.

## 2 Background & Related Work

In this section, we contextualize our work by (a) setting the necessary legal background, (b) highlighting some of the key non-compliance of platforms with regulations, and (c) reporting prior research in computer science related to the current work.

### 2.1 Legal Background

While several provisions of the General Data Protection Regulation (GDPR) [16], Digital Services Act (DSA) [17], and Digital Markets Act (DMA) [5] are broadly relevant, covering all of them is beyond the scope of this work. Therefore, we focus on the GDPR provisions that are most pertinent to our study.

**Principles of data processing:** The processing of personal data by online platforms within the European Union (EU) is primarily governed by the GDPR, which establishes foundational *principles* of data processing. Central to these principles is Article 5, mandating that personal data be processed lawfully, fairly, and transparently (Article 5(1)(a)). Furthermore, purpose limitation (Article 5(1)(b)) requires data to be collected for specified, explicit, and legitimate purposes and not further processed in a manner incompatible with those purposes. Finally, data minimization (Article 5(1)(c)) dictates that data collection must be adequate, relevant, and limited to what is necessary in relation to the purposes for which it is processed.

**Lawfulness of data processing:** Article 6 of the GDPR mandates the core principles of data processing by listing valid legal bases. Among these bases, *legitimate interests* (Article 6(1)(f)) is the most pertinent to platform operations concerning advertising and personalization. Platforms might invoke *legitimate interests* to justify ongoing data collection for service personalization even in ad-free setting. However, such justification requires a careful three-part assessment: identifying a legitimate interest, ensuring the processing is necessary to achieve it, and critically balancing this against the fundamental rights and freedoms of the data subject – a balance that, we posit, *may shift when users are directly paying for a service*.

<sup>1</sup><https://prolific.com>

**Rights of access by users:** The GDPR also enables users to exercise rights to verify whether the principles and lawfulness of data processing are upheld by platforms. To this end, a cornerstone for such user empowerment is the *Right of Access* (Article 15). Specifically, Article 15(3) grants the right to obtain a copy of the personal data undergoing processing, enabling users and researchers alike to scrutinize actual data being processed.

## 2.2 Violations by Tech Platforms

Major technology platforms have repeatedly faced regulatory actions in the EU, highlighting systemic concerns over data practices, particularly in advertising and personalization. For instance, as a violation of the GDPR, Google was fined in 2019 for not properly disclosing to users about its data collection and processing practices in the context of personalized ads [35]. Amazon was also fined by the data protection body of Luxembourg for manipulating user consent in their advertising practices [32]. These cases indicate a robust regulatory environment in which personal data collection for profiling, personalization, and advertising remains under constant scrutiny, regardless of the service model.

Meta's introduction of a 'consent or pay' model for Facebook and Instagram in the EU highlights the tension between data-driven business models and privacy regulation [1]. Increasingly adopted by platforms [28, 31], this model offers users a choice between consenting to extensive data tracking used for personalization and advertising [7, 10], or paying to limit such processing. However, Data Protection Authorities questioned whether consent obtained under this scheme is freely given as required by GDPR Article 7, noting that subscription fees may unduly pressure users. In response, the European Data Protection Board emphasized that valid 'consent or pay' models must ensure a genuinely voluntary and informed choice [14]. In 2025, the European Commission fined Meta for non-compliance of this model with GDPR and DMA consent requirements [4, 5, 16].

## 2.3 Related Work

We contextualize our work within two strands of prior research: (1) user perception of data collection and targeted advertising, and (2) platform audits using GDPR data exports.

**User perception of data collection and targeted advertising:** Users recognize content personalization relies on data collection and view online tracking as a way for platforms to improve their services [12, 23]. Moreover, they often view targeted advertising as beneficial for discovering relevant content and enhancing their overall experience [26, 33]. At the same time, concerns persist about how personal data are collected and used. For instance, many users perceive targeted advertising as privacy-invasive or reliant on deceptive design practices [11, 15, 24, 27, 36, 39]. Furthermore, users believe targeted advertising can contribute to the spread of harmful content, among other issues [25, 33, 34, 42]. Our work builds on this line of research by studying users' normative and descriptive expectations of platform data collection across subscription models. **Platform audits using GDPR data exports:** Researchers have been leveraging GDPR Right of Access to recruit participants who voluntarily donate their data exports. Using these donated GDPR data exports, prior works have conducted research on data export

reliability [22], ad targeting mechanisms [38], recommendation algorithms [37], user behavior analysis [41], and societal topics ranging from news and politics to personal health and safety [8, 9, 20, 21, 40]. We contribute to this line of research by employing data exports to examine platform data collection practices across subscription models.

## 3 Platform Data Gathering Practices Across Subscription Models

In this section, we investigate data collection practices of online platforms across subscription models and present our findings.

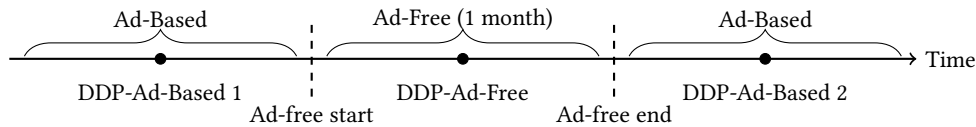
### 3.1 Data Collection Methodology

To study data collection practices of online platforms across various subscription models, we need a method that allows us to obtain personal data that online platforms collect, store, and process. This task is challenging because online platforms are generally non-transparent about their data collection practices, and the information disclosed in privacy policies is often incomplete.

To overcome these challenges, we leverage the power of GDPR's Right of Access by the data subject (Article 15), which entitles individuals to obtain a copy of their data being held and processed by online platforms. Platforms typically provide this information in the form of a Data Download Package (DDP). These DDPs provide a comprehensive view of what a platform collects, stores, and processes about the user who exercised their GDPR right. Crucially, DDPs capture data across devices (e.g., web and mobile), as well as information gathered through third parties, offering an invaluable perspective that is otherwise inaccessible. By grounding our methodology in the information that platforms are legally required to disclose, we ensure that our analysis relies on data they demonstrably hold, thereby avoiding ambiguities around whether particular data were collected or retained. For these reasons, we argue that DDPs are a suitable and reliable source for systematically studying data collection practices.

Building on this foundation, we designed a procedure to obtain comparable data from both ad-based and ad-free subscription modes. Figure 1 depicts a schematic overview of the steps we took for collecting such data. First, two co-authors of this work created accounts in the platforms' default ad-based model from within the European Union. We then subscribed to the ad-free model for one month and continued to use the service under this model. After the subscription expired, we returned to the default ad-based model. At each step of this procedure, we obtained a DDP: an initial ad-based DDP (DDP-Ad-Based-1) before switching to the ad-free subscription, an ad-free DDP (DDP-Ad-Free) three weeks into the subscription period, and a second ad-based DDP (DDP-Ad-Based-2) one week after returning to the ad-based model.

**Platforms in this work:** We focus on Instagram, Facebook, and X (formerly Twitter) in our study. We chose these online platforms for multiple reasons. First and foremost, all these platforms provide ad-based and ad-free subscription models, which is essential for our comparative analysis. Second, they operate their own advertising infrastructures rather than relying on external providers, meaning that the data we examine is generated and processed within the platforms' own ecosystems. This makes them particularly relevant



**Figure 1: Timeline of requested Data Download Packages (DDPs) across ad-based and ad-free subscription models of an online platform. Initially, we start by obtaining an ad-based DDP (DDP-Ad-Based-1). Then, we subscribe to the platform’s ad-free model and request ad-free DDP (DDP-Ad-Free) throughout our subscription period. Finally, we obtain a second ad-based DDP (DDP-Ad-Based 2) after our ad-free subscription ends and the account switches back to the ad-based model.**

**Table 1: Number of Ad-only, Non-ad, and Mixed files.**

Platform	#Files across the three categories			Total
	Ad-only	Non-ad	Mixed	
Instagram	7	21	51	79
Facebook	12	54	137	203
X	6	35	42	83

for studying how subscription models affect data collection. Furthermore, DSA designates these platforms as Very Large Online Platforms (VLOPs) [17]. Therefore, they are subject to heightened regulatory obligations and scrutiny.

### 3.2 Categorizing Files Contained in DDPs

The DDPs from Instagram, Facebook, and X contain a large number of files, covering a wide variety of personal data categories. A more detailed characterization of what is included in the DDPs is available in Appendix B. While comprehensive, the diversity of these files makes systematic analysis challenging, as not all data collection serves the same purpose. For instance, some files reflect information used exclusively for advertising, others pertain to personalization or platform functionality, and others contain data relevant to both. Distinguishing between these categories is essential for our analysis: without it, we cannot determine whether ad-free subscriptions meaningfully change the scope of advertising-related data collection or whether platforms continue to collect comparable information for other purposes.

To enable such comparisons, we systematically categorized each file in the DDPs according to its relationship to advertising. Two co-authors independently categorized each file as follows: (1) **Ad-only** (the file is exclusively related to advertising and is not used for other non-advertising purposes), (2) **Non-ad** (the file is unrelated to advertising), and (3) **Mixed** (the file contains data that can be used for both advertising and other non-advertising purposes, such as content personalization). In cases of disagreement, a third co-author independently made the final decision. The categorization relied on multiple sources of information, including (a) the platforms’ privacy policies; (b) README or explanatory documents provided within the DDPs (only applicable to X); (c) the folder structure used to organize files; and (d) file descriptions or embedded metadata. The three co-authors conducting this classification are experienced researchers familiar with DDPs. In total, we categorized 365 files across the DDPs. The two co-authors agreed on 336 files with a Fleiss’ Kappa score of 0.85, indicating an almost perfect

inter-annotator agreement between them [19]. The third co-author then independently annotated the 29 files with disagreements to reach the final classification (based on majority agreement). Table 1 presents an overview of our DDP categorization.

### 3.3 Analyzing DDPs

Our analysis focuses on systematically comparing pairs of DDPs obtained from the same platform under different subscription models. The comparison operates at multiple analytical levels. At the file system level, we examine the overall structure of the DDPs, identifying differences in the presence, absence, or modification of files and directories. At the data structure level, we traverse JSON files to detect changes in their schema and organization. Together, these analyses allow us to capture both coarse and fine-grained differences in data collection practices across subscription models.

To support this process, we developed a browser-based analyzer tool that automated these comparisons and generated a structured JSON report summarizing the observed differences. The tool is generalizable, as the implementation allows it to handle DDPs from multiple online platforms with almost no platform-specific customization. We implemented the tool in JavaScript with client-side processing to ensure all data remains local to the user’s machine. To facilitate future research, we make our analyzer tool available.<sup>2</sup>

### 3.4 Findings

Here, we present findings from our analysis. First, we begin with insights drawn from Ad-only files, then discuss Mixed files, and conclude with Non-ad files.

**Ad-only files:** Ideally, once ads are no longer served, Ad-only files should demonstrate reduced data collection. This may manifest either through their removal from the DDPs or by ceasing additional data collection during the ad-free subscription period. To this end, we first examine whether Ad-only files are removed, and then assess whether the files that persist through both ad-based and ad-free periods exhibit reduced data collection.

Table 2 provides an overview of all Ad-only files from Instagram, Facebook, and X, and their presence in DDP-Ad-Based 1 and DDP-Ad-Free. We observe that Instagram and Facebook removed 4 and 5 Ad-only files in their DDP-Ad-Free, corresponding to 57% and 42% of all Ad-only files in their DDPs, respectively, whereas X did not remove any. This removal indicates that subscribing to the ad-free model stops Instagram and Facebook from processing the information included in these files. To confirm, we compared DDP-Ad-Based 1 and DDP-Ad-Based 2 and found a lack of significant

<sup>2</sup><https://github.com/sepehroumouavi/subscription-data-practices>

**Table 2: Presence of Ad-only files of Instagram, Facebook, and X in both DDP-Ad-Based 1 and DDP-Ad-Free.**

Platform	File	Present in DDP-Ad-Based 1	Present in DDP-Ad-Free
Instagram	advertisers_using_your_activity_or_information.json, ads_about_meta.json, ads_clicked.json, ads_viewed.json	✓	✗
	other_categories_used_to_reach_you.json, subscription_for_no_ads.json, consents.json	✓	✓
Facebook	advertisers_using_your_activity_or_information.json, ads_about_meta.json, advertisers_you've_interacted_with.json, ads_interests.json, ads_personalization_consent.json	✓	✗
	ad_preferences.json, meta_ad_library_accounts.json, other_categories_used_to_reach_you.json, subscription_for_no_ads.json, your_consent_settings.json, your_meta_business_suite_guidance_interactions.json, consents.json	✓	✓
	ad-engagements.js, ad-impressions.js, ad-mobile-conversions-attributed.js, ad-mobile-conversions-unattributed.js, ad-online-conversions-attributed.js, ad-online-conversions-unattributed.js	✓	✓

overlap between the two instances, suggesting that the contents of these files were indeed deleted after the ad-free subscription.

In addition, ideally, the remaining Ad-only files should cease data collection, particularly for user profiling. To assess this, we examined whether such files continued collecting profiling-related data in DDP-Ad-Free. Our analysis found that two Ad-only files on Meta platforms (one present in both Instagram and Facebook, and one unique to Facebook) and one Ad-only file on X continued to retain or collect profiling data. These correspond to 14%, 17%, and 17% of Ad-only files on Instagram, Facebook, and X, respectively.

In the case of Meta platforms, the file *other\_categories\_used\_to\_reach\_you.json*, present in both Instagram and Facebook, contains automatically inferred categories used by advertisers to target their audience [6]. We observed that this file retained its existing content in DDP-Ad-Free, but showed no signs of new data collection. The retention of these categories allows Meta platforms to continue leveraging previous ad-related profiling data, even if they cease new data collection in the ad-free model. Also, the file *ad\_preferences.json* in Facebook contains information related to advertising preferences. This includes fields that have data on ads and advertisers hidden by the user, along with other user preferences. Our analysis of DDP-Ad-Free revealed new entries in the hidden ads and hidden advertisers fields, indicating the propagation of profiling based on actions taken during the ad-based subscription period. Conversely, we did not observe new data collection by the remaining Ad-only files in the Meta platforms. Overall, these findings illustrate that, despite subscription to the ad-free model, Meta platforms continue to retain or collect some profiling data.

For X, the *ad-mobile-conversions-unattributed.js* file stores mobile app events associated with an account that may become attributable to an advertisement in the future. During the ad-free subscription period, this file recorded events (e.g., install and login) from two mobile applications, while the other Ad-only files showed no new data collection. This indicates that X continues to collect exclusive ad-related profiling data even under the ad-free model.

**Mixed files:** Mixed files contain data used for both advertising and non-advertising purposes (e.g., content personalization). Hence, platforms are not necessarily required to alter their data collection practices for these files once users subscribe to the ad-free model.

Indeed, we did not find any such file being removed from DDP-Ad-Free across the three platforms. However, Mixed files may include fields exclusively related to advertising, which ideally should no longer be processed once an account switches to the ad-free model. To systematically assess this, we identified ad-related fields and examined whether they continued to be processed in DDP-Ad-Free. To identify ad-related fields, we searched among Mixed files for keys containing the substring “ad”. This yielded 2 fields in Facebook and 5 in X, while none was found in Instagram.

In the case of Facebook, we identified *advertiser\_id* and *Google advertiser ID* fields inside *mobile\_devices.json* and *your\_devices.json* files, respectively. Moreover, we observed that *Google advertiser ID* was empty in all DDP instances. However, *advertiser\_id* retained its value in DDP-Ad-Free. This identifier, which is associated with each mobile device, allows advertisers to target their audience and can be used to collect behavioral data of users [2]. Although no ads are served under the ad-free model, the presence of this identifier still allows user tracking.

For X, we identified the *audienceAndAdvertisers* field inside the *personalization.js* file, which contains four nested ad-related fields: *lookalikeAdvertisers*, *advertisers*, *doNotReachAdvertisers*, and *catalogAudienceAdvertisers*. These nested fields collectively contain information exclusively related to advertising purposes. Moreover, they were all empty across DDP instances, except for *lookalikeAdvertisers*. In particular, *lookalikeAdvertisers* in DDP-Ad-Free retained a non-empty list of advertisers that own look-alike audiences that include our X accounts. Furthermore, these advertiser lists were identical between DDP-Ad-Based 1 and DDP-Ad-Free. This raises the question of why X continued to include our accounts in look-alike audiences of advertisers despite the ad-free subscription. It is essential to note that X provides advertisers with estimates of the “reach” (i.e., size) of their custom audiences. If these estimates include a substantial number of ad-free users who cannot see ads, X risks misleading advertisers.

**Non-ad files:** Naturally, subscribing to the ad-free model should not affect data collection practices of Non-ad files. Indeed, our analysis confirms this expectation. Specifically, we did not find any such file being removed from DDP-Ad-Free across the three platforms. Moreover, we observed no meaningful or significant changes across

the three platforms in how such files retain and collect data. These observations suggest that the ad-free subscription model does not influence data collection practices of Non-ad files.

#### Main takeaways:

- 👉 Instagram and Facebook, upon subscription to the ad-free model, partially reduce their data collection practices by removing 57% and 42% of all Ad-only files in their DDPs, respectively, whereas X does not remove any such files (see Table 2).
- 👉 However, we find evidence that all three platforms continue to retain or collect some personal data that are specifically categorized for advertising purposes under the ad-free model.

## 4 User Expectation on Data Gathering Practices Across Subscription Models

Having examined platform data collection practices, we next investigate user expectations by conducting a survey on perceptions of data collection under different subscription models.

### 4.1 Survey Design

Our survey comprised two parts. The first examined users' normative and descriptive expectations regarding data collection practices across subscription models. The second assessed whether differences or reductions in data collection affect participants' willingness to select an ad-free subscription. The full questionnaire is provided in Appendix C, and we detail the survey design below.

**Expectation of reduced data collection:** First, we investigated participants' normative (what they believe platforms *should* do) and descriptive (what they believe platforms *would* do) expectations of data collection practices adopted by platforms based on the ad-based and ad-free subscription models. To this extent, we asked our participants the following two questions:

- *Q\_Normative*: "If you subscribe to an ad-free plan of a platform, should the platform collect less data about you?"
- *Q\_Descriptive*: "If you subscribe to an ad-free plan of a platform, do you think that the platform would collect less data about you?"

We framed the above two questions as multiple-choice questions with the following three choices: (1) *Yes*, (2) *No*, and (3) *I'm not sure*. Furthermore, to account for potential order-effect bias caused by the ordering of these two questions, we randomly shuffled these two questions for each participant [30]. Also, to gain deeper insights into participants' reasoning, we placed a free-response question after each of these two questions. Specifically, we asked our participants to justify their choice: "Please justify your above choice."

**Choice of subscription model:** The second part of our survey examined whether data collection practices of online platforms across subscription models might affect participants' willingness to choose an ad-free subscription. We asked our participants:

- *Q\_Influence*: "Would it affect your decision to choose an ad-free plan if the platform continued collecting the same amount of data as it does in the ad-based plan?"

Similar to the first part of the survey, we also framed *Q\_Influence* as a multiple-choice question with the same three choices and accompanied it with a free-response question to understand the reasoning of our participants.

### 4.2 Participant Recruitment

We recruited three distinct user groups: (1) legal experts, (2) technology experts, and (3) members of the general public. Legal experts provide us with insights into normative and regulatory aspects of user privacy and data collection practices. Technology experts offer informed views on what is technically feasible as well as what is being commonly implemented by online platforms. Finally, members of the general public reflect end-user perceptions and concerns with respect to data collection practices. By surveying these three groups, we aimed to understand how expectations vary across legal, technical, and end-user perspectives.

We used the online crowd-sourcing platform Prolific<sup>3</sup> and chose multiple screening criteria to recruit our participants. To apply the above expertise-based distinction, we used Prolific's "Work Function" screener, which identifies participants' professional roles within their organizations. This screener includes 21 work functions: we selected "Legal" for legal experts, "Engineering (e.g., software)" and "IT / Information Networking / Information Security" for technology experts, and the remaining 18 for members of the general public. Since the legal basis of our work relies on the European digital regulation (e.g., GDPR), we targeted users who live in the European Union, where these regulations are applicable. Moreover, as our questionnaire was in English, we recruited participants who were fluent in English. We were interested in recruiting active and reliable participants. Hence, we looked for users who had at least 20 Prolific submissions with an approval rate of at least 95%.

We conducted our survey in May 2025. The median survey completion time was 10 minutes, and we reimbursed each participant with £2, which is equivalent to £12 per hour, more than the recommended remuneration rate by Prolific.<sup>4</sup> We aimed to recruit 100 participants for each of the three user groups. While we successfully reached this target for both tech experts and members of the general public, we were only able to recruit 85 legal experts, resulting in a total of 285 participants. This shortcoming was due to the limited number of Prolific users who satisfied our screening requirements. In particular, only 190 active Prolific users met our screening criteria for the legal experts group. Therefore, our recruitment capacity was constrained for this specific user group. We provide demographic characteristics of our participants in Appendix A.

**Ethical Considerations:** The Ethical Review Board (ERB) of Saarland university approved our study. Participants took part in our survey voluntarily and with informed consent. Furthermore, pursuant to the ERB guidelines, we did not collect any personally identifiable information (PII) from our participants.

### 4.3 Survey Validation and Qualitative Coding of Free-response Questions

We improved our survey after performing pilots with two of our colleagues and obtaining their feedback. Then, we conducted three pilot studies on Prolific, each with 40 participants. This is a much higher sample size than the recommended baseline for survey validation [29]. We ensured that our pool of participants in these pilot studies differed from that in our final study.

<sup>3</sup><https://prolific.com>

<sup>4</sup><https://researcher-help.prolific.com/en/article/9cd998>

We used qualitative open coding methodology to analyze the free-response questions and identify recurring themes in participants' explanations. We grouped responses by their corresponding multiple-choice answer (Yes, No, or I'm not sure), resulting in nine sets. Then, the first author of this work reviewed and coded these sets. We provide further details and examples in Appendix D.

Furthermore, in line with prior work, we relied on answers to the three free-response questions in our survey to detect participants who were inattentive [13, 18]. We compared each participant's multiple-choice and free-response answers, and flagged a participant as inattentive if the free-response answer (1) did not match the multiple-choice answer, (2) was missing, or (3) contained only a couple of words. In the end, we found 30 inattentive participants and discarded their responses. Hence, we analyzed the responses of the remaining 255 participants, consisting of 78 legal experts, 90 technology experts, and 87 members of the general public.

#### 4.4 Survey Results

This subsection reports the results of our descriptive survey which aims to quantify the gap between users' normative and descriptive expectations. Hence, we focus on characterizing these distributions at the population level and do not conduct inferential statistical analyses to support our main claims.

**Normative expectation of reduced data collection:** Figure 2a shows the distribution of responses to Q\_Normative. We observe that, overall, 69% of participants believed that online platforms should collect less data from users subscribed to the ad-free model compared to the ad-based model. Moreover, we see that the normative expectations of participants are closely aligned across the three user groups. Specifically, 68% of legal experts, 72% of technology experts, and 67% of the general public responded positively to Q\_Normative. The substantial majority of respondents who normatively expect less data collection also signal a desire for stronger digital privacy beyond just ad removal. This highlights a clear gap between users' normative expectations and the actual platform practices observed in Section 3.

To further understand the reasoning of participants, we analyzed their answers to the free-response question that followed Q\_Normative. Since the normative expectations of participants are closely aligned across the three user groups, we combined the answers to the free-response question across the three user groups. Among the participants who responded positively to Q\_Normative, 37% ( $n = 65$ ) mentioned that *data is no longer needed for targeted advertising*, 28% ( $n = 49$ ) stated that *ad-free subscription payment should stop monetization of their data*, and 13% ( $n = 23$ ) believed that *privacy should prevail as a benefit of paying the subscription fee*.

Among the participants who responded negatively (No) to this question, 28% ( $n = 11$ ) thought that *data collection is needed for other platform functionalities beyond advertising*, 23% ( $n = 9$ ) mentioned their *indifference to data collection*, and 21% ( $n = 8$ ) stated that *the scope of the ad-free subscription only covers ad removal*.

**Descriptive expectation of reduced data collection:** Interestingly, although the majority of participants normatively expect a reduction in data collection, only 17% of our participants believed that online platforms *would* actually collect less data from users subscribed to the ad-free model compared to the ad-based model

(see Figure 2b). Furthermore, we see that this expectation of participants is comparable across the three user groups. Particularly, 22% of legal experts, 13% of technology experts, and 17% of general public responded positively to Q\_Descriptive.

Importantly, among the participants who normatively expect platforms to reduce data collection, 63% expressed doubt that platforms would actually reduce data gathering. This skepticism aligns with the observed data collection practices of platforms (see Section 3) and highlights users' limited trust in online services.

Next, we analyzed participants' answers to the free-response question following Q\_Descriptive by aggregating responses across the three user groups. Among the participants who responded positively (Yes) to Q\_Descriptive, 55% ( $n = 24$ ) mentioned that *data is no longer needed for serving ads* and 20% ( $n = 9$ ) stated that *ad-free subscription payment should stop monetization of their data*.

Moreover, among the participants who responded negatively (No) to Q\_Descriptive, 38% ( $n = 65$ ) said that *data collection is profitable for platforms*, 21% ( $n = 36$ ) believed that *data collection is a standard practice of platforms*, and 13% ( $n = 23$ ) thought that *the scope of the ad-free subscription only covers ad removal*.

**Choice of subscription model:** Figure 2c shows the distribution of participants' responses to Q\_Influence. Overall, 34% of participants mentioned that continued collection of the same amount of data would deter their decision to choose an ad-free plan. Moreover, we see that the prevalence of this belief is quite comparable across the three user groups. Specifically, 42% of legal experts, 31% of technology experts, and 29% of the general public responded positively to Q\_Influence.

To understand the reasoning of participants, we analyzed their answers to the free-response question following Q\_Influence by aggregating responses across the three user groups. Among the participants who responded positively (Yes) to Q\_Influence, 43% ( $n = 37$ ) mentioned that *paying for an ad-free subscription is not financially worth it because platforms still collect the same amount of data* and 30% ( $n = 26$ ) stated *privacy and data protection concerns*. Furthermore, among the participants who responded negatively (No) to Q\_Influence, 30% ( $n = 39$ ) stated that *data collection is an inevitable practice of online platforms*, 28% ( $n = 37$ ) primarily wanted to *only avoid viewing ads*, and 27% ( $n = 35$ ) raised *their indifference and lack of concern*.

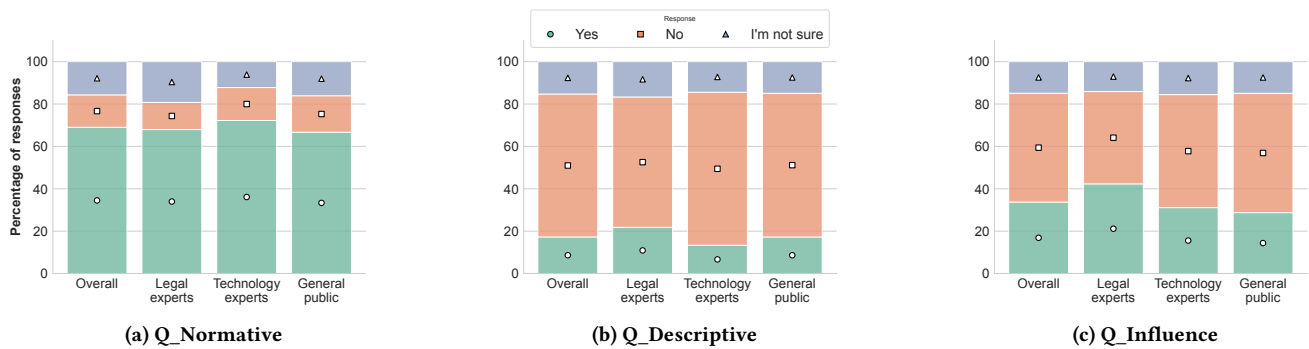
#### Main takeaways:

- 👉 The majority of surveyed participants (69%) normatively expect platforms to reduce data collection in ad-free model.
- 👉 However, only 17% of our participants descriptively expect platforms to reduce data collection in ad-free model.
- 👉 Furthermore, 34% of participants mentioned that the continued collection of the same amount of data in the ad-free model as in the ad-based model would deter their decision to subscribe to the ad-free plan.

## 5 Concluding Discussion

### 5.1 Summary

We studied how platforms differ in their data collection practices across different subscription models and how users perceive the



**Figure 2: Distribution of participants' responses to Q\_Normative, Q\_Descriptive, and Q\_Influence. We observe that, overall, 69% of participants believed that online platforms should collect less user data in the ad-free model compared to the ad-based model. However, only 17% of participants believed that online platforms would actually reduce their data collection practices in reality. Furthermore, 34% of participants stated that the continued collection of the same amount of data in the ad-free model as in the ad-based model would deter their decision to subscribe to the ad-free model.**

data collection practices associated with these subscription models. Through exercising GDPR *Right of access* (Article 15(3)), we empirically studied the data collection practices of Instagram, Facebook, and X across their ad-based and ad-free subscription models. Our analysis revealed some evidence that Instagram and Facebook reduced their data collection by removing certain ad-related data under the ad-free model. However, we observed that all three platforms continue to retain or collect some personal data that we classify as advertising-related, even under the ad-free model. Next, by surveying 255 participants, we investigated user perception of data collection practices. We found that 69% of our participants normatively believe that platforms should reduce data collection in ad-free subscription model. Whereas, only 17% of them expect platforms to reduce their data collection in practice.

## 5.2 Limitations

Despite its contributions, our study has some limitations. First, our study is limited to two author-owned accounts on three platforms, which cannot capture the full range of user behaviors (e.g., device diversity or feature usage) and constrains the robustness of our findings across platforms. However, this restriction is deliberate, as analyzing third-party users' DDPs would require processing sensitive PII and creating multiple fake accounts would violate platforms' terms of service [3]. Hence, our findings establish a lower bound on what data continues to be collected or retrained in ad-free models. Also, to evaluate robustness across platforms, we already analyzed YouTube. However, we excluded it as its DDP doesn't include any ad-related information other than viewed ads, which prevents us from drawing comparable conclusions. Second, although GDPR Article 15(3) mandates comprehensive data disclosure, the completeness of DDPs cannot be independently verified due to the proprietary nature of such practices. Hence, our analysis is limited to the data available within these DDPs and our findings should be interpreted within the visibility scope of DDPs.

## 5.3 Discussion

Despite the above limitations, our insights reveal a clear gap between user demands and platform behavior. While 69% of surveyed participants believe platforms should collect less data under the ad-free model, our empirical analysis shows that platforms continue to retain or collect certain data that are processed exclusively for advertising purposes. This mismatch means that users are not receiving what they believe they are paying for, as subscribing to an ad-free model does not translate into reduced data collection. Notably, 34% of participants reported that the continued collection of the same amount of data in the ad-free model as in the ad-based model would deter their decision to pay for the ad-free subscription. This observation highlights how this gap undermines user trust and risks misinterpretation of subscription offerings.

Furthermore, our findings point to two broader implications. First, data collection practices under the ad-free model raises concerns about compliance with core GDPR principles, such as purpose limitation and data minimization (Article 5(1)). This calls for closer regulatory scrutiny of whether platforms meaningfully adjust their data practices across subscription models, alongside enforcement efforts that help ensure that "ad-free" reflects a substantive shift rather than a rhetorical label. Second, our results indicate at potential internal misalignment within platforms themselves. Even when companies aim to provide more transparency or comply with regulation, technical and organizational silos may prevent the implementation of fully-compliant solutions. Moving forward, we believe that there is a need for more coordination across legal, engineering, and product teams to ensure that platform-wide data practices align with both user demands and regulatory requirements.

## Acknowledgments

We used generative AI tools to refine the writing, improve clarity, and assist with coding throughout this work.

## References

- [1] 2024. Facebook and Instagram to Offer Subscription for No Ads in Europe. <https://about.fb.com/news/2024/11/facebook-and-instagram-to-offer->

- subscription-for-no-ads-in-europe/ Accessed: 22.01.2026.
- [2] 2025. About mobile advertising IDs for custom audiences. <https://business.facebook.com/business/help/570474483033581> Accessed: 22.01.2026.
  - [3] 2025. Can I create multiple Facebook accounts? <https://www.facebook.com/help/975828035803295> Accessed: 22.01.2026.
  - [4] 2025. Commission finds Apple and Meta in breach of the Digital Markets Act. [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_25\\_1085](https://ec.europa.eu/commission/presscorner/detail/en/ip_25_1085) Accessed: 22.01.2026.
  - [5] 2025. The Digital Markets Act. [https://digital-markets-act.ec.europa.eu/index\\_en](https://digital-markets-act.ec.europa.eu/index_en) Accessed: 22.01.2026.
  - [6] 2025. Settings that you can manage across your accounts in an Accounts Centre. <https://www.meta.com/en-gb/help/accounts-center/1894490827562393/> Accessed: 22.01.2026.
  - [7] Sila Ada, Nadia Abou Nabout, and Elea McDonnell Feit. 2022. Context information can increase revenue in online display advertising auctions: Evidence from a policy change. *Journal of Marketing Research* 59, 5 (2022), 1040–1058.
  - [8] Ashwaq Alsoubai, Afsaneh Razi, Zainab Agha, Shiza Ali, Gianluca Stringhini, Munmun De Choudhury, and Pamela J Wisniewski. 2024. Profiling the offline and online risk experiences of youth to develop targeted interventions for online safety. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024).
  - [9] Sina Blassnig, Eliza Mitova, Nico Pfiffner, and Michael V Reiss. 2023. Googling referendum campaigns: analyzing online search patterns regarding swiss direct-democratic votes. *Media and Communication* 11, 1 (2023).
  - [10] Tomasz Bujlow, Valentín Carela-Español, Josep Sole-Pareta, and Pere Barlet-Ros. 2017. A survey on web tracking: Mechanisms, implications, and defenses. *Proc. IEEE* 105, 8 (2017), 1476–1510.
  - [11] Juan Miguel Carrascosa, Jakub Mikians, Ruben Cuevas, Vijay Erramilli, and Nikolaos Laoutaris. 2015. I always feel like somebody’s watching me: measuring online behavioural advertising. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*. 1–13.
  - [12] Farah Chanchary and Sonia Chiasson. 2015. User perceptions of sharing, advertising, and tracking. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. 53–67.
  - [13] Abhisek Dash, Abhijnan Chakraborty, Saptarshi Ghosh, Animesh Mukherjee, and Krishna P Gummadi. 2024. Investigating Nudges toward Related Sellers on E-commerce Marketplaces: A Case Study on Amazon. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–31.
  - [14] EDPB. 2024. EDPB: ‘Consent or Pay’ models should offer real choice. [https://www.edpb.europa.eu/news/news/2024/edpb-consent-or-pay-models-should-offer-real-choice\\_en](https://www.edpb.europa.eu/news/news/2024/edpb-consent-or-pay-models-should-offer-real-choice_en) Accessed: 22.01.2026.
  - [15] 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*. 1–13.
  - [16] European Commission. 2016. General Data Protection Regulation. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>. Accessed: 22.01.2026.
  - [17] European Commission. 2025. The Digital Services Act package. <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>. Accessed: 22.01.2026.
  - [18] Florian M Farke, David G Balash, Maximilian Golla, and Adam J Aviv. 2023. How does connecting online activities to advertising inferences impact privacy perceptions? *arXiv preprint arXiv:2312.13813* (2023).
  - [19] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971), 378.
  - [20] Mario Haim, Dominik Leiner, and Valerie Hase. 2023. Integrating Data Donations in Online Surveys. *M&K Medien & Kommunikationswissenschaft* 71, 1-2 (2023).
  - [21] Valerie Hase and Mario Haim. 2024. Can we get rid of bias? Mitigating systematic error in data donation studies through survey design strategies. *Computational Communication Research* 6, 2 (2024).
  - [22] Sai Keerthana Karnam, Abhisek Dash, Antariksh Das, Sepehr Mousavi, Stefan Bechtold, Krishna P. Gummadi, Animesh Mukherjee, Ingmar Weber, and Savvas Zannettou. 2026. Setting the Course, but Forgetting to Steer: Analyzing Compliance with GDPR’s Right of Access to Data by Instagram, TikTok, and YouTube. In *IEEE Symposium on Security and Privacy (SP)*.
  - [23] Anastasia Kozyreva, Philipp Lorenz-Spreen, Ralph Hertwig, Stephan Lewandowsky, and Stefan M Herzog. 2021. Public attitudes towards algorithmic personalization and use of personal data online: Evidence from Germany, Great Britain, and the United States. *Humanities and Social Sciences Communications* 8, 1 (2021), 1–11.
  - [24] Lin Kyi, Sushil Ammanaghatta Shivakumar, Cristiana Teixeira Santos, Franziska Roesner, Frederike Zufall, and Asia J Biega. 2023. Investigating deceptive design in GDPR’s legitimate interest. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
  - [25] Zhou Li, Kehuan Zhang, Yinglian Xie, Fang Yu, and XiaoFeng Wang. 2012. Knowing your enemy: understanding and detecting malicious web advertising. In *Proceedings of the 2012 ACM conference on Computer and communications security*. 674–686.
  - [26] Aleksandar Matic, Martin Pielot, and Nuria Oliver. 2017. “OMG! How did it know that?” Reactions to Highly-Personalized Ads. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. 41–46.
  - [27] Aleecia M McDonald and Lorrie Faith Cranor. 2010. Americans’ attitudes about internet behavioral advertising practices. In *Proceedings of the 9th annual ACM workshop on Privacy in the electronic society*. 63–72.
  - [28] Victor Morel, Cristiana Santos, Viktor Fredholm, and Adam Thunberg. 2023. Legitimate interest is the new consent-large-scale measurement and legal compliance of IAB Europe TCF paywalls. In *Proceedings of the 22nd Workshop on Privacy in the Electronic Society*. 153–158.
  - [29] Thomas V Perneger, Delphine S Courvoisier, Patricia M Hudelson, and Angèle Gayet-Ageron. 2015. Sample size for pre-tests of questionnaires. *Quality of life Research* 24 (2015), 147–151.
  - [30] William D Perreault. 1975. Controlling order-effect bias. *The Public Opinion Quarterly* 39, 4 (1975), 544–551.
  - [31] Ali Rasaii, Devashish Gosain, and Oliver Gasser. 2023. Thou shalt not reject: analyzing accept-or-pay cookie banners on the web. In *Proceedings of the 2023 ACM on Internet Measurement Conference*. 154–161.
  - [32] Reuters. 2025. Amazon loses court fight against record 812 mln Luxembourg privacy fine. <https://www.reuters.com/technology/amazon-loses-court-fight-against-record-812-mln-luxembourg-privacy-fine-2025-03-19/> Accessed: 22.01.2026.
  - [33] Tanusree Sharma, Smirity Kaushik, Yaman Yu, Syed Ishtiaque Ahmed, and Yang Wang. 2023. User perceptions and experiences of targeted ads on social media platforms: Learning from bangladesh and india. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.
  - [34] Karthika Subramani, Xingzi Yuan, Omid Setayeshfar, Phani Vadrevu, Kyu Hyung Lee, and Roberto Perdisci. 2020. When push comes to ads: Measuring the rise of (malicious) push advertising. In *Proceedings of the ACM Internet Measurement Conference*. 724–737.
  - [35] The New York Times. 2019. Google Is Fined 57 Million Under Europe’s Data Privacy Law. <https://www.nytimes.com/2019/01/21/technology/google-europe-gdpr-fine.html> Accessed: 22.01.2026.
  - [36] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *Proceedings of the eighth symposium on usable privacy and security*. 1–15.
  - [37] Karan Vombatkere, Sepehr Mousavi, Savvas Zannettou, Franziska Roesner, and Krishna P Gummadi. 2024. TikTok and the Art of Personalization: Investigating Exploration and Exploitation on Social Media Feeds. In *Proceedings of the ACM on Web Conference 2024*. 3789–3797.
  - [38] Miranda Wei, Madison Stamos, Sophie Veys, Nathan Reitingger, Justin Goodman, Margot Herman, Dorota Filipczuk, Ben Weinschel, Michelle L Mazurek, and Blase Ur. 2020. What Twitter knows: Characterizing ad targeting practices, user perceptions, and ad explanations through users’ own Twitter data. In *29th USENIX Security Symposium (USENIX Security 20)*.
  - [39] Ben Weinschel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L Mazurek, and Blase Ur. 2019. Oh, the places you’ve been! User reactions to longitudinal transparency about third-party web tracking and inferring. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. 149–166.
  - [40] Cai Yang, Sepehr Mousavi, Abhisek Dash, Krishna P. Gummadi, and Ingmar Weber. 2025. Studying Behavioral Addiction by Combining Surveys and Digital Traces: A Case Study of TikTok. In *Proceedings of the International AAAI Conference on Web and Social Media*.
  - [41] Savvas Zannettou, Olivia Nemes-Nemeth, Oshrat Ayalon, Angelica Goetzen, Krishna P Gummadi, Elissa M Redmiles, and Franziska Roesner. 2024. Analyzing User Engagement with TikTok’s Short Format Video Recommendations using Data Donations. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–16.
  - [42] Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. 2021. What makes a “bad” ad? user perceptions of problematic online advertising. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–24.

## A Demographic statistics of survey participants

Table 3 summarizes two main demographic characteristics of our participants, namely gender and age. We observe that the gender of our recruited participants is skewed towards Male users. Furthermore, we see that the age distribution of our participants is skewed towards users who are young and at most 34 years old.

## B Information contained in the DDPs of Instagram, Facebook, and X

Table 4 contains detailed descriptions of the fields of information made available in the DDPs of Instagram, Facebook, and X.

**Table 3: Demographic distribution of survey participants.**

Characteristic	Category	Count	Percentage
Gender	Male	167	65.5
	Female	88	34.5
Age	18-24	41	16.1
	25-34	128	50.2
	34-44	47	18.4
	45-64	38	14.9
	65+	1	0.4

### C Survey questions

In this section, we provide our questionnaire.

*This section focuses on your views regarding how data collection practices **should** differ between a platform's ad-based and ad-free subscription plans.*

- (1) If you subscribe to an ad-free plan of a platform, **should** the platform collect less data about you?
  - Yes
  - No
  - I'm not sure

- (2) Please justify your above choice.

*This section asks **what you think the potential differences will be** in user data collection between ad-based and ad-free subscription plans on a platform.*

- (1) If you subscribe to an ad-free plan of a platform, **do you think** that the platform **would** collect less data about you?
  - Yes
  - No
  - I'm not sure

- (2) Please justify your above choice.

*This section focuses on whether continued collection of the same amount of data between ad-based and ad-free subscription plans affects your willingness to choose an ad-free subscription plan.*

- (1) Would it affect your decision to choose an ad-free plan if the platform continued collecting the same amount of data as it does in the ad-based plan?
  - Yes
  - No
  - I'm not sure

- (2) Please justify your above choice.

### D Qualitative coding of free-response questions

For each free-response question, we grouped responses according to the corresponding multiple-choice answer (Yes, No, or I'm not sure), resulting in nine sets in total. Then, one co-author independently reviewed each set and identified all the broad themes that captured ideas expressed in the responses. We present the three most common themes for each set, along with a random participant response for each theme in Table 5.

**Table 4: Fields of information available in the data download packages of Instagram, Facebook, and X, along with their descriptions.**

<b>Online Platform</b>	<b>Field of Information</b>	<b>Description</b>
Instagram/Facebook	<b>Your Instagram/Facebook activity</b>	Information about user activity, such as content created, messages and more
	<b>Personal information</b>	Information about user profile
	<b>Connections</b>	Who and how user has connected with people
	<b>Logged information</b>	Information that Meta logs about user activity, including things like user location and search history
	<b>Security and login information</b>	Technical information and login activity
	<b>Apps and websites off of Instagram/Facebook</b>	Apps user owns and activity Instagram/Facebook receives from apps and websites off of Instagram/Facebook
X (formerly Twitter)	<b>Preferences</b>	Actions user has taken to customize their experience
	<b>Ads information</b>	User interactions with ads and advertisers
	<b>Sensory information</b>	Audio, electronic, visual, and similar information
	<b>Identifiers</b>	Real name, alias, postal address, telephone number, unique identifiers (such as a device identifier, cookies, mobile ad identifiers), customer number, Internet Protocol address, email address, account name, and other similar identifiers
	<b>Online activity</b>	Internet and other electronic network activity information, including, but not limited to, information regarding interactions with websites, applications, or advertisements
	<b>Inferences</b>	Inferences drawn to create a profile about the user reflecting their preferences, characteristics, predispositions, behavior, and attitudes
	<b>Protected classifications</b>	Characteristics of certain legally protected classifications

**Table 5: The three most common themes observed for each set after conducting qualitative coding of free-response answers.**

Question	Multiple-choice answer	Coded free response	Example free-response answer
Q_Normative	Yes	Reduced need (ad-free)	<i>They don't need data for personal advertising, so they could stop collecting them.</i>
		Payment principle	<i>If something is free, you are the product.</i>
		Privacy expectation	<i>If I am paying for a service it should come with benefits in terms of privacy as well</i>
	No	Data needed (other functions)	<i>They need data to provide other features</i>
		User indifference	<i>Dont care</i>
		Scope of ad-free plan	<i>I think you're just paying to not have ads.</i>
	I'm not sure	Lack of knowledge	<i>i don't understand these data things</i>
		Depends on specifics	<i>Maybe it just depends on the agreement.</i>
		Skepticism	<i>I cant control the data collected by the platform, but probably they collect every single click on ad-free or ad-based</i>
Q_Descriptive	Yes	Reduced need (ad-free)	<i>Because without ads, less data is needed for targeting and personalization.</i>
		Payment changes incentive	<i>Since I'm paying and the platform is making a profit, why does it still collect so much data?</i>
		Expected benefit of plan	<i>a premium service has certain benefits and I think this was the case</i>
	No	Data value	<i>The payment for the service are data</i>
		Standard practice	<i>I think they still collect the same data as before.</i>
		Scope of plan	<i>I don't think it matters to them, ad free means just that, not "more privacy".</i>
	I'm not sure	Lack of knowledge	<i>I honestly am not very informed about the data collection on ad-free platforms</i>
		Skepticism	<i>I'm not sure as I don't really trust how the platforms collect our data</i>
		Conflicting beliefs	<i>I am thinking it twice maybe they collect my personal data but I don't see relevant ads</i>
Q_Influence	Yes	Value proposition	<i>If I know it will collect more or the same amount, I just might refuse the service as it would not be worth it</i>
		Privacy expectation	<i>Because one of the main reasons to pay for ad-free is to reduce data collection and protect privacy.</i>
		Reduced need (ad-free)	<i>if there's no ad, they dont need my data</i>
	No	Inevitability	<i>It is so universal that it is impossible to fight anymore</i>
		Focus on ad removal	<i>Because when buying such a package I am primarily interested in the lack of advertisements</i>
		User indifference	<i>I just wouldnt care about it, as long as I get no ads</i>
	I'm not sure	Depends on specifics	<i>Depends. I'm using adblocker everywhere, so ad-free plans don't have a lot of value for me</i>
		Lack of knowledge	<i>not sure have no knowledge in this really</i>
		Focus on ad removal	<i>I would probably use the plain mainly to not see ads, the data collection is secondary</i>