User-Centric Textual Descriptions of Privacy-Enhancing Technologies for Ad Tracking and Analytics

Lu Xian
University of Michigan
xianl@umich.edu

Song Mi Lee-Kan
University of Michigan
songmil@umich.edu

Jane Im
University of Michigan
imjane@umich.edu

Florian Schaub
University of Michigan
fschaub@umich.edu

Abstract

Describing Privacy Enhancing Technologies (PETs) to the general public is challenging but essential to convey the privacy protections they provide. Existing research has explored the explanation of differential privacy in health contexts. Our study adapts well-performing textual descriptions of local differential privacy from prior work to a new context and broadens the investigation to the descriptions of additional PETs. Specifically, we develop user-centric textual descriptions for popular PETs in ad tracking and analytics, including local differential privacy, federated learning with and without local differential privacy, and Google's Topics. We examine the applicability of previous findings to these expanded contexts, and evaluate the PET descriptions with quantitative and qualitative survey data (n=306). We find that adapting a process- and implications-focused approach to the ad tracking and analytics context achieved similar effects in facilitating user understanding compared to health contexts, and that our descriptions developed with this process+implications approach for the additional, understudied PETs help users understand PETs' processes. We also find that incorporating an implications statement into PET descriptions did not hurt user comprehension but also did not achieve a significant positive effect, which contrasts prior findings in health contexts. We note that the use of technical terms as well as the machine learning aspect of PETs, even without delving into specifics, led to confusion for some respondents. Based on our findings, we offer recommendations and insights for crafting effective user-centric descriptions of privacy-enhancing technologies.

Keywords

Privacy-enhancing technologies, local differential privacy, federated learning, Google Topics, ad tracking, analytics.

1 Introduction

Various organizations and companies are increasingly incorporating Privacy Enhancing Technologies (PETs) into their services and products. In response to increased privacy concerns and regulations, platforms involved in ad tracking and analytics are moving away from third-party cookies, which enabled individual-level tracking and targeting [30], towards more privacy-focused approaches [12, 23]. Apple and Google use local differential privacy *(LDP)* for analytics in Safari [21, 68], Chrome [26], and across other services [3, 13, 33], adding statistical noise to user data on individual devices before it is transmitted to central servers. Google also employs federated learning *(FL)* [40, 56], which trains machine learning models on decentralized devices to update global models without centralizing data [40, 49, 50, 56]. FL has expanded to other applications, including IBM's data analytics services [42]. FL is often paired with LDP *(FL+LDP)* to enhance privacy in model training, such as in Google's Smart Text Selection and Apple's Siri personalization [32, 38]. This method incorporates noise into shared models to protect user data [70]. Google has launched Topics *(GT)* as part of its Privacy Sandbox initiative [31, 59], replacing the preceding federated learning of cohorts *(FLoC)* [57]. Unlike FLoC, which was criticized for reducing browsing activities into behavioral labels and potentially enabling predatory ad targeting [5, 7, 12, 19], GT analyzes browsing data directly on the device to identify interests. This allows advertisers to target ads based on a subset of these interests without tracking individuals across websites [59]. However, GT also faces criticism for not providing users sufficient control over how their browsing history-derived information is shared [15].

Given the potential impact of PETs on user privacy regarding online ads and analytics, providing intuitive descriptions of the employed PETs that effectively explain to users how the PET increases privacy protections and its limitations is crucial for fostering users' informed decision-making. Prior research primarily conducted in health settings shows that users' understanding of PET descriptions significantly affects their willingness to share data [8, 18, 34, 53, 55]. Studies have focused predominantly on effective communication methods for differential privacy (DP) and LDP within the health domain [28, 45, 51, 72, 73]. While PETs can be presented through various formats like images and interactive tools, textual descriptions are most commonly used by PET vendors, implementers (e.g., see Table 5 in Appendix A.3), and in privacy policy documents. A promising approach for textually describing PETs, developed by Xiong et al. [72], combines an explanation of how a PET works on a *process* level, such as random modification of data, with a statement on its privacy *implications*, such as reducing the risk from server-side data breaches. This method focuses on conveying the PET's core process in a textual description, rather than delving into technical details, for instance omitting the ϵ parameter in a description of LDP, to avoid confusion [72]. Both Xiong et al. [72] and

Proceedings on Privacy Enhancing Technologies 2025(1), 379–398

https://doi.org/10.56553/popets-2025-0021

Kühtreiber et al.'s replication study [47] found that including a privacy implications statement in DP and LDP descriptions for health apps significantly enhances users' objective comprehension and willingness to share data. Similarly, Smart et al. [63] found a positive effect, although insignificant, of an explanation that includes privacy parameter's implications for DP on user comprehension.

However, due to the context-dependent nature of privacy and information sensitivity [1], it is unclear if these findings in the health domain apply to other domains, such as ad tracking and analytics. For instance, social media users often face heightened privacy risks due to the platform's default privacy settings, prevailing norms around self-disclosure, and frequent exposure to third-party ads [24, 46]. They may perceive behavioral data used for advertising in a different light compared to health data. Furthermore, while previous studies have focused on user-centric descriptions of DP/LDP, how to effectively describe other PETs, such as federated learning combined with local differential privacy or Google Topics, has not yet been studied despite their growing use.

Our research investigates the effectiveness of a PET description approach that emphasizes processes and includes a privacy implications statement, which has been shown to be effective for fostering understanding of DP/LDP used in health apps, in a different context (ad tracking and analytics) and for PETs beyond DP/LDP. The goal here is to ascertain the extent to which this process+implications approach generalizes to PETs used in other contexts. We adapted and refined Xiong et al.'s [72] LDP description to the ad tracking and analytics context and also developed similar descriptions for a new set of PETs: FL, FL+LDP, and GT. We conducted an online survey experiment with 306 U.S. adults to assess their understanding of these PET descriptions, both with and without a privacy implications statement.

Summary of contributions. We find that:

- The LDP descriptions we developed for ad tracking and analytics yielded user comprehension levels similar to those seen with existing health context descriptions. This demonstrates that the process+implications approach to PET descriptions can be successfully adopted in other domains.
- The process- and implications-focused approach is also adaptable to other PETs (FL, FL+LDP, and GT) in the ad tracking and analytics context.
- In contrast to prior findings in the health context, adding an implications statement to LDP and other PET descriptions (FL, FL+LDP, and GT) only marginally improved understanding in our study, but also did not negatively affect user comprehension or confidence.
- Our qualitative analysis provides insights about aspects in textual descriptions that fostered understanding or led to confusion. For instance, participants understood user data shared with the platform is modified or not shared due to the use of LDP and FL, respectively. Yet, terms such as "adding noise" or "machine learning" still elicited confusion, despite our efforts to reduce and avoid jargon.

Our findings validate the utility of emphasizing process in describing PETs, but also highlight the complexities of describing PETs. Based on our findings, we provide recommendations for crafting effective user-centric PET descriptions.

2 Related Work

We discuss existing literature on strategies to communicate Privacy-Enhancing Technologies (PETs) in a way that is accessible and relatable to non-experts. This includes exploring the content of what needs to be explained (e.g., PET processes, privacy implications) and determining the effective formats for such explanations (e.g., texts, visualizations, interactive interfaces). Our work addresses challenges in explaining PETs by focusing on process- and implication-focused textual descriptions.

2.1 Challenges in Explaining PETs

PETs encompass a range of technical measures explicitly designed to safeguard diverse privacy aspects of user identities and behaviors [27, 41]. These measures range from data minimization, encryption, and network traffic anonymization methods, to transparency-enhancing tools, and identity management solutions [43]. As privacy is a many-faceted concept, each PET addresses distinct aspects of information privacy and thus may find application in different settings [41, 72] (see [43] for a taxonomy of PETs).

However, misconceptions prevail among laypersons, who often perceive PETs as supplementary features of existing technologies or abstract, on-demand services that require technical training [16, 25, 29]. Such misconceptions are shaped by many cognitive, affective, and socio-contextual factors [66]. Consistent with the generally low technological literacy among the public [17], there is a widespread misunderstanding of common technical terms in the privacy domain, such as "local storage" or "Do Not Track" [67]. Many users regard privacy protection as a secondary concern when using information and communication technologies [39]. A subset of users perceives the use of PETs as "extreme" due to the belief that they have "nothing to hide" or consider it socially undesirable as it might mistakenly imply engagement in illicit activities [16, 39]. Some users further question the ethics of using PETs, equating techniques like data obfuscation with "lying" [8]. These misunderstandings demonstrate the complexity of privacy concerns, requiring diverse solutions that address behavioral variables like limited attention, self-efficacy, social norms, and the salience of privacy cues [2, 20].

One crucial challenge to tackle is effectively communicating the nature and function of PETs to end users to help them develop an accurate understanding of the privacy guarantees or risks they are subject to. Promoting accurate comprehension of the technical aspects of PETs is one of the prerequisites for users to gain privacy literacy [69]. When it comes to communicating technical aspects, striking a balance between simplifying technical complexity to make the information *relevant, actionable, and understandable* [60] and maintaining fidelity is a tricky yet indispensable task [62, 71, 72]. This balance is pivotal in facilitating well-informed choice of and trust in technologies, allowing individuals to fully leverage their benefits while understanding the limitations and pitfalls [16, 17, 61, 71]. Research in explainable machine learning highlights two key decisions for crafting descriptions that guarantee both accuracy and comprehensibility: determining *what to explain* (i.e., content) and *how to explain* it (i.e., delivery format) [22, 64]. Our study addresses the challenge of effectively communicating PETs by exploring *what to explain* to users.

2.2 What to Explain: Process and Implications

Existing literature cautions against the oversimplification of system logic. Such simplification risks depriving individuals, particularly non-experts, of the chance to develop accurate and sophisticated mental models [22]. End users are notably interested in understanding the data protection processes in PETs [16, 47]. To balance the need for more information and potential information overload, some studies recommend having users pinpoint specific confusing vocabularies in the given text and then providing clarifications for those terms [54, 67].

Meanwhile, several studies underscore the importance of incorporating privacy implications into PET descriptions to enhance comprehension, especially for those without a technical background [47, 62, 72]. Descriptions that focus solely on definitions and data processing techniques can be challenging for this audience. Xiong et al. [72] conducted online experiments comparing participants' data sharing intentions, subjective understanding (i.e., how easy they perceive it to understand), and objective understanding (i.e., correct responses to comprehension check questions) when presented with different descriptions of differential privacy (DP) and local differential privacy (LDP). They assessed four types of descriptions: 1) baseline definitions as the control; 2) descriptions emphasizing LDP's data perturbation process; 3) descriptions naming widely-known companies like Google and Apple that use these PETs; and 4) descriptions that include privacy implications. Their findings suggest that descriptions with privacy implications improve correct understanding and encourage data sharing under LDP, which is considered safer than DP [72]. Kühtreiber et al. [47] replicated these findings with a German sample, a demographic typically regarded as more privacy-conscious, confirming that incorporating implications statements enhances willingness to share data. Similarly, Smart et al. [63] found that outcome-focused explanations, which offer higher transparency about privacy implications, marginally outperformed explanations that only focused on the process of how DP/LDP works in enhancing user comprehension.

These foundational studies primarily focused on describing DP and LDP within the context of data sharing decisions for medical or health-related apps [e.g., 18, 62, 72]. Explainability research highlights that the context of use heavily influences how users assess explanations, particularly concerning their perceived truthfulness and completeness, such as the level of detail and generalizability [22, 64]. Furthermore, users weigh the sensitivity of the information (e.g., medical records vs. exercise self-reports) and the trustworthiness of companies handling the data when interpreting descriptions [72]. Given the context-specific nature of privacy, our study extends the application of process- and implications-focused PET descriptions to a novel context—personalized advertising and analytics—and to other prevalent PETs (FL, FL+LDP, GT) for which textual descriptions have not been studied yet.

2.3 How to Explain: Text Vs. Multimodal

Explanations can be presented through various modalities and styles, such as numerical summaries, textual technical descriptions, argumentation, metaphors, icons with labels, diagrams, interactive interfaces, or a combination thereof [37, 62, 64]. For DP in particular, visual illustrations, interactive visualization tools, along with the effects of different ϵ parameters, have been a popular area of study (e.g., [35, 44, 45, 52, 53, 73]). In the privacy field, Habib et al. [37] tested different combinations of icons and textual descriptions that communicate privacy choices to users, and found that icons alone might not sufficiently convey complex privacy concepts or data flows. Their results affirmed that icons evoking simpler concepts, such as choices, can be beneficial but not sufficient for user comprehension.

The chosen method of presentation significantly influences user attention and understanding due to the varying cognitive efforts required. Explainability research suggests that while visual elements can be mentally processed more quickly, plain textual descriptions might provide equally effective communication [22], especially for abstract or unfamiliar concepts. While various formats, including images and interactive tools, have been proposed to present and explain PETs, there is no consensus on the most effective medium or method for such presentations. Currently, textual descriptions remain the most common approach used by PET vendors, implementers (e.g., see Table 5 in Appendix A.3), and in privacy policy documents. Our study focuses on investigating and developing textual descriptions to enhance user comprehension of PETs.

3 Methods

Our study investigates the effectiveness of a textual PET description approach that combines a process-focused explanation of how a PET functions with a statement of its privacy implications [47, 72]. We adopted the process- and implications-focused approach for LDP from prior work's health settings to a new context—ad tracking and analytics. We further applied it to craft new descriptions for other PETs relevant to ad tracking and analytics, namely for federated learning (FL), federated learning with LDP (FL+LDP), and Google's Topics (GT). We then evaluated user comprehension of these descriptions and compared the effects of process-only versus process+implications descriptions. We asked:

RQ1: How does user comprehension differ between our refined LDP description and the minimally modified description from Xiong et al. for the ad tracking and analytics context?

RQ2: How does including an implications statement, found effective in Xiong et al.'s health context [72], affect user comprehension of the LDP description for ad tracking and analytics?

RQ3: How does including an implications statement affect user comprehension of the descriptions of FL, FL+LDP, and GT for ad tracking and analytics?

RQ4: What aspects of the text describing PETs for ad tracking and analytics do users comprehend accurately and inaccurately?

In this section, we begin by outlining the development of the PET descriptions employed in our study. Following this, we detail the recruitment and procedure of our survey experiment, along with our quantitative and qualitative measures and analyses. Following an overview of participant demographics and our approaches to ensuring data quality, we then discuss the limitations of our methodology.

3.1 Development of PET Descriptions

We developed process-focused textual descriptions for different PETs (LDP, FL both with and without LDP, and GT) in the context of ad tracking and analytics, aiming to make them understandable to users without a technical background. The descriptions for LDP were adapted from Xiong et al. [72]. For FL, FL+LDP, and GT, which lack established descriptions, we drafted descriptions based on industry documents (see Table 5 for examples) and academic research. In order to test the effect of including implementation statements, which were shown to be effective in prior work in the health context [47, 72], we created two versions of each PET description: one describing the PET's data processing approach, and another combining this with a privacy implications statement ("impl"). We refined our descriptions through think-aloud interviews (*n*=5) and multiple rounds of pilot testing with Prolific participants (*n*=21 in total), all conducted with informed consent. We also incorporated feedback from an industry privacy engineer. Table 1 shows the 10 finalized descriptions evaluated in our survey, including a control description we developed that neither explains how privacy is protected nor contains an implications statement. Below, we describe the specific adaptations and modifications we made to Xiong et al.'s LDP description, along with our development process for the descriptions of FL, FL+LDP, and GT. Figure 3 in Appendix A.1 presents a flow figure illustrating the steps we took to develop these descriptions.

3.1.1 *Local differential privacy descriptions.*

LDP modifies user data directly on the user's device by adding statistical noise before transmitting it to a centralized system [21]. This system can then infer statistics and behavioral patterns from the aggregated data. This approach protects user privacy by making user data differentially private before its distribution [21]. We formulated and assessed three descriptions for LDP: 1) a modified version of Xiong et al.'s most effective process+impl description [72] tailored for ad tracking and analytics (*LDP-Xiong*), 2) our process-only description (*LDP*), and 3) our description with an added sentence about privacy implications (*LDP-impl*).

LDP-Xiong. We replaced terms from Xiong et al.'s original text, such as "the app" and "the cellphone," with terms more relevant to the ad tracking and analytics context, like "the company." The references to "data" were specified as *"behavioral data (e.g., your interaction with the platform and with other apps/websites)."* (We consistently refer to data in this way in all descriptions.) We were interested in user understanding of PETs based on the provided descriptions rather than their pre-existing conceptions. Therefore, in our study, PETs are described without their names. The reference to "local differential privacy" in Xiong et al.'s [72] original LDP description was modified to "additional privacy technique." We highlight the modifications in italicized text of LDP-Xiong in Table 4 in Appendix A.2.

LDP. Our LDP description expanded on LDP-Xiong in several aspects. First, while LDP-Xiong focuses on the data sharing aspect of LDP, our version explicitly outlined its data modification process: 1) how user data is randomly modified with the addition of noise before sharing with the platform, and 2) how the sharing includes a randomly selected subset of this altered data (i.e., *"...some of your actual data is used whereas some of it is random and not representative of your behavior"*). The extent to which the organization has access to and can infer behavioral patterns is also specified: *"Your exact behavioral data is never sent to the organization,"* and *"The organization can still infer patterns from the noisy data across a large number of users."* This refinement of our LDP description from Xiong et al.'s [72] was informed by think-aloud interviews (*n*=5).

LDP-impl. The implications statement for our LDP description was adapted from Xiong et al.'s [72] to describe potential consequences of a security breach: *"the organization still learns [...] but not your exact behavior, which protects your privacy against the organization's employees or if the organization's database is compromised."* This implications statement is also used in the "impl" descriptions of other PETs with minor modifications for consistency. To maintain accessibility and relevance, we deliberately avoided intricate technical details (e.g., the significance of the privacy loss parameter ϵ [51]), and even the name "local differential privacy."

3.1.2 *Federated learning descriptions.*

FL trains local machine learning models on users' devices using their data, then aggregates the parameters of these local models to create a global model for a group of users [49]. This method enhances privacy by keeping user data on the device; however, incorporating LDP into the process can significantly improve privacy. Thus, we developed federated learning descriptions with (FL+LDP) and without LDP (FL) to assess whether that nuanced difference becomes clear in our descriptions.

FL. Taking the same approach used for developing LDP descriptions, we drafted FL descriptions based on industry sources, notably Google's initial announcement of FL [49] and IBM's blog post [6]. These drafts were then refined through think-aloud interviews. For example, a participant was confused by the use of "pattern" in the initial phrase *"machine learning on the user's device to identify patterns."* To improve clarity, we substituted it with "inferred interests." Our descriptions highlighted the use and sharing of machine learning models instead of raw behavioral data to deduce user interests. We also clarified the process of merging machine learning models to infer collective behavioral patterns.

FL+LDP. FL+LDP operates by training local machine learning models on user devices with their data (the FL part) and by incorporating statistical noise into these models' parameters (the LDP part) [4]. This process occurs before the parameters are aggregated to create a global model, ensuring that individual contributions are effectively masked. In developing our descriptions, we referenced Meta's research report [4] and Wei et al.'s work [70] to simplify technical details. Our focus was on emphasizing the role of machine learning models in inferring user interests, data modification through noise addition, and the fact that users' actual behavioral data is not shared with the platform.

FL-impl & FL+LDP-impl. The implications statements for both followed the LDP version with minor modifications to account for differing privacy features, e.g., the organization learning "your interests" (FL) versus "aggregated interests across its users" (FL+LDP).

3.1.3 *Google Topics descriptions.*

GT is a PET developed by Google for improving privacy in ad tracking/targeting [31, 59]. GT analyzes users' recent activities within a given timeframe to identify their interests, termed as "topics." It then randomly selects one of the top five identified topics for sharing with the platform [31, 58]. There is a chance that the shared topic is not directly derived from user activities but is randomly selected from a predefined list of advertising topics. This approach enhances privacy by sharing inferred topics

Table 1: PET descriptions tested in the survey experiment. Shown are the combined process+implications descriptions. Here, the implications statements are in bold text in square brackets for clarity but were not bolded in the actual survey shown to participants. For process-only conditions, the implications statements were left out.

Condition	Description
Control	To protect your information, the organization stores all of your behavioral data for targeting ads (e.g., your interaction with the platform and with other apps/websites) securely on their servers.
LDP-Xiong	To respect your personal information privacy and ensure the best user experience, the behavioral data (e.g., your interaction with the platform and with other apps/websites) shared with the company will be processed via an additional privacy technique. That is, your behavioral data will be randomly modified before it is sent to the company. Since the company stores only the modified version of your personal information, your privacy is protected even if the company's database is compromised.
LDP[**-impl**]	To protect your information, the organization adds noise to your behavioral data (e.g., your interaction with the platform and with other apps/websites) before being sent to the organization for targeting ads. This means that your data is randomly modified, so that some of your actual data is used whereas some of it is random and not representative of your behavior. Your exact behavioral data is never sent to the organization, instead a subset of your noisy data is randomly selected and sent. The organization can still infer patterns from the noisy data across a large number of users. [**This way, the organization still learns aggregated interests across users but not your exact behavior, which protects your privacy against the organization's employees or if the organization's database is compromised.**]
FL[**-impl**]	To protect your information, the organization uses machine learning on your device to infer interests from your behavioral data (e.g., your interaction with the platform and with other apps/websites) for targeting ads. Your exact behavioral data is never sent to the organization and only a machine learning model representing your inferred interests will be sent. Then, to infer patterns across a large number of users, your model is merged with other users' models. [**This way, the organization still learns your interests but not your exact behavior, which protects your privacy against the organization's employees or if the organization's database is compromised.**]
FL+LDP[**-impl**]	To protect your information, the organization uses machine learning on your device to infer interests from your behavioral data (e.g., your interaction with the platform and with other apps/websites) for targeting ads. Noise will be added to your behavioral data so that it is randomly modified before being used for training a machine learning model representing your inferred interests. This means that, for training the model, some of your actual data is used whereas some of it is random and not representative of your behavior. Your exact behavioral data is never sent to the organization and only the model representing your inferred interests will be sent. Then, to infer patterns across a large number of users, your model is merged with other users' models. [**This way, the organization still learns aggregated interests across its users but not your exact behavior, which protects your privacy against the organization's employees or if the organization's database is compromised.**]
GT[**-impl**]	To protect your information, the organization uses machine learning on your device to infer interests from your behavioral data (e.g., your interaction with the platform and with other apps/websites) for targeting ads. This means that the technology records inferred topics you may be interested in from your behavioral data only on your device. Your exact behavioral data is never sent to the organization, instead from your top topics of the last week, a small number are randomly selected and sent; there is also a small chance a random topic will be selected instead of one of yours. [**This way, the organization still learns some of your interests but not your exact behavior, which protects your privacy against the organization's employees or if the organization's database is compromised.**]

rather than actual user data, and by incorporating a random topic selection process that avoids solely sharing specific user interests.

GT. Given that GT has been developed and implemented by Google, our primary source of descriptions was Google's documents [31, 58]. We aimed to match the level of technical detail included in our other PET descriptions. Our descriptions highlighted the use of a machine learning model on the device to infer user interests, and the sharing of random and possibly modified subsets of these inferred interests with websites the user visits.

GT-impl. The implications statement sits between those of LDP and FL in that we highlighted that "some" of your interests are learned by the organization.

3.2 Online Survey Experiment

To evaluate the comprehensibility of our PET descriptions and test the effectiveness of the process-plus-implications approach in the ad tracking and analytics context, we conducted an online survey experiment using Qualtrics. We gathered both quantitative and qualitative data to assess participants' reactions to these descriptions. This study received an exemption from our university's Institutional Review Board. We pilot-tested the survey instrument along with our PET descriptions with Prolific participants (n=21 in total), all conducted with informed consent, as mentioned in Section 3.1.

3.2.1 Participant recruitment. A total of 357 participants were recruited through Prolific for the main study, which was advertised as an investigation into user perceptions of a social media platform to minimize self-selection bias. All participants were English-speaking U.S. adults, aged 18 years or older. No additional eligibility criteria were applied. Participants were compensated USD $4.25 for their time. The median survey completion time was 10.75 minutes. Due to exceptionally short completion time, low-effort answers to open-ended questions, and instances of time-out, 51 responses were excluded from the analysis. Consequently, we analyzed 306 responses.

3.2.2 Procedure. After providing consent, participants were randomly assigned to one of ten experimental conditions, each featuring a PET description as the stimulus. These conditions included a control group, LDP-Xiong, and descriptions both with and without "impl" (implications statement) for LDP, FL both with and without LDP, and GT (see Table 1). Our survey procedure was inspired by and adapted from that used by Smart et al. [62]. In each condition, participants were presented with a scenario involving a fictitious social media platform that generates revenue through advertising and uses a specific PET for user data protection, i.e., the PET description. Participants first answered a Likert-scale question assessing their confidence in deciding to use the described platform. They then explained in their own words how the platform safeguards their data. This was followed by five true/false questions testing their comprehension of the respective PET. Participants were also asked to articulate their understanding of specific concepts in the PET description, rate their subjective comprehension on a Likert scale, and identify any confusing parts of the description. At the end of the survey, participants answered a series of demographic questions, along with an additional question regarding whether they have an educational background or employment in computer science/engineering or information technology. The complete survey instrument is provided in Appendix B.

3.2.3 Quantitative measures and analyses. We conducted statistical tests to compare pairs of experimental conditions for each PET, quantitatively measuring the effect of the implications statement in PET descriptions. We used non-parametric Mann–Whitney tests for our non-normally distributed data.

Confidence in platform use. Participants rated their confidence in deciding to use the fictional platform on a 5-point scale (1=Not at all confident, 5=Very confident). This metric helps gauge the potential influence of PET understanding, among other factors, on the decision to adopt the platform.

Objective comprehension score. Participants responded to five comprehension questions adapted from Smart et al. [62] for the social media context, with options including "true," "false," and "don't know:" (Q1) *"An employee working for the platform, such as a data analyst, could be able to see my exact behavioral data."* (Q2) *"A criminal or foreign government that hacks the platform could learn my behavioral data."* (Q3) *"A law enforcement organization could access my behavioral data with a court order requesting this data from the company."* (Q4) *"Graphs or informational charts created using information given to the platform could reveal my behavioral data."* (Q5) *"Data that the platform shares with its partner organizations could reveal my behavioral data."* The correct answer to all questions is "False" across conditions, except for the control condition. Scores from questions were aggregated to form a total score ranging from 0 to 5 that objectively quantifies participants' understanding of privacy technologies in general. Further, we analyzed individual questions' correctness rates as a crude measure of PET descriptions' comprehensibility. Differences in correctness rates between conditions with and without "impl" for each PET and question were assessed using z tests.

Subjective comprehension. Participants rated their subjective confidence in understanding the given PET description using a 5-point scale (1=Not at all confident, 5=Very confident). Alongside open-ended interpretation responses, this metric helps identify instances of misplaced confidence. Participants who rated their confidence below the "confident" threshold were prompted to specify which words or sentences were hard to understand (see *Confusing phrasing within PET descriptions* in Section 3.2.4).

Prior PET familiarity. This was measured as a control variable. Responses to the question, *"Have you ever heard of the following technologies? (select all the apply),"* were analyzed in two ways: by the count of technologies selected and the distribution of these selections. To determine if the average number of selections varied significantly across conditions, we conducted a Kruskal-Wallis test. Due to the lack of viable statistical tests for comparing distributions in check-all-that-apply responses, we visually assessed if specific PETs were selected more frequently by participants in related conditions using a proportion bar plot.

PET identification. Since the PET descriptions were provided without mentioning their names, participants were asked to identify the type of PET described. Correct identifications were scored as 1, and incorrect ones as 0. For the FL+LDP condition, selecting either FL or LDP was considered correct; for the GT condition, choosing either Topics or FLoC was considered correct. We also included a decoy option, "deliquescent security," which doesn't exist.

3.2.4 Qualitative measures and analyses. Participants provided four types of qualitative responses throughout the survey, as detailed below. Each type was analyzed separately. The first author singled-coded all responses using inductive coding methods[48]. The detailed codebooks are available in Appendix C.

Perceived protection of user data. Participants were asked to describe, in at least two sentences and in their own words, how they believe the fictional social media platform protects user data. We first analyzed responses from each experimental condition separately to understand the effect of including an implications statement in PET descriptions on users' perceptions. Comparisons between responses from conditions with and without "impl" for each PET showed no significant variation. Therefore, we developed specific codebooks for the responses corresponding to each PET.

Interpretation of behavioral data. Participants were asked to articulate their interpretation of "behavioral data" in the given PET description. These responses were expected to uncover perceived privacy risks associated with this type of data, offering insights for comparison with health data privacy concerns discussed in existing literature. They also helped contextualize participants' understanding of PETs. Responses from each PET's conditions were first analyzed separately. Given the lack of variation in responses, a comprehensive codebook was developed by analyzing responses across all conditions.

Confusing phrasing within PET descriptions. Participants rating subjective comprehension below "confident" were asked to pinpoint confusing words or phrases in the description. We considered responses above this confidence threshold to indicate a comfortable understanding of our PET descriptions, and focused on responses below this threshold to identify confusing aspects of the PET descriptions. These responses were coded by condition.

Interpretation of PET description segments. Participants were asked to explain their interpretations of specific segments from the PET description provided, with questions tailored to each experimental condition (see Table 6 for the exact questions asked per condition). Participants in the LDP and FL+LDP conditions were asked two questions targeting different aspects of the descriptions to ensure comprehensive coverage of response content, addressing deficiencies identified in pilot tests. We labeled responses by correctness to determine accurate and inaccurate understandings of PETs. This involved a two-stage labeling process: we first did coarse coding, and then refined the codes to categorize the responses as either accurate or inaccurate. Responses were coded separately for each PET.

Due to the varying nature of PETs, it is difficult to compare across PETs. Our RQ2 and RQ3 thus center on comparing the effects of descriptions with and without an implications statement for each PET on user comprehension, and RQ4 centers on users' mental models of each PET. The qualitative answers in our survey offer insights for indirectly comparing user perceptions of various PETs.

Table 2: Demographic statistics of survey respondents (*n*=306). Census statistics are exported from or computed based on American Community Survey five-year estimates [9–11].

	Category	Sample (%)	Census %
Age	18 – 24 years	46 (14.9%)	13.3%
	25 – 34 years	104 (33.8%)	13.7%
	35 – 44 years	81 (26.3%)	12.9%
	45 – 54 years	24 (7.8%)	12.4%
	55 – 64 years	22 (7.2%)	12.9%
	65 years and over	29 (9.4%)	16.5%
Gender	Woman	143 (46.6 %)	51.1%
	Man	154 (50.3%)	48.9%
	Non-binary	5 (1.6%)	N/A
	Prefer not to answer	4 (1.3%)	N/A
Race	Asian, Indigenous Peoples	21 (6.8%)	6.0%
	Black or African American	24 (7.8%)	12.5%
	Hispanic, Latino, or Spanish	15 (5.2%)	18.7 %
	White	230 (75.2%)	65.9%
	Two or more races	7 (2.3%)	8.8%
	Prefer not to answer	8 (2.4%)	N/A
Education	Advanced degree	47 (15.3 %)	12.4%
	Associate's degree	26 (8.5%)	7.7%
	Bachelor's degree	120 (39.2%)	19.0%
	High school graduate	37 (12.0%)	27.2%
	Some college but no degree	72 (23.5%)	13.8%
	Prefer not to answer	4 (1.3%)	N/A
Tech background	Yes	53 (17.3 %)	N/A
	No	242 (79.1%)	N/A
	Prefer not to answer	12 (3.6%)	N/A

3.3 Participant Demographics

Table 2 shows our participants' demographics (n = 306). Our sample encompassed a broad age range, with a median of 35, was predominantly white, and about half were women (51%). Participants also had marginally higher education levels compared to general census data. Participants were randomly assigned to experimental conditions to control for potential confounding variables. Analysis confirmed no significant irregularities in the distribution across conditions based on age, gender, race, and education level. Most did not have a technological background, with 6% to 30% of participants across conditions having tech-related expertise, peaking at 30% in the FL+LDP condition. However, no significant effect was found from varying tech backgrounds on the outcomes.

Additionally, we analyzed the following measures—prior PET familiarity and PET identification—to examine if significant differences in the pre-conception of PETs exist between participants across experimental conditions:

Prior familiarity with PETs among our participants was generally low, with only about 5% recognizing the PET names. GT was an outlier, with awareness spiking to 25%, whereas no participants recognized FLoC despite its relevance to GT. This discrepancy likely results from people being familiar with the term "topics" in the name but not necessarily recognizing it as a specific privacy technology. LDP achieved the highest familiarity, possibly benefiting from the explicit inclusion of the term "privacy" in its name.

In identifying the PET described, LDP-related conditions showed the highest correctness rates (53.3% for LDP-Xiong; 40.0% for LDP; 48.3% for LDP-impl), with FL+LDP at 55.7%. Lower correctness rates were noted for FL (12.9%) and GT (11%). The higher recognition of LDP may be due to its longer history and public exposure through notable applications such as the U.S. Census and Apple's privacy initiatives. These observations highlight the influence of pre-existing public knowledge and question design on responses in survey-based privacy technology research.

3.3.1 *Data quality.*

To ensure the quality of participants' responses, we triangulated their quantitative answers from (1) the Likert-scale question assessing their self-reported confidence in PET comprehension, and (2) the true/false question measuring their objective comprehension of PETs, and their qualitative responses to (3) the open-ended question where they interpreted specific segments of PET descriptions in their own words. We filtered out timed-out, low-effort, and exceptionally fast survey responses and analyzed a total of 306 responses. Among those, we found no instances of incoherence, such as high subjective confidence paired with low objective comprehension scores and low-quality or self-contradictory qualitative responses. Additionally, there were no significant variations in coherence levels across experimental conditions.

3.4 Limitations

We acknowledge several limitations in our study. First, our survey did not account for factors besides PET descriptions that might influence users' perception of PETs, such as risk preference and perception, trust, and mental models of ad tracking and analytics services. We noticed that a few participants expressed skepticism toward ad services in their open-ended responses, which could skew their perception of the provided PET descriptions. These unaddressed factors may be valuable for future research to further explore the relationships between PET descriptions and user perceptions of PETs.

Another limitation is the potential discrepancy between users' perceptions of PETs in real-world settings versus their responses to PET descriptions in a survey setting. However, the controlled nature of our survey experiment, with its varied and paired textual descriptions across different conditions, supports the internal validity of our findings. Moreover, the structured survey environment facilitates participants' attentive engagement with PET descriptions. Lastly, our survey's geographical scope was limited to the U.S., which may restrict the generalizability of our findings across different cultural and regulatory contexts concerning privacy.

4 Results

Section 4.1 addresses RQ1 by evaluating the effectiveness of LDP descriptions adapted and developed for the ad tracking and analytics context. Sections 4.2 to 4.4 examine the impact of including implications statements into the descriptions for LDP, FL, and GT, respectively, covering RQ2 and RQ3. Finally, Sections 4.5 to 4.7 discuss participants' accurate and inaccurate understandings of the descriptions for LDP, FL, and GT, respectively, in response to RQ4. Table 3 summarizes our main findings.

4.1 Process- and implications-focused LDP description works well in the ad tracking and analytics context *(RQ1)*

Overall, our LDP-impl description, which includes both the process and implications, and LDP-Xiong—an adapted version of Xiong et

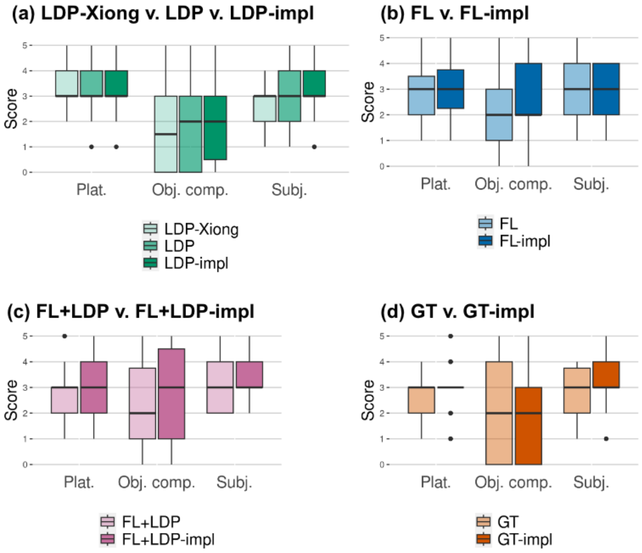


Figure 1: Grouped boxplots comparing confidence in platform use (Plat.), objective comprehension (Obj.), and subjective comprehension (Subj.) across condition pairs in RQs 1–3. Mann-Whitney test results for RQs 1–3 are summarized in Table 7 in Appendix D. Apart from the confidence in platform use comparison in RQ1, none of the other comparisons reached significance at the 0.05 level.

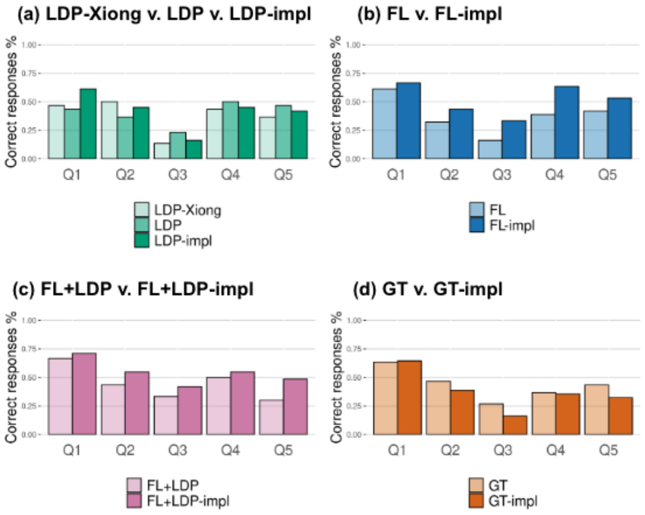


Figure 2: Grouped bar plots comparing the proportions of correct responses of each objective comprehension question (Q1-5) across condition pairs in RQs 1–3. None of the comparisons reaches statistical significance at the 0.05 level in *z* tests.

al.'s LDP description that also combines process and implications—showed similar effects on user comprehension. However, our LDP-impl description was associated with significantly higher subjective comprehension (r=0.7, p<0.05, scale=$[1, 5]$) compared to LDP-Xiong. Despite this, there were no significant differences in objective comprehension score or confidence in platform use between the two conditions (Figure 1(a)). In objective comprehension scores, participants in the LDP-impl condition (median=2.0, scale=$[0, 5]$) slightly outperformed those in LDP-Xiong (median=1.5, scale=$[0, 5]$), with similar correctness rates on individual comprehension questions (Figure 2a). Both groups reported comparable levels of confidence in platform use (median=3.0, scale=$[1, 5]$). About 81% of participants (25 out of 31) in the LDP-Xiong condition reported a confidence level below four on subjective comprehension, compared to 61% (19 out of 31) in the LDP-impl condition. We considered responses above the confidence threshold of four, which corresponds to "confident" in our Likert-scale question on subjective comprehension, to indicate a comfortable understanding of our PET descriptions, and focused on responses below this threshold to identify confusing aspects of our PET descriptions. In general, data-related terms, such as "subset" and "aggregated data," caused confusion. The use of the term "subset" led some participants to question the total size and proportion of data transmitted.

In the LDP-impl condition, some participants found the terms "noise" and "noisy data" difficult to understand. This confusion may be due to the different meanings of "noise" in technical versus everyday contexts. One participant noted: *"It seems like an interesting word choice to refer to a technical term."* Some participants specifically mentioned not understanding how noise is added to or "*merged with personal data.*"

In contrast, when details about the random modification of data were omitted, as in LDP-Xiong, more participants demanded clarity. About half wanted to know the specifics of random modification (e.g., *"I'd like to know exactly how my data would be modified. Random is too broad of a word."*). Questions also arose about how data modification protects privacy (e.g., *"I would like to know more about how the modification protects privacy, like what the modification is exactly."*). Participants seeking more information on data modification generally reported lower subjective confidence in their understanding of the PET, with five rating a confidence level of 1. Further, the phrase "additional privacy technique" led to confusion for some participants, with one commenting: *"The additional privacy technique statement is confusing. Technique is a word not used much in describing software terminology."* Despite our intention to align LDP-Xiong with other PET descriptions, this feedback suggests that the phrase was perceived as too vague.

Summary RQ1: Our findings indicate that LDP-Xiong and LDP-impl have a comparable impact on participants, showing similar levels of objective comprehension and confidence in platform use decisions. However, participants in LDP-impl reported greater subjective confidence in their understanding of the PET. This suggests that our LDP-impl description is at least as effective, if not more so, than LDP-Xiong adjusted for the ad tracking and analytics context. The results show that the approach we took to develop PETs, which expanded the explanation of the data modification process in the refinement of Xiong et al.'s LDP description, performs marginally better than Xiong et al.'s original approach. Moreover, by comparing descriptions of LDP where parallels from existing work exist, we observed that the process- and implications-focused approach to PET descriptions, which was found effective in health settings, is also effective in the context of ad tracking and analytics.

4.2 LDP implications statement has limited effect in ad tracking/analytics context *(RQ2)*

Our findings indicate that the process-only LDP description is as effective as the version that includes an implications statement in terms of enhancing user comprehension. No significant differences were observed in objective comprehension scores, subjective comprehension levels, or confidence in using the platform between the LDP and LDP-impl conditions. Both conditions showed the same median values for confidence in platform use (*median*=3.0, *scale*=$[1,5]$), objective comprehension score (*median*=2.0, *scale*= $[1,5]$), and subjective comprehension (*median*=3.0, *scale*=$[1,5]$) (Figure 1(a)). The correctness rates of individual objective comprehension questions did not significantly differ between the two conditions (Figure 2(a)). In the LDP condition, 60% of participants (18 out of 30) reported a confidence level below four on subjective comprehension and highlighted parts of the description they found challenging, similar to 61% (19 out of 31) in the LDP-impl condition. More than half of the participants in the LDP condition expressed confusion regarding the concept of "noise" or "noisy data." In both conditions, some were puzzled by the term "subset" and the extent of user data shared with the platform.

Summary RQ2: Participants showed moderate levels of comprehension of the process-focused LDP descriptions, and terms like "noise" and "subset" in the process descriptions confused users. Although having an implications statement did not improve user comprehension in our study, which contrasts with prior work's findings, we also did not observe negative effects or additional confusion from including implications.

4.3 Process+implications approach can be adapted to federated learning *(RQ3)*

Overall, we found that our process+implications descriptions for FL and FL+LDP were understandable for participants; however, incorporating an implications statement did not show a significant impact here either. There were no notable differences in confidence in platform use, subjective comprehension, and objective comprehension scores between the descriptions with and without an implications statement for FL and FL+LDP. Both conditions with and without "impl" in FL recorded the same median scores for objective comprehension (*median*=2.0, *scale*=$[0,5]$), subjective comprehension (*median*=3.0, *scale*=$[1,5]$), and confidence in platform use (*median*=3.0, *scale*=$[1,5]$) (Figure 1(b)). Although not statistically significant, the median of objective comprehension scores of the FL+LDP-impl condition (*median*=3, *scale*=$[1,5]$) was slightly higher compared to FL+LDP (*median*=2, *scale*=$[1,5]$). This trend is further corroborated by the correctness rates of individual objective comprehension questions (Figure 2(c)). Incorporating an implications statement in the FL+LDP description may therefore have a marginally positive effect on user comprehension. Qualitative analysis of open-ended responses shows that the confusing aspects of descriptions with and without "impl" are largely consistent across FL variants, suggesting that the descriptions similarly influence subjective comprehension. We discuss points of confusion below.

4.3.1 FL: Confusion about machine learning models. In the FL condition, 18 of 31 participants (58.1%) reported a confidence level below four on subjective comprehension, and in the FL-impl condition, 21 of 30 (70.0%) did so. Among these participants, more than half in both conditions were confused about how machine learning models are trained using user behavioral data. This confusion stemmed partly from unfamiliarity with the term "machine learning." One of them wrote, *"Machine learning—is this AI?"* Another questioned the data processing, stating, *"I do not understand how it will infer patterns if it does not track my specific behavioral data...I just do not see how it can even [give] a general idea without first getting and tracking specifics."* Additionally, there was confusion about how the exact behavioral data used by the models differs from the model outputs shared with the platform, regarding the sentence contrasting the FL and FL-impl descriptions. One respondent wrote, *"I would want some more information about how non-personal the final inferred interests are as it is not clearly stated."* Less frequently, participants in both conditions indicated a lack of a clear understanding of how machine learning models can be merged.

4.3.2 FL+LDP: Confusion about "noise" and machine learning models. In the FL+LDP condition, 20 out of 30 participants (66.7%) reported a confidence level below 4, while in the FL+LDP-impl condition, 22 of 31 (70.9%) did so. Similar to the LDP conditions, participants in both FL+LDP conditions expressed confusion over terms like "noise" and "noisy data." Additionally, they sought clarity on how adding noise protects user privacy, with questions like, *"if there's a way to easily remove the noise data in order to see an individual's behavior"* and *"how obvious is it to not be fake data which could be discarded?"* As in the FL-related conditions, confusion also arose regarding how the machine learning model processes user behavioral data and infers user interests. Occasionally, participants in the FL+LDP condition questioned the utility of user data altered by noise for an organization aiming to understand user interests.

In summary, while the descriptions we developed for FL and FL+LDP that focus on the processes are understandable for some participants, incorporating an implications statement into the descriptions did not significantly improve user comprehension. Additionally, the mention of machine learning in FL descriptions and the data modification by noise addition process in FL+LDP descriptions, which are intended to enhance privacy, caused confusion for some.

4.4 Process+implications approach also adaptable to Google Topics *(RQ3)*

Overall, focusing on the process in PET descriptions was also effective for GT, though incorporating an implications statement did not significantly impact the outcomes. Comparisons between GT and GT-impl showed no significant differences in objective comprehension (*median*=2.0, *scale*=$[0,5]$), confidence in platform use (*median*=3.0, *scale*=$[1,5]$), and subjective comprehension (*median* =3.0, *scale*=$[1,5]$) (Figure 1(d)). Qualitative analysis indicated confusion among participants about the on-device inference of user interests in both GT conditions.

In the GT condition, 22 of 30 participants (73.3%) reported a confidence level below four, and in the GT-impl condition, 19 of 31 (61.3%) did so. Participants in both conditions were puzzled by the platform's method of deducing their interests without direct access to user behavioral data. One participant questioned: *"It says that the platform [uses] your behavioral data from your device but then it* *says that the exact behavioral data is never sent to the organization so it seems contradictory. Is the data sent or not?"* This reflects a misunderstanding about how behavioral data is utilized to deduce interests and the specific implications of processing "on your device." For some, this confusion extended to the mechanics of data storage and usage, questioning whether data remains solely on the user's device or if the technology uses only the data generated by the device. Other concerns arose about the data collection process itself, with one participant noting, *"It is conflicting to understand privacy technology...that monitors or collects information based on your personal search history...If the information is not sent to an organization then who is collecting it?"*

In summary, our process-focused GT descriptions are understandable for some users, although incorporating an implications statement did not improve user understanding. Some participants found the local inference of user interests, the core functionality of the PET, challenging to understand, among other points of confusion.

Summary RQ3: Our findings suggest that the process-focused description approach also works well with other PETs, specifically for FL, FL+LDP, and GT in the ad tracking and analytics context. However, including an implications statement in those PET descriptions did not significantly improve comprehension, but it also did not introduce further confusion. Technical terms like "machine learning" (in FL) and "noise" (in FL+LDP) and the phrase of on-device inference of user interests (in GT) confused some participants. Comparing across PETs, our results indicate that GT might be more challenging for users to understand, as shown by the distribution of objective comprehension scores.

4.5 (In)accurate understandings of LDP *(RQ4)*

Unlike RQs 1–3 which evaluated the implications statement's effect on user perception and the extent to which the process-oriented approach extends to other contexts and PETs, RQ4 delves into user perception of specific segments of PET descriptions. In this and the following subsections, we detail both accurate and inaccurate understandings by participants for each PET studied.

4.5.1 LDP-Xiong: Accurate understanding of data sharing but not data modification. With LDP-Xiong, which closely matches Xiong et al's original LDP description for the health context [72], some participants accurately understood the data sharing aspect of LDP, recognizing that user data transmitted to the platform is altered. One participant explained, *"Aspects of your data will be jumbled up randomly before the company gets it. The company won't be able to see your data."* Some understood the protective attributes of modified data, noting it *"will not align with your original data"* and is *"not directly associated with you,"* thereby the company gains some information about users *"not knowing who exactly you are."*

However, the concept of "random" data modification confused about half of the participants. Some interpreted "randomly modified" as altering personally identifiable or sensitive information to prevent recognition, noting it would *"change identifiers...so they can't tell it's you"* or involve *"removing anything that would identify you,"* like names or other specific identifiers. While some recognized the privacy-preserving benefit of data aggregation, they overlooked the crucial source of privacy protection—individual randomization of data before aggregation.

4.5.2 Our LDP descriptions: Improved understanding of data modification, with confusion around "noise". Compared to LDP-Xiong, our LDP descriptions led to a qualitatively better understanding of LDP's data modification through noise addition. One participant described it as, *"Adding noise to the data means that the data is collected and modified, so some parts may be changed, adding new things and/or taking some away."* Furthermore, some correctly recognized this method as a means of enhancing user privacy, with one stating, *"the company basically adds a layer of protection into our data."* They also understood that the modified data reflects general trends but does not precisely replicate individual user behavior.

However, some participants were curious about the specific type of "noise" added to data and exhibited a nuanced yet not entirely accurate grasp of the data modification process. While broadly correct, their perceptions diverged from the precise technical processes due to the abstraction inherent in these statistical and technical operations. A more accurate interpretation of "noise" by some participants included content they never interacted with or misrepresented time spent on a page. Others imprecisely associated noise addition with obscuring sensitive details like IP addresses, describing it as *"Random or blank data will be added to block out things that are important in our datasets."* This indicates an understanding that data modification protects privacy, but with a misconception that it selectively protects crucial information while leaving non-sensitive behavioral data clear for analysis and use.

With our LDP descriptions, similar to LDP-Xiong, most participants correctly understood the data sharing aspect and recognized that the organization utilizes merged data across users. Some acknowledged that organizations can identify user behavioral patterns without accessing exact data or directly linking behaviors to specific users. Some were aware of the organization's capability to deduce user preferences from modified or anonymized data. While participants recognized the merging of user data, they associated the addition of noise with data merging, interpreting it as *"blending my data with other users' data."*

In summary, both LDP-Xiong and our LDP descriptions effectively conveyed the data sharing aspect of local differential privacy to many participants. By providing a more detailed explanation of data modification through noise addition, our descriptions enhanced participants' comprehension of how user data is altered to safeguard privacy. Yet, consistent with findings from RQs 1–3, the term "noise" caused confusion for some participants.

4.6 (In)accurate understandings of FL and FL+LDP *(RQ4)*

Since LDP adds privacy-preserving features to FL in FL+LDP, we present results for both FL with and without LDP below, and briefly discuss the influence of the additional privacy-preserving feature of LDP on user comprehension through a rough comparison at the end of this subsection.

4.6.1 FL: Understanding model merging is challenging, possibly due to confusion around machine learning. The distinction between merging user data and merging models highlights a significant difference between FL and LDP regarding the organization's access to user data. However, understanding the sharing and merging of models that represent user interests in federated learning was challenging for many participants. Some comprehended well that with federated learning machine learning models deduce user interests on the user's device before being merged with other users' models. They also clearly understood the model sharing aspect, noting that specific behavioral data is not transmitted to the organization; rather, the individual models are. One participant explained, *"It means that a machine learning model acts as a middleman. Direct data is not sent to the organization."*

However, many did not realize that precise user data is not necessary for organizations to learn about user interests. About half thought user data is shared and then aggregated, and some associated the merging of user data with anonymization, thinking *"It means that your data is pooled with other users so that everyone remains anonymous."* Others believed that merging data safeguards privacy by ensuring that no single user's data can be isolated.

Some participants misunderstood the model merging process, believing that machine learning models are trained on aggregated user data from multiple individuals, as one stated, *"[your data] is blended with a bunch of [others'] data to make patterns clear, rather than inferring them from your direct history."* Rather than seeing these models as intermediate processes whose outputs are aggregated for further analysis, some viewed them as technologies for tracking and collecting user behavior: *"the platform will use an AI model to track patterns in online behavior."*

4.6.2 FL+LDP: Accurate understanding of modified data in machine learning training; "noise" and model merging challenges persist. For the combination of FL and LDP, over half of participants understood that data used for machine learning training is modified for privacy: *"the platform purposefully introduces random changes (noise) to your behavioral data to protect your privacy before using it to build a machine learning model that forecasts your interests."* Some correctly linked these modifications to privacy protection, noting, *"it helps protect users from being completely identifiable by the platform based on behavior and interests"* and *"[it] will protect you from the organization's employees or being compromised."*

Despite grasping the general idea of data modification, as in LDP, some participants misunderstood "noise", viewing it as random actions not representative of actual behavior, like *"random clicking or random long pauses where the site almost times out"*, as the introduction of external data that (*"isn't exactly yours"*), or as a method of anonymization by introducing "fake" data to mask true information.

Meanwhile, some participants correctly recognized the distinction between sharing direct data and sharing insights derived from models. Some also accurately understood that these learned models, which represent user interests, are shared and merged with other model representations of user interests, noting, *"Your behavioral data is only sent as a model to the organization. This model is also merged with others to create patterns."* However, about half of the participants misunderstood model merging, believing user data is combined with data from other users: *"Aggregated data is being used to train the ML model. Apparently this reduces the risk to an [individual's] privacy."* Some, assuming that the organization could access precise user data, interpreted large-scale data aggregation as a means of anonymizing individual data.

In summary, understanding "models“ and model merging remains challenging in both FL and FL+LDP. In FL conditions, we observed confusion about the functionalities of machine learning models and misunderstanding about model/data sharing. In FL+LDP conditions, many respondents grasped the role of machine learning models and model merging well, yet these concepts still posed significant challenges for others, as observed in the FL conditions. The use of the term "noise" in data modification introduced further difficulties, mirroring those in LDP conditions.

4.7 (In)accurate understandings of GT *(RQ4)*

Regarding Google Topics, participants had mixed understanding of on-device inference of user interests. About half of the participants correctly understood that behavioral information is processed on the user's device to identify top ad topics, with only these top inferred topics being shared. They also noted that the shared topics include a degree of randomness and are not exclusively based on the user's data. One participant explained: *"The platform takes small samples of data that come from your top topics in the previous week or from a completely random topic to use for what you may be interested in. This way content can more or less be near what a user wants, while still allowing for variability."* Moreover, some accurately understood that due to this randomness, the shared topics might not precisely reflect the individual user.

In contrast, some participants were unclear about or overlooked how ad topics are derived directly on users' devices, instead focusing on concerns about surveillance and the perceived intrusiveness of technology. This confusion often stemmed from a misunderstanding of machine learning models' role, with some mistakenly believing that these models—referred to as "AI" or "the machine"—actively track and record user behavior. One participant stated, *"it 'records' my behavior to figure out use patterns, likes, pages viewed, activities or shopping, etc. It literally records all of what I do online."* Additionally, some incorrectly assumed that a portion of user data is directly transmitted to organizations by those models, noting, *"it looks at what you are checking out online. However, it's only selecting a few bits here and there to send to the owners of the company."*

In summary, half of the participants correctly grasped on-device machine learning for inferring user interests, while others were distracted by concerns about user tracking. This aligns with common confusion with machine learning models in LDP, FL, and FL+LDP.

Summary RQ4: The PET descriptions we developed worked generally well, and most respondents were able to accurately grasp the key features of PETs, such as data modification, data/model sharing, and data/model merging. Comparing across PETs, we find that respondents are more likely to correctly understand the concept of merging user data than merging models, including the associated training and sharing of these models. Meanwhile, respondents often perceive the merging of individual data with that of others as the source of privacy protection, expecting individual data to become unidentifiable post-merging. This tendency persists even in FL and FL+LDP, where data merging neither occurs nor is mentioned in our descriptions.

Table 3: Summary of findings: Evaluating our process-and-implications descriptions of PETs—LDP, FL, FL+LDP, and GT—in ad tracking and analytics, including the LDP-Xiong [72] description we adapted from the health domain.

RQ	Findings
RQ1 (LDP-Xiong v. LDP-impl)	- Objective Comprehension & Platform Use Confidence: LDP-impl and LDP-Xiong show comparable effects. - Subjective Confidence: LDP-impl > LDP-Xiong. - Our refinement of Xiong et al.'s LDP description, with expanded data modification explanations, marginally outperformed the original. - The process+implications approach we used to develop our PET descriptions worked well in both ad tracking/analytics and health.
RQ2 (LDP v. LDP-impl)	- Moderate understanding of process-focused LDP descriptions among participants. - Adding an implications statement did not significantly enhance, but did not negatively affect, user comprehension. - Terms like "noise" and "subset" in the descriptions added some confusion.
RQ3 FL(+LDP) v. FL(+LDP)-impl, GT v. GT-impl	- Our process+implications approach can be adapted to FL, FL+LDP, and GT. - Including an implications statement in these PET descriptions did not significantly enhance comprehension but also did not add confusion. - Confusing technical terms: FL–"machine learning"; FL+LDP–"noise"; GT–on-device inference of user interests.
RQ4 (users' mental models)	- Many respondents accurately grasped key PET features like data modification and data/model sharing. - Respondents showed better comprehension of merging user data over merging models. - Challenging aspects of model merging include the related machine learning training processes and model sharing with platforms.

5 Discussion

5.1 Key findings

General audiences often struggle to understand the privacy protection mechanisms of PETs, which hinders the development of a nuanced understanding of their ability to protect privacy [18, 36, 65, 67, 72]. Much of the existing work on communicating PETs to users has focused on (local) differential privacy and often in the health context [28, 45, 51, 73]. We expanded the promising approach of process- and implications-focused descriptions into the context of ad tracking and analytics and developed and tested descriptions for FL, FL+LDP, and GT—PETs for which descriptions were previously not studied, yet which are increasingly used in practice. Our findings, summarized in Table 3, highlight the challenges in describing and understanding PETs. Based on our findings we offer recommendations for effectively developing PET descriptions to enhance comprehension.

The effect of our refinement of LDP descriptions. Our comparative analysis shows that our LDP description tailored for ad tracking and analytics, which elaborates on the PET's process, increases users' subjective comprehension, though it does not significantly improve confidence in platform use and objective comprehension, compared to Xiong et al.'s health-app based LDP description [72]. Most participants appreciated the explanation of the data modification process in our description. In contrast, when such details were omitted as in LDP-Xiong, about half found the concept confusing. Our additional clarification about the sharing of modified user data also increased participants' confidence in their own understanding. Nevertheless, the term "noise" in our descriptions posed difficulties for some participants, leading to somewhat inaccurate assumptions about its manifestation in user behavioral data. Our findings corroborate Xiong et al. [72] regarding users' struggles with technical terms and provide a more detailed account of how users may misunderstand such terms. Together, our results suggest that the process- and implications-focused approach to describing PETs is effective not only in health settings but also in other contexts, such as ad tracking and analytics.

The effect of the implications statement in PET descriptions. Prior work has found positive effects of implications statements for DP/LDP in the health context [47, 72]; yet, we could not replicate a significant impact of adding implications for LDP and other PETs in ad tracking and analytics. Although the process-focused descriptions of LDP, FL, FL+LDP, and GT are helpful for users, incorporating an implications statement into the descriptions did not measurably increase user understanding in terms of objective comprehension score, subjective comprehension, and confidence in platform use. Despite the lack of a significant positive effect, the implications statement also did not introduce confusion, except for GT. This result may suggest that implications statements that we adapted from the health context are less effective in ad tracking and analytics. More research is needed to craft implications statements specifically suited for behavioral data and further examine their phrasings and salience, such as the location within the description, to improve usefulness.

Users' perception of PET descriptions. We are among the first to develop and evaluate user-centric descriptions for FL, FL+LDP, and GT, with a focus on clarifying how these PETs protect user privacy. We also provide comprehensive insights into users' mental models of these PETs, observing both accurate and inaccurate understandings, that extend beyond those from prior work, which has focused on users' misconceptions of PET operations and their impacts [36, 65, 67]. Our findings reveal a mixed understanding among participants. For FL and FL+LDP, some accurately understood the role of machine learning models in discerning user interests and the concept of model sharing. However, many struggled with the concept of on-device model training and the idea of merging models to derive general behavioral patterns. Similar to the issues we saw for LDP descriptions, the term "noise" in FL+LDP descriptions posed comprehension challenges for some participants. For GT, while some participants clearly understood the on-device processing of data, others failed to recognize this feature and instead expressed concerns about potential data tracking and collection practices. These findings underscore the complexities in effectively communicating the functionalities and roles of machine learning and "models" in PET descriptions, marking an area for future research aimed at improving user comprehension. However, overall, the findings from our LDP, FL, FL+LDP, and GT descriptions show that the process+implications approach to describing PETs is generally effective. While it is difficult to compare across PETs due to their varying nature, it seems that the process of GT is more challenging to describe and understand accurately.

5.2 Implications for describing PETs to users

To enhance the clarity of specific functional aspects of PETs, like data modification in LDP, we attempted to avoid technical jargon in the descriptions. Here, we outline potential strategies to further improve the comprehensibility of PET descriptions for LDP, FL, and GT within the context of ad tracking and analytics.

Avoid noise, machine learning, and other jargon. Technical details like noise, random modification, and machine learning models often posed comprehension challenges in our study. While we avoided statistical terminology, the term "noise" still confused users, leading to a superficial understanding of data modification processes. Users expressed a desire to comprehend the post-noise modification state of their data and its practical implications, such as changes in behavioral metrics, and the term "noise" alone was insufficient for a clear understanding of PETs' workings. To address this, alternative methods of explaining the concept of statistical noise and its impact, as suggested by Bullek et al. [8], could be more effective than textual explanations. Furthermore, referring to the use of machine learning in PET descriptions sometimes redirected users' attention towards concerns about data tracking and recording, overshadowing the privacy-preserving aspects of these technologies. To prevent this, descriptions could highlight even more how these technologies protect privacy in addition to how they function.

Help users pinpoint the source of privacy protection. Enhancing user understanding of PETs requires a robust explanation of their protective mechanisms in the implications statement. Participants often struggled to identify the source of privacy protection, whether through data modification and the sharing of the modified data (LDP), sharing machine learning models without actual data (FL), sharing machine learning models trained on modified data (FL+LDP), or sharing subsets of inferred topics that may include random ones (GT).

Our study highlights the need for clearly delineating the risks that privacy protections are designed to mitigate, including the risk of identity disclosure, sensitive information leaks, third-party sharing, and data breaches. Enhancing clarity regarding the origin of privacy protection can improve users' comprehension of both the privacy risks and the protections provided by PETs. Accurate understanding of these is crucial to prevent unintended harms that may arise from misconceptions about what PETs can or cannot mitigate [36, 65, 67].

Our results suggest that the implications statements we adopted from Xiong et al. [72] might not sufficiently clarify these risks. Additionally, explaining "behavioral data" within the context of ad tracking and analytics, and differentiating between levels of data sensitivity, could further aid understanding. In the context of ad tracking and analytics, what might provoke privacy concerns could be perceived as less sensitive than data in health contexts [72].

Since our study, Google Topics has become a default feature in Google Chrome, as detailed in Chrome version 122.0.6261.69, listed in Table 5 in Appendix A.3. Considering our findings, Chrome's description of GT could be improved by providing clearer information about the specific data Chrome collects and utilizes and how it deduces user interests from user data. This clarification is crucial as our findings suggest that users might incorrectly assume that their precise behavioral data is always shared with Google. Furthermore, although Chrome's current description effectively outlines the technology's mechanisms, it lacks specific references to the sources of privacy protection and does not include a statement on its implications, which could leave users uncertain about the balance between the intrusiveness of technology and its protective benefits.

Provide more specificity about user data. Common confusions and misconceptions across PETs we observed were often associated with users' desire to understand the nuances of data handling—what data is collected, where it is analyzed, and by whom. This clarity is vital, particularly for PETs like FL and FL+LDP, where data processing occurs on the user's device rather than a centralized system accessible to the organization. Users need a basic understanding of where their data is stored when not shared with the organization—typically on their device—to appreciate the privacy protections offered. Furthermore, although not directly related to PETs, users frequently seek clearer explanations on how their behavioral data is tracked and collected, whether by the company directly, through machine learning models, or by other entities. There is also a significant demand for information about the types of data organizations aim to extract, how they acquire it, and why this data remains valuable to them even when modified.

6 Conclusion

Our study adapted well-performing textual descriptions of local differential privacy (LDP) from the health context to ad tracking and analytics. We further developed new user-centric descriptions for other prevalent PETs in this context, namely federated learning (FL) both with and without LDP, and Google Topics (GT). Our survey experiment (n=306) examined the applicability of previous findings to these expanded contexts, the effect of incorporating an implications statement in PET descriptions, and users' perception of PET descriptions. We found that the process- and implications-focused PET description approach was generally effective for describing PETs in the ad tracking and analytics context. The process+implications descriptions we adapted and developed for LDP, FL both with and without LDP, and GT in this new context were helpful for user understanding, although the implications statements found crucial in health contexts [47, 72] had no significant effect on user comprehension in our study. This suggests a potential area for future research to explore more context-specific phrasing and presentation of implications statements in non-health PET descriptions. Our findings also provide new insights into misconceptions about privacy protection mechanisms and the challenges in accurately conveying the sources of privacy protection in the descriptions, which would help further improve user-centric PET descriptions.

Acknowledgments

This research has been partially supported by a Google Privacy Research Faculty Award and by the National Science Foundation under Award No. 2105734. Jane Im has been supported by a Meta Graduate Research Student Fellowship. We thank the anonymous reviewers for their constructive feedback, as well as Alain Forget for his input and suggestions. The authors used Grammarly for proofreading.

Contribution Statement

The authors confirm contribution to the paper as follows:

- Lu: Quantitative and qualitative analysis (lead), writing (lead).
- Song: Quantitative and qualitative analysis (support), writing (support).
- Jane: Study conception, study design, data collection, quantitative analysis (support), proofreading, funding acquisition (support).
- Florian: Study conception, study design, analysis, writing, guidance, funding acquisition (lead), project supervision.

References

[1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (2015), 509–514.

[2] Idris Adjerid, Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2013. Sleights of privacy: Framing, disclosures, and the limits of transparency. In *Proceedings of the ninth symposium on usable privacy and security*. 1–11.

[3] Apple. n.a.. Apple Differential Privacy Technical Overview. https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf Accessed on 14 February 2024.

[4] Engineering at Meta. 2022. Applying federated learning to protect data on mobile devices. https://engineering.fb.com/2022/06/14/production-engineering/federated-learning-differential-privacy/ Accessed on 14 February 2024.

[5] Alex Berke and Dan Calacci. 2022. Privacy limitations of interest-based advertising on the web: A post-mortem empirical analysis of Google's FLoC. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*. 337–349.

[6] IBM Research Blog. 2022. What is federated learning? https://research.ibm.com/blog/what-is-federated-learning Accessed on 14 February 2024.

[7] Dieter Bohn. 2021. Privacy and ads in Chrome are about to become FLoCing complicated. https://www.theverge.com/2021/3/30/22358287/privacy-ads-google-chrome-floc-cookies-cookiepocalypse-finger-printing Accessed on 14 February 2024.

[8] Brooke Bullek, Stephanie Garboski, Darakhshan J Mir, and Evan M Peck. 2017. Towards understanding differential privacy: When do people trust randomized response technique?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 3833–3837.

[9] U.S. Census Bureau. 2022. ACS Demographic and Housing Estimates. U.S. Census Bureau. https://data.census.gov/table/ACSDP5Y2022.DP05?q=UnitedStatessex&g=010XX00US&moe=false Accessed on 14 February 2024.

[10] U.S. Census Bureau. 2022. Age and Sex. U.S. Census Bureau. https://data.census.gov/table/ACSST5Y2022.S0101?q=UnitedStatesagegrouppercentage&g=010XX00US&moe=false Accessed on 14 February 2024.

[11] U.S. Census Bureau. 2022. Educational Attainment. U.S. Census Bureau. https://data.census.gov/table/ACSST1Y2022.S1501?q=education&moe=false Accessed on 14 February 2024.

[12] Matt Burgess. 2021. Google's cookie ban and FLoC, explained. https://www.wired.co.uk/article/google-cookies-floc Accessed on 14 February 2024.

[13] Alisa Chang and Pritish Kamath. 2021. Practical Differentially Private Clustering. https://blog.research.google/2021/10/practical-differentially-private.html Accessed on 14 February 2024.

[14] Google Chrome. 2024. More about ad topics. chrome://settings/adPrivacy/interests Accessed on 14 February 2024.

[15] Thomas Claburn. 2023. Shot down: Google's grand fancy plan for pro-privacy targeted ads. https://www.theregister.com/2023/01/18/google_topics_api/ Accessed on 14 February 2024.

[16] Kovila PL Coopamootoo. 2020. Usage patterns of privacy-enhancing technologies. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*. 1371–1390.

[17] National Research Council et al. 2002. *Technically speaking: Why all Americans need to know more about technology*. National Academies Press.

[18] Rachel Cummings, Gabriel Kaptchuk, and Elissa M Redmiles. 2021. "I need a better description": An Investigation Into User Expectations For Differential Privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. 3037–3052.

[19] Bennett Cyphers. 2021. Google's FLoC Is a Terrible Idea. https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea Accessed on 14 February 2024.

[20] Tobias Dehling and Ali Sunyaev. 2023. A design theory for transparency of information privacy practices. *Information Systems Research* (2023).

[21] Apple Differential Privacy Team. 2017. Learning with Privacy at Scale. https://machinelearning.apple.com/research/learning-with-privacy-at-scale Accessed on 14 February 2024.

[22] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing transparency design into practice. In *23rd international conference on intelligent user interfaces*. 211–223.

[23] David Eliot and David Murakami Wood. 2022. Culling the FLoC: Market forces, regulatory regimes and Google's (mis) steps on the path away from targeted advertising 1. *Information Polity* 27, 2 (2022), 259–274.

[24] Houda Elmimouni, Erica Racine, Patrick Skeba, Eric PS Baumer, and Andrea Forte. 2010. Does privacy still matter in the era of Web 2.0? A qualitative study of user behavior towards online social networking activities. In *Proceedings of Pacific Asia Conference on Information Systems (PACIS 2010)*. 591–602.

[25] Houda Elmimouni, Erica Racine, Patrick Skeba, Eric PS Baumer, and Andrea Forte. 2020. What are PETs for Privacy Experts and Non-experts. In *USENIX Symposium on Usable Privacy and Security (SOUPS)*.

[26] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. 1054–1067.

[27] Simone Fischer-Hbner and Stefan Berthold. 2017. Privacy-enhancing technologies. In *Computer and information security Handbook*. Elsevier, 759–778.

[28] Daniel Franzen, Saskia Nuñez von Voigt, Peter Sörries, Florian Tschorsch, and Claudia Müller-Birn. 2022. " Am I Private and If So, how Many?"–Using Risk Communication Formats for Making Differential Privacy Understandable. *arXiv preprint arXiv:2204.04061* (2022).

[29] Kevin Gallagher, Sameer Patil, and Nasir Memon. 2017. New Me: Understanding Expert and Non-Expert Perceptions and Usage of the Tor Anonymity Network. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. 385–398.

[30] 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. (2020).

[31] Vinay Goel. 2022. Get to know the new Topics API for Privacy Sandbox. https://blog.google/products/chrome/get-know-new-topics-api-privacy-sandbox/ Accessed on 14 February 2024.

[32] Google. 2023. Distributed differential privacy for federated learning. Google Research. https://research.google/blog/distributed-differential-privacy-for-federated-learning/ Accessed on 14 February 2024.

[33] Miguel Guevara. 2022. Expanding access to Differential Privacy to create a safer online ecosystem. https://developers.googleblog.com/2022/01/expanding-access-to-differential-privacy.html Accessed on 14 February 2024.

[34] Nadine Guhr, Oliver Werth, Philip Peter Hermann Blacha, and Michael H Breitner. 2020. Privacy concerns in the smart home context. *SN Applied Sciences* 2 (2020), 1–12.

[35] Yeting Guo, Fang Liu, Tongqing Zhou, Zhiping Cai, and Nong Xiao. 2023. Seeing is believing: Towards interactive visual exploration of data privacy in federated learning. *Information Processing & Management* 60, 2 (2023), 103162.

[36] Vicki Ha, Kori Inkpen, Farah Al Shaar, and Lina Hdeib. 2006. An examination of user perception and misconception of internet cookies. In *CHI'06 extended abstracts on Human factors in computing systems*. 833–838.

[37] Hana Habib, Yixin Zou, Yaxing Yao, Alessandro Acquisti, Lorrie Cranor, Joel Reidenberg, Norman Sadeh, and Florian Schaub. 2021. Toggles, Dollar Signs, and Triangles: How to (In)Effectively Convey Privacy Choices with Icons and Link Texts. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–25. https://doi.org/10.1145/3411764.3445387

[38] Karen Hao. 2019. How Apple personalizes Siri without hoovering up your data. MIT Technology Review. https://www.technologyreview.com/2019/12/11/131629/apple-ai-personalizes-siri-federated-learning/ Accessed on 14 February 2024.

[39] David Harborth, Sebastian Pape, and Kai Rannenberg. 2020. Explaining the Technology Use Behavior of Privacy-Enhancing Technologies: The Case of Tor and JonDonym. *Proceedings on Privacy Enhancing Technologies* 2020, 2 (2020), 111–128.

[40] Andrew Hard, Kanishka Rao, Rajiv Mathews, Swaroop Ramaswamy, Françoise Beaufays, Sean Augenstein, Hubert Eichner, Chloé Kiddon, and Daniel Ramage. 2018. Federated learning for mobile keyboard prediction. *arXiv preprint arXiv:1811.03604* (2018).

[41] Johannes Heurix, Peter Zimmermann, Thomas Neubauer, and Stefan Fenz. 2015. A taxonomy for privacy enhancing technologies. *Computers & Security* 53 (2015), 1–17.

[42] IBM. 2021. IBM Launches New Watson Capabilities to Help Businesses Build Trustworthy AI. https://newsroom.ibm.com/2021-04-21-IBM-Launches-New-Watson-Capabilities-to-Help-Businesses-Build-Trustworthy-AI Accessed on 14 February 2024.

[43] Nesrine Kaaniche, Maryline Laurent, and Sana Belguith. 2020. Privacy enhancing technologies for solving the privacy-personalization paradox: Taxonomy and survey. *Journal of Network and Computer Applications* 171 (2020), 102807.

[44] Farzaneh Karegar, Ala Sarah Alaqra, and Simone Fischer-Hübner. 2022. Exploring {User-Suitable} Metaphors for Differentially Private Data Analyses. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. 175–193.

[45] Farzaneh Karegar and Simone Fischer-Hübner. 2021. Vision: A noisy picture or a picker wheel to spin? exploring suitable metaphors for differentially private data analyses. In *Proceedings of the 2021 European Symposium on Usable Security*. 29–35.

[46] Jana Korunovska, Bernadette Kamleitner, and Sarah Spiekermann. 2020. The challenges and impact of privacy policy comprehension. *arXiv preprint arXiv:2005.08967* (2020).

[47] Patrick Kühtreiber, Viktoriya Pak, and Delphine Reinhardt. 2022. Replication: the effect of differential privacy communication on german users' comprehension and data sharing attitudes. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. 117–134.

[48] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research methods in human-computer interaction*. Morgan Kaufmann.

[49] Brendan McMahan and Daniel Ramage. 2017. Federated Learning: Collaborative Machine Learning without Centralized Training Data. https://blog.research.google/2017/04/federated-learning-collaborative.html Accessed on 14 February 2024.

[50] Eric Miraglia. 2019. Privacy that works for everyone. https://blog.google/technology/safety-security/privacy-everyone-io/ Accessed on 14 February 2024.

[51] Priyanka Nanayakkara, Johes Bater, Xi He, Jessica Hullman, and Jennie Rogers. 2022. Visualizing Privacy-Utility Trade-Offs in Differentially Private Data Releases. http://arxiv.org/abs/2201.05964 arXiv:2201.05964 [cs].

[52] Priyanka Nanayakkara, Johes Bater, Xi He, Jessica Hullman, and Jennie Rogers. 2022. Visualizing privacy-utility trade-offs in differentially private data releases. *arXiv preprint arXiv:2201.05964* (2022).

[53] Priyanka Nanayakkara, Mary Anne Smart, Rachel Cummings, Gabriel Kaptchuk, and Elissa M Redmiles. 2023. What are the chances? explaining the epsilon parameter in differential privacy. In *32nd USENIX Security Symposium (USENIX Security 23)*. 1613–1630.

[54] Irene Pollach. 2005. A typology of communicative strategies in online privacy policies: Ethics, power and informed consent. *Journal of Business Ethics* 62 (2005), 221–235.

[55] Ismini Psychoula, Deepika Singh, Liming Chen, Feng Chen, Andreas Holzinger, and Huansheng Ning. 2018. Users' privacy concerns in IoT based applications. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 1887–1894.

[56] Swaroop Ramaswamy, Rajiv Mathews, Kanishka Rao, and Françoise Beaufays. 2019. Federated learning for emoji prediction in a mobile keyboard.

[57] The Privacy Sandbox. 2022. Federated Learning of Cohorts (FLoC). https://privacysandbox.com/proposals/floc/ Accessed on 14 February 2024.

[58] The Privacy Sandbox. 2022. Topics API overview. https://developers.google.com/privacy-sandbox/relevance/topics Accessed on 14 February 2024.

[59] The Privacy Sandbox. 2023. Topics API: Relevant Ads Without Cookies. https://privacysandbox.com/proposals/topics/ Accessed on 14 February 2024.

[60] Florian Schaub, Rebecca Balebako, and Lorrie Faith Cranor. 2017. Designing effective privacy notices and controls. *IEEE Internet Computing* 21, 3 (2017), 70–77.

[61] Florian Schaub, Rebecca Balebako, Adam L Durity, and Lorrie Faith Cranor. 2015. A design space for effective privacy notices. In *Eleventh symposium on usable privacy and security (SOUPS 2015)*. SOUPS, 1–17.

[62] Mary Anne Smart, Priyanka Nanayakkara, Rachel Cummings, Gabriel Kaptchuk, and Elissa Redmiles. 2024. Models matter: Setting accurate privacy expectations for local and central differential privacy. *arXiv preprint arXiv:2408.08475* (2024).

[63] Mary Anne Smart, Dhruv Sood, and Kristen Vaccaro. 2022. Understanding risks of privacy theater with differential privacy. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–24.

[64] Kacper Sokol and Peter Flach. 2020. Explainability fact sheets: A framework for systematic assessment of explainable approaches. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 56–67.

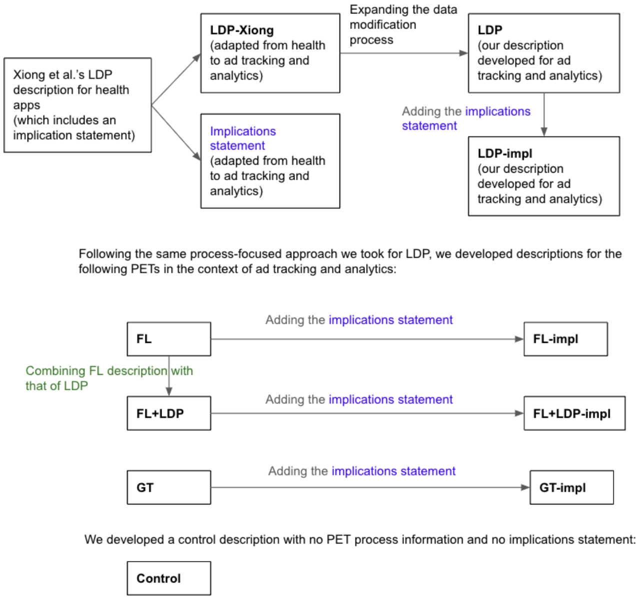
[65] Peter Story, Daniel Smullen, Yaxing Yao, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2021. Awareness, adoption, and misconceptions of web privacy tools. *Proceedings on Privacy Enhancing Technologies* (2021).

[66] Evan T Straub. 2009. Understanding technology adoption: Theory and future directions for informal learning. *Review of educational research* 79, 2 (2009), 625–649.

[67] Jenny Tang, Hannah Shoemaker, Ada Lerner, and Eleanor Birrell. 2021. Defining privacy: How users interpret technical terms in privacy policies. *Proceedings on Privacy Enhancing Technologies* 2021, 3 (2021).

[68] ADP Team et al. 2017. Learning with privacy at scale. *Apple Mach. Learn. J* 1, 8 (2017), 1–25.

[69] Sabine Trepte, Doris Teutsch, Philipp K Masur, Carolin Eicher, Mona Fischer, Alisa Hennhöfer, and Fabienne Lind. 2015. Do people know about privacy and data protection strategies? Towards the "Online Privacy Literacy Scale"(OPLIS). *Reforming European data protection law* (2015), 333–365.



Xiong et al.'s LDP description for health apps (which includes an implication statement)
LDP-Xiong (adapted from health to ad tracking and analytics)
Implications statement (adapted from health to ad tracking and analytics)
Expanding the data modification process
LDP (our description developed for ad tracking and analytics)
Adding the implications statement
LDP-impl (our description developed for ad tracking and analytics)
Following the same process-focused approach we took for LDP, we developed descriptions for the following PETs in the context of ad tracking and analytics:
FL
Adding the implications statement
FL-impl
Combining FL description with that of LDP
FL+LDP
Adding the implications statement
FL+LDP-impl
GT
Adding the implications statement
GT-impl
We developed a control description with no PET process information and no implications statement:
Control

Figure 3: Steps we took to develop the process- and implications-focused PET descriptions in our study, as described in more detail in Section 3.1. The PET descriptions we used in our survey experiment are represented by bold text in text boxes, each of which corresponds to an experimental condition in our survey. The exact PET descriptions used in our survey experiment are presented in Table 1.

[70] Kang Wei, Jun Li, Ming Ding, Chuan Ma, Howard H Yang, Farhad Farokhi, Shi Jin, Tony QS Quek, and H Vincent Poor. 2020. Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security* 15 (2020), 3454–3469.

[71] Claire Woodcock, Brent Mittelstadt, Dan Busbridge, and Grant Blank. 2021. The impact of explanations on layperson trust in artificial intelligence–driven symptom checker apps: experimental study. *Journal of medical Internet research* 23, 11 (2021), e29386.

[72] Aiping Xiong, Tianhao Wang, Ninghui Li, and Somesh Jha. 2020. Towards effective differential privacy communication for users' data sharing decision and comprehension. In *2020 IEEE Symposium on Security and Privacy (SP)*. 392–410. https://doi.org/10.1109/SP40000.2020.00088 ISSN: 2375-1207.

[73] Aiping Xiong, Chuhao Wu, Tianhao Wang, Robert W Proctor, Jeremiah Blocki, Ninghui Li, and Somesh Jha. 2022. Using Illustrations to Communicate Differential Privacy Trust Models: An Investigation of Users' Comprehension, Perception, and Data Sharing Decision. In *arXiv preprint arXiv:2202.10014*.

A PET descriptions

A.1 Steps for developing PET descriptions

We present how we developed the textual descriptions of LDP, FL, FL+LDP, and GT in the context of ad tracking and analytics in Figure 3. These nine PET descriptions, and a control description that does not explain how privacy is protected or privacy implications, correspond to the ten experimental conditions of our survey experiment.

A.2 LDP descriptions

For comparison purposes, we show Xiong et al.'s [72] original LDP description that includes an implications statement, the version of this description that we adapted to the advertising and analytics context (LDP-Xiong) and our LDP descriptions with and without an implications statement (LDP and LDP-impl) in Table 4.

A.3 In-the-wild descriptions of other PETs

Table 5 provides examples of in-the-wild descriptions of FL, FL+LDP, and GT provided by major tech companies. The sources include the companies' research webpages, announcements, and product settings. Among these, the description of Topics was incorporated by Google in the Ad Topics settings of Chrome [14] after we designed the survey and collected our data.

B Survey Instrument

We use the *control* condition as an example to list the survey instructions and questions. We provide additional information and mark the variations in survey questions shown to respondents assigned to other conditions in italicized texts.

Survey instruction.

In this survey, we will ask you a series of questions about a hypothetical scenario. Please do your best to imagine yourself in this scenario and answer the questions.

Scenario description.

Imagine that you came across the following description of a social platform.

The platform makes revenue by showing users personalized ads via inferring users' interests from their online activity tracked on the platform and other businesses' websites/apps.

To protect your information, the organization stores all of your behavioral data for targeting ads (e.g., your interaction with the platform and with other apps/websites) securely on their servers.

Please see Table 1 for the PET description under other experiment conditions.

Imagine you are trying to decide whether you would like to use this platform.

Confidence in platform use question.

- How confident are you about deciding whether to use this platform?
  - ○ Very confident
  - ○ Confident
  - ○ Moderately confident
  - ○ Slightly confident
  - ○ Not at all confident

Perceived protection of user data.

- How would you explain to other people how the platform protects users' data? Please write at least two clear sentences. [Space for open-ended responses was provided.]

Objective comprehension questions.

For each of the following statements, please indicate if you expect the following to be true or false if you would use the platform described above.

- An employee working for the platform, such as a data analyst, could be able to see my exact behavioral data (e.g., my interaction with the platform and with other apps/websites).
  - ○ True
  - ○ False
  - ○ I don't know
- A criminal or foreign government that hacks the platform could learn my behavioral data (e.g., my interaction with the platform and with other apps/websites).
  - ○ True
  - ○ False
  - ○ I don't know
- A law enforcement organization could access my behavioral data (e.g., my interaction with the platform and with other apps/websites) with a court order requesting this data from the company.
  - ○ True
  - ○ False
  - ○ I don't know
- Graphs or informational charts created using information given to the platform could reveal my behavioral data (e.g., my interaction with the platform and with other apps/websites).
  - ○ True
  - ○ False
  - ○ I don't know
- Data that the platform shares with its partner organizations could reveal my behavioral data (e.g., my interaction with the platform and with other apps/websites).
  - ○ True
  - ○ False
  - ○ I don't know

Interpretation of PET description segments.

- In your own words, describe what "behavioral data (e.g., your interaction with the platform and with other apps/websites)" in the above description means. Please write at least two clear sentences.
- In your own words, describe what "To protect your information, the organization stores all of your behavioral data for targeting ads (e.g., your interaction with the platform and with other apps/websites) securely on their servers" in the above description means. Please write at least two clear sentences.

  This question is different under different experiment conditions; this question is specific to the PET description provided in the scenario description section of the survey (see Table 6).

Subjective comprehension questions.

- How confident are you in your understanding of the privacy technology used by the platform's company?
  - ○ Very confident
  - ○ Confident
  - ○ Moderately confident
  - ○ Slightly confident
  - ○ Not at all confident
- You indicated that the description of privacy technology used by the platform was not easy to understand. Please indicate which words or sentences were hard to understand, or you wished you had more details about. [This question was asked to respondents who gave a rating below the "confident" threshold.]

Prior PET familiarity

- Have you ever heard of the following technologies? (select all that apply)
  - ○ Differential privacy

Table 4: We present three types of LDP descriptions relevant to this paper: (1) LDP Imp. is the original LDP description from Table XII in Xiong et al. [72]. We have modified this to create LDP-Xiong, LDP, and LDP-impl. (2) LDP-Xiong is our adaptation of Xiong et al.'s original LDP description tailored for the context of ad tracking and analytics. Modifications to Xiong et al.'s original LDP description in LDP-Xiong are highlighted in the italicized text but were not italicized in the actual survey shown to participants. (3) LDP and LDP-impl are our refined versions, distinguished by the inclusion or exclusion of an implications statement.

PETs	Description
LDP Imp. in Xiong et al. [72]	To respect your personal information privacy and ensure best user experience, the data shared with the app will be processed via the local differential privacy (LDP) technique. That is, the app will randomly modify your data on your cellphone before sending it to the app server. Since the app server stores only the modified version of your personal information, your privacy is protected even if the app server's database is compromised.
LDP-Xiong	To respect your personal information privacy and ensure best user experience, *the behavioral data (e.g., your interaction with the platform and with other apps/websites)* shared with *the company* will be processed via *an additional privacy technique.* That is, *your behavioral data* will be randomly modified before it is sent to *the company.* Since *the company* stores only the modified version of your personal information, your privacy is protected even if *the company's* is compromised.
LDP	To protect your information, the organization adds noise to your behavioral data (e.g., your interaction with the platform and with other apps/websites) before being sent to the organization for targeting ads. This means that your data is randomly modified, so that some of your actual data is used whereas some of it is random and not representative of your behavior. Your exact behavioral data is never sent to the organization, instead a subset of your noisy data is randomly selected and sent. The organization can still infer patterns from the noisy data across a large number of users.
LDP-impl	To protect your information, the organization adds noise to your behavioral data (e.g., your interaction with the platform and with other apps/websites) before being sent to the organization for targeting ads. This means that your data is randomly modified, so that some of your actual data is used whereas some of it is random and not representative of your behavior. Your exact behavioral data is never sent to the organization, instead a subset of your noisy data is randomly selected and sent. The organization can still infer patterns from the noisy data across a large number of users. This way, the organization still learns aggregated interests across users but not your exact behavior, which protects your privacy against the organization's employees or if the organization's database is compromised.

Table 5: Examples of industry descriptions of FL, FL+LDP, and GT. The last description of Topics [14] (last row) in the table was incorporated in the Ad Topics settings of Chrome by Google after our data collection was completed. We list the description of Topics from Chrome Version 122.0.6261.69 [14] in the table.

PET	Description
FL [49]	Federated Learning enables mobile phones to collaboratively learn a shared prediction model while keeping all the training data on device, decoupling the ability to do machine learning from the need to store the data in the cloud. This goes beyond the use of local models that make predictions on mobile devices (like the Mobile Vision API and On-Device Smart Reply) by bringing model training to the device as well. It works like this: your device downloads the current model, improves it by learning from data on your phone, and then summarizes the changes as a small focused update. Only this update to the model is sent to the cloud, using encrypted communication, where it is immediately averaged with other user updates to improve the shared model. All the training data remains on your device, and no individual updates are stored in the cloud.
FL [21]	Differential privacy provides a mathematically rigorous definition of privacy and is one of the strongest guarantees of privacy available. It is rooted in the idea that carefully calibrated noise can mask a user's data. When many people submit data, the noise that has been added averages out and meaningful information emerges.
FL+LDP [4]	Federated learning with differential privacy (FL-DP) is one of the latest privacy-enhancing technologies being evaluated at Meta as we constantly work to enhance user privacy and further safeguard users' data in the products we design, build, and maintain. FL-DP enhances privacy in two important ways: 1. It allows machine learning (ML) models to be trained in a distributed way so that users' data remains on their mobile devices. 2. It adds noise to reduce the risk of an ML model memorizing user data.
GT [31]	With Topics, your browser determines a handful of topics, like "Fitness" or "Travel & Transportation," that represent your top interests for that week based on your browsing history. Topics are kept for only three weeks and old topics are deleted. Topics are selected entirely on your device without involving any external servers, including Google servers. When you visit a participating site, Topics picks just three topics, one topic from each of the past three weeks, to share with the site and its advertising partners. Topics enables browsers to give you meaningful transparency and control over this data, and in Chrome, we're building user controls that let you see the topics, remove any you don't like or disable the feature completely.
GT [14]	Chrome notes topics of interest based on your browsing history from the last few weeks. Later, a site you visit can ask Chrome for your topics to personalize the ads you see. Chrome shares up to 3 topics while protecting your browsing history and identity. Chrome auto-deletes topics that are older than 4 weeks. As you keep browsing, a topic might reappear on the list. Or you can block topics you don't want Chrome to share with sites. Learn more about managing your ad privacy in Chrome.

- ○ End-to-end encryption
- ○ Secure multi-party computation
- ○ Deliquescent security
- ○ Federated learning
- ○ Topics
- ○ FLoC (Federated Learning of Cohorts)
- ○ None of the above

PET identification

- Which of these technologies do you think was described in the scenario?
  - ○ Differential privacy
  - ○ End-to-end encryption
  - ○ Secure multi-party computation
  - ○ Deliquescent security
  - ○ Federated learning
  - ○ Topics
  - ○ FLoC (Federated Learning of Cohorts)
  - ○ None of the above

Demographic questions

- In what year were you born? (four digits please)
- What is your gender?
  - ○ Man
  - ○ Woman
  - ○ Non-binary
  - ○ Prefer to self-describe
  - ○ Prefer not to answer
- Please specify your race/ethnicity (select all that apply)
  - ○ Hispanic, Latino, or Spanish

  - Black or African American
  - White
  - American Indian or Alaska Native
  - Asian, Native Hawaiian, or Pacific Islander
  - Prefer to self-describe
  - Prefer not to answer
- What is the highest level of school you have completed or the highest degree you have received?
  - Less than high school degree
  - High school graduate (high school diploma or equivalent including GED)
  - Some college but no degree
  - Associate's degree
  - Bachelor's degree
  - Advanced degree (e.g., Master's, doctorat)
  - Prefer not to answer
- Which of the following best describes your educational background or job field?
  - I DO NOT have an education in, nor do I work in, the field of computer science, computer engineering or IT.
  - I have an education in, nor do I work in, the field of computer science, computer engineering or IT.

C Codebook

We show the codebooks for data discussed in the paper below. For presentation purposes, we list the codebooks by data type and by PET.

C.1 Codebook for confusing phrasing within PET descriptions

LDP.

- How data is modified
  Example: *"I have an okay understanding of it, I think, but I would like more clarification about what the noise is and how that works and what makes the information randomly modified. "*
- How large is the subset
  Example: *"I wish there were more details about the subset of data that is sent. How much of your data is really being sent"*
- Interpretation of noise
  Example: *"I'd like to know more details about how the noisy data is added and if it is hard to distinguish from the real data."*
- How data modification protects privacy
  Example: *"I don't see how obscuring some parts of my data will actually hide any info being mined about me, since it doesn't seem like my privacy is being protected throughout the entire process."*
- Why modified data is still useful
  Example: *"how the data is useful for the companies if it's all going to be switched around and modified before being sent in."*

FL.

- How machine learning models work
  Example: *"It could be helpful to explain how exactly machine learning is used on the user's device to infer interests. Providing a brief overview of the algorithms or techniques involved might make this clearer."*
- How machine learning models protect privacy
  Example: *"I would like to better understand how it works for them to send my data to a machine learning platform without my data being accessible to any human."*
- Data shared with the organization
  Example: *"I would want some more information about how non-personal the final inferred interests are as it is not clearly stated."*
- How model merging works
  Example: *"The concept of merging user models to infer patterns could be elaborated upon. For instance, you might want to know more about how this merging process works and how it ensures individual user data privacy."*

FL+LDP.

- How machine learning models work
  Example: *"How does the ML do the inferring of the data? I feel like inferring is still very strongly linked to your data"*
- Interpretation of noise
  Example: *"Noise will be added. What type of noise, how obvious is it to not be fake data which could be discarded?"*
- Why modified data is still useful
  Example: *"I don't understand the value of my data when it's been modified. Wouldn't that make it useless?"*
- Why adding noise protects privacy
  Example: *"I am not sure how exactly adding noise in makes it better for me."*

GT.

- How the company infers interests without access to exact data
  Example: *"I don't understand how the company only uses my inferred likes and interests without accessing/sharing all the information."*
- How machine learning models work
  Example: *"What does "machine learning" mean?"*
- On users' device
  Example: *"I do not understand the phrase "only on your device." You can visit a social platform on more than one device."*
- Data shared with the organization
  Example: *"I don't see much difference in inferred interest and actual data. I'm not sure I believe it."*

C.2 Codebook for interpretation of PET description segments

LDP.

- Random modification of data
  Example: *"Aspects of your data will be jumbled up randomly before the company gets it. The company won't be able to see your data."*
- Sources of privacy protection
  Example: *"In order to protect the privacy of its users, the platform adds noise to the data, which is a modification of the data that prevents users from being completely tracked, which in turn preserves the privacy of users."*

Table 6: Open-ended questions about specific segments of the PET descriptions, which vary across experimental conditions.

Condition	Survey question
Control	In your own words, describe what "To protect your information, the organization stores all of your behavioral data for targeting ads (e.g., your interaction with the platform and with other apps/websites) securely on their servers" in the above description means. Please write at least two clear sentences.
LDP-Xiong	In your own words, describe what "... the behavioral data (e.g., your interaction with the platform and with other apps/websites) shared with the company will be processed via an additional privacy technique. That is, your behavioral data will be randomly modified before it is sent to the company" in the above description means. Please write at least two clear sentences.
LDP	(1) In your own words, describe what "To protect your information, the organization adds noise to your behavioral data (e.g., your interaction with the platform and with other apps/websites) before being sent to the organization for targeting ads. This means that your data is randomly modified, so that some of your actual data is used whereas some of it is random and not representative of your behavior" in the above description means. Please write at least two clear sentences. (2) In your own words, describe what "Your exact behavioral data is never sent to the organization, instead a subset of your noisy data is randomly selected and sent. The organization can still infer patterns from the noisy data across a large number of users" in the above description means. Please write at least two clear sentences.
FL	In your own words, describe what "Your exact behavioral data is never sent to the organization and only the model representing your inferred interests will be sent. Then, to infer patterns across a large number of users, your model is merged with other users' models" in the above description means. Please write at least two clear sentences.
FL+LDP	(1) In your own words, describe what "Noise will be added to your behavioral data so that it is randomly modified before being used for training a model representing your inferred interests. This means that, for training the model, some of your actual data is used whereas some of it is random and not representative of your behavior" in the above description means. Please write at least two clear sentences. (2) In your own words, describe what "Your exact behavioral data is never sent to the organization and only the model representing your inferred interests will be sent. Then, to infer patterns across a large number of users, your model is merged with other users' models" in the above description means. Please write at least two clear sentences.
GT	In your own words, describe what "... the technology records inferred topics from your behavioral data only on your device. Your exact behavioral data is never sent to the organization, instead from your top topics of the last week, a small number are randomly selected and sent; there is also a small chance a random topic will be selected instead of one of yours" Please write at least two clear sentences.

- Personal identifiable or sensitive information gets removed or modified
  Example: *"I mean maybe it puts you in groups with people with similar behavioral data but then it removes anything that would identify you exactly like your name or other identifying information that would tie it to your exact identity."*
- Unsure about how data is "modified"
  Example: *"I can't really explain this one, I don't know what "randomly modified" means so I certainly don't know how I'm supposed to feel safer"*
- Encryption of user data
  Example: *"This means that the data is essentially encrypted. The encryption will be random and then sent to the company."*
- Specific interpretation of data modification: noise
  Example: *"It means that exact interactions will be shielded from the organization. So perhaps types of pages you visit/interact with will be randomly changed to reflect broad categories as (the Bernie Sanders campaign page changed to "political campaign" for example)."*
- Specific interpretation of data modification: data is modified by being mixed with other users' data Example: *"I think that this means that the data will be slightly altered based on other people's behavioral data"*
- Users' data gets merged
  Example: *"I'm honestly not sure how that works, but I assume that the data from many people will be randomly mixed together."*
- Mention of the company sharing users' data with a third party
  Example: *"I assume when apps sell data, such as your interests to thirds parties, it is skewed in some way."*

FL.

- User's data gets merged
  Example: *"My data is mixed with the data of other users."*
- Vague notion of anonymization
  Example: *"The data is made anonymous before it is sent."*
- Vague notion of machine learning model outputs
  Example: *"The exact pages you visit maybe masked over by the machine data gathering sequence."*
- Machine learning models are trained on the combined data of users' and other data
  Example: *"Then it basically compiles all of its collected data into one model to create a broader model to cover multiple users."*
- Machine learning models are trained on the user's device
  Example: *"I believe that this description is saying that instead of showing your exact behavioral data to the company, a machine will condense/alter that information in a way that defines one's "inferred interests"."*
- Machine learning models are shared with the organization
  Example: *"This is then merged with other users' models to disguise your exact information even further, and will then be used by the website. "*
- AI, machine, or a machine learning model tracks users' online behaviors
  Example: *"I think instead it would follow person X and others*

and say what sites people of the same age as person X go to or something similar.”

FL w/ LDP.

- User’s data gets merged
  Example: *‘This data is then merged with others uses so it will be difficult to isolate your data.”*
- Machine learning models learn users’ interests from modified data
  Example: *“Noise is added to some of your data so it is randomly modified before being to training model. This will protect you from the organization’s employees or being compromised.”*
- Machine learning models are trained on the combined data of users’ and other data
  Example: *“Aggregated data is being used to train the ML model. Apparently this reduces the risk to an individuals privacy.”*
- Machine learning models are merged
  Example: *“Your model is integrated or mixed with models from many other users in order to comprehend more general user trends and preferences.”*
- Specific interpretation of data modification: noise
  Example: *“Noise I would assume would mean random clicking or random long pauses where the site almost times out.”*
- Merging data makes data unidentifiable
  Example: *“They’re trying to give me anonymity to the employees. They’re trying to merge data and add fake data so nothing specific can truly be tied to an individual.”*
- User data is shared with the company
  Example: *“This means that my exact data is never fully used, but only part of my data is sent, and the model has to infer the rest (make up what it THINKS I would like from the data it does have).”*
- The mix of users’ data and other data is shared with the company
  Example: *“The organization never gets data that is purely yours, but is mixed with other random data.”*

GT.

- User data is processed on device and not shared with the organization
  Example: *“The explanation means that the device the person is using is going to be the only device that holds in their behavioral data so it never goes to the company itself.”*
- User interests are inferred
  Example: *“The platform takes small samples of data that come from your top topics in the previous week or from a completely random topic to use for what you may be interested in.”*
- Randomness in topics shared with the organization
  Example: *“Sometimes even a random topic is sent to protect my identity.”*
- AI/tech tracks users’ online behaviors
  Example: *“The technology tracks how you are interacting with different posts to create topics that you might like.”*
- Users’ data is shared with the company
  Example: *“It means that my actual patterns are never sent in full. Only a select set is sent to the parent organization, along with some random data to hide it.”*

Table 7: Mann-Whitney test statistics for RQ1-3, with effect sizes in parentheses. *p<0.05; **p<0.01; ***p<0.005.

	Confidence in platform use	**Objective comprehension**	**Subjective confidence**
RQ1	499.5 (0.53)	438.5 (0.47)	651.0 (0.70)***
RQ2	483.5 (0.52)	453.0 (0.49)	506.5 (0.54)
RQ3:FL	388.5 (0.42)	358.5 (0.39)	480.5 (0.52)
RQ3:FL+LDP	373.5 (0.40)	398.5 (0.43)	450.5 (0.48)
RQ3:GT	430.0 (0.54)	500.5 (0.46)	368.0 (0.40)

D Test results

We list the test results for RQ1-3 with effect sizes in parentheses in Table 7.