“Because I didn’t touch these and even don’t know why I should to change these”: Why App Developers Do (Not) Update Apple’s Privacy Labels

Arwa Alsahdi
The George Washington University
aalsahdi@gwu.edu

Jialiang Yan
The George Washington University
jy843@gwmail.gwu.edu

Monica Kodwani
The George Washington University
monicakodwani@gwmail.gwu.edu

Matthias Fassl
The George Washington University
matthias.fassl@gwu.edu

Chris Kanich
University of Illinois Chicago
ckanich@uic.edu

Adam J. Aviv
The George Washington University
aaviv@gwu.edu

Abstract

Apple introduced app-based privacy labels in 2020 to improve apps’ communication of their data practices. However, most developers appear to treat privacy labels as a “set-once” mechanism. To better understand the dynamics of this system, we first analyzed a four-year longitudinal dataset of Apple’s Privacy Label. Next, we conducted an email survey of developers who have changed (or not changed) their privacy labels during this period, and finally, performed follow-up interviews with developers from each group. We find that only 51,364 apps (less than 6%) over this period have made any changes to their privacy labels, many of them changing their initial *Do Not Collect* label to more refined classification. From the emails and interviews, the “black box” of third-party data practice may lead developers to underreport their app’s data practices. Many developers reported that privacy labels are a valuable marketing tool for promoting their apps as privacy-friendly. Privacy labels may appear stable not necessarily because practices are stable, but because ambiguity encourages minimal or optimistic disclosure. To improve privacy label maintenance, we recommend enhanced transparency mechanisms for third-party libraries, stronger workflow integration, and platform support that guides developers and strengthens users’ control.

Keywords

app store, privacy labels, privacy policies, developers, longitudinal, email survey, interviews

1 Introduction

The web’s reliance on advertising as a business model resulted in pervasive data collection in mobile app ecosystems [12, 38, 40, 41]. As a result, informing users about data practices has become critical for digital privacy. For decades, notice and choice has served as a key approach to maintaining privacy, primarily through the use of privacy policies [15]. However, these legal documents are often lengthy and difficult for users to comprehend [16, 36]. To address these limitations, Kelley et al. proposed “privacy nutrition labels” as a standardized and concise alternative to facilitate more effective privacy communication [26–28]. Informed by this, in December 2020, Apple introduced “Apple Privacy Labels” (APL), which mandate developers to disclose their apps’ data practices within a standardized “App Privacy” section in the App Store (Figure 1) [3]. Following Apple’s new practice, Google introduced a similar section called “Data Safety Section” (DSS) for Google Play Store in early 2022 [21].

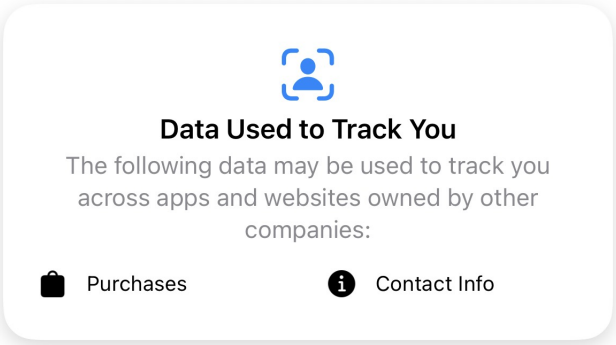

Figure 1: An example of Apple privacy labels (APL)

The effectiveness of privacy communication highly depends on the accuracy [34]. However, prior research indicates that APL may not always reflect the app’s data practices [2, 33, 44]. A longitudinal analysis from 2021 to 2022 showed that most developers did not update their privacy labels over a two-year period, treating APL as a “set-once” mechanism [11]. One contributing factor may be that developers struggle to interpret Apple’s privacy terminology and classification scheme [34]. Hence, the process of creating privacy labels and keeping them up to date poses challenges, restricting privacy labels’ usefulness.

In this work, we investigate the privacy label creation and maintenance as a dynamic process in the iOS ecosystem. We pair behavioral evidence from real-world privacy label updates with qualitative self-reports from developers, allowing us to examine both what changes occur and how developers interpret and manage the maintenance process. A deeper understanding of this process allows us to propose evidence-based interventions aimed at improving the

Proceedings on Privacy Enhancing Technologies YYYY(X), 1–21

https://doi.org/XXXXXXX.XXXXXXX

accuracy of privacy labels. The following three research questions guide our research approach:

- RQ1: How many apps have changed or not changed their privacy labels and in what form?
- RQ2: What are the driving factors behind the changes or non-changes of privacy labels?
- RQ3: What are developers' views of maintaining privacy label accuracy?

We used a three-part mixed-methods approach to answer these research questions. First, we conducted a longitudinal analysis that extends Balash et al. [11] by analyzing APL changes of 926,240 unique apps from 2021 to 2024. Using this dataset, we investigated the prevalence of privacy label changes, how these labels change, which apps change their label, and if changes improve alignment with their privacy policies. Second, we deployed email surveys to 7,000 developers identified through the longitudinal analysis to identify the driving factors behind (not) changing privacy labels. Finally, we conducted semi-structured interviews with 19 iOS developers to gain an in-depth understanding of their privacy reporting workflow, perceptions of privacy labels, and the barriers to maintaining labels accurately.

Over the four-year period, only a small fraction update their privacy labels. When changes occur, they predominantly disclose expanded data practices. 41.7% of observed changes newly added *Data Used to Track You*. Apps offering in-app purchases were associated with a 46.5% higher expected number of label changes. Despite these revisions, changes did not substantially improve alignment between privacy labels and underlying privacy policies. Survey responses suggest that privacy label changes are primarily driven by functionality changes and monetization strategies, particularly the addition or removal of advertising, analytics, or third-party SDKs. Developers wanted to comply with Apple's requirements and improve transparency for users. The interview highlighted how organizational structure shapes the maintenance of accurate privacy labels: In smaller teams, developers interpret data practices and update labels alone, whereas in larger organizations management, legal, and product roles take over. Limited visibility into third-party data practices may lead to unintentional under-reporting.

Inaccuracies or delays in privacy label updates appear to be less a matter of developer intent and more a result of uncertainty, organizational fragmentation, and limited tool support. Improving the usefulness of privacy labels will requires better integration into development workflows, clearer guidance around third-party data practices, and tools to help developers keep privacy labels in sync as apps evolve.

2 Related Work and Background

Privacy Nutrition Labels. Prior research has shown privacy policies to be lengthy and difficult to understand [15, 25, 36]. Privacy nutrition labels were suggested to provide a standardized format to facilitate more effective communication of privacy policies [26, 27]. Kelley et al. found that clearly presented labels on mobile phones can help users make more informed privacy decisions [28].

Apple Privacy Labels (APL). Apple introduced privacy labels to its app store in 2020 to provide users with a concise summary of app privacy practices, representing an industrial-scale implementation of the "privacy nutrition label" concept. The labels, as shown in Figure 1, are presented in the "App Privacy" section for every app in the store. The labels categorize data practices into three groups based on their privacy implications: (1) *Data Used to Track You*, which refers to data used by the app or third-party advertisers to track users' activities across different apps and websites over time; (2) *Data linked to you*, containing any information tied to the user's identity and explicitly including any personal information as defined under relevant privacy regulations; (3) *Data not linked to you*, include data collected but not linked to a user's identity [7]. In the default compact view, these practices are displayed as high-level blocks containing specific types of collected data. Users can further access the purposes for each data type in the expanded label view.

Privacy Label Measurement. Following the introduction of APL and Google's own version of them, called Data Safety Sections (DSS), these labels have gained significant attention from the academic community and have been the subject of large-scale measurement studies [2, 9, 11, 24, 29, 31, 33, 34, 42–44]. Specifically, Li et al. conducted the first large-scale measurement study of APL. By collecting all iOS apps' privacy labels in the app store, they found that only 2.7% apps voluntarily adopted privacy labels in 2021 [34]. Using similar collection methods, Scoccia et al. found that free apps tend to collect more sensitive data than paid ones [43]. Subsequent studies have identified significant discrepancies between self-reported labels and data practices by comparing traffic behaviors [33], binary code [44], and privacy policy text [2, 24]. Similarly, the discrepancies between reported practices and actual behavior have also been found in DSS [9, 31]. Since both Apple and Google have introduced privacy labels in their app stores, previous research has also compared labels of the same apps across platforms, finding discrepancies and inconsistencies in self-reported data practices across platforms [29, 42]. Finally, to characterize the APL ecosystem, Balash et al. collected weekly snapshots of over 1.6 million apps over a two-year period from 2021 to 2022. Their findings suggested that privacy labels are often treated as a "set-once" task with few updates [11]. Building upon this work, our study extends the analysis to a four-year period (2021-2025). By analyzing these longitudinal data and contacting developers identified within the dataset, this research aims to provide a more comprehensive and nuanced understanding of the factors driving or hindering privacy label updates.

Privacy Label Perceptions. Beyond measurement-based analyses, several user studies have investigated how users interact with and perceive privacy labels [10, 20, 35, 46, 47]. Zhang et al. conducted an interview study with APL users, identifying issues of APL such as confusing structures and unfamiliar terms [46]. Similarly, Lin et al. examined how the design differences between APL and DSS impacted user comprehension and perceptions through an interview study with both Android and iOS users. Their findings indicated that DSS also suffered a lot of issues identified from APL before [35]. Furthermore, Zhang et al. analyzed whether privacy labels (both DSS and APL) could answer privacy-related questions from users, revealing that over half of the questions were not addressed [47]. In addition, research has explored the impact of privacy labels on user behavior. Balash et al. conducted a survey investigating how APL affected user perception of app behavior, finding that many

privacy label attributes could raise participants' risk perception and lower their willingness to install an app [10]. Most recently, Frik et al. investigated the influence of telehealth app's DSS on users' privacy expectations and behavioral intensions, revealing that DSS altered users' privacy expectations, often inaccurately [20]. Distinct from this body of user study research, our study investigates the developer perspective on how do they maintain and update labels in the context of the larger iOS ecosystem dynamics.

Developers' Perspectives on Privacy Labels. Prior work has also investigated developers' perspectives on privacy label implementation, specifically for APL [34] and DSS [30]. Li et al. observed APL creation process from 12 iOS developers and interviewed them [34]. They found that developers were highly involved in creating APL but mostly made errors such as missing third-party data use. Additionally, they reported challenges developers faced, including known factors (e.g., lack of resources), unknown factors (e.g., preconceptions), and label complexities. Compared to their work, which focused on the APL creation process, this study aims to explore the factors behind the evolution of APL. Khandelwal et al. [30] performed a longitudinal study to analyze the evolution of DSS, complemented by an email survey to identify the factors of label changes. They found that developers faced challenges such as frequently changing DSS rules, unclear guidelines and review processes by Google, label complexities, and limited transparency regarding third-party data practices. While prior work has explored the DSS ecosystem, our study extends this work by focusing on APL. Furthermore, we go beyond survey-based investigations by conducting semi-structured interviews to elicit in-depth insights.

3 Methodology

We use a three-part approach to understanding why developers do (not) update Apple's privacy labels: (1) We analyze privacy label changes of the 926,240 unique apps in a 2021–2025 dataset of app–store metadata, (2) we email developers to get a broad overview of what drives changes to privacy labels, and (3) we conduct in-depth interviews to understand barriers to maintaining accurate privacy labels. We contacted 5,000 developers who changed privacy labels in the last two years and 2,000 who did not. From the 128 developers who responded, we recruited 19 for in-depth interviews. Figure 2 shows an overview of the three study parts and their participants.

3.1 Privacy Label Changes

Prior work [11] suggested that developers treat privacy labels as a "set once" mechanism that they do not change as the privacy policy changes. To understand the prevalence of privacy label changes across apps and the common characteristics of apps with such changes, we analyze a four-year longitudinal dataset of App Store privacy labels

Dataset. This work builds upon our prior longitudinal study [11], which conducted weekly measurements from July 15, 2021, to October 25, 2022. Since then, we have continued the same scraping methodology and made all data available in our observatory.[1] We started our recruitment with a dataset snapshot from October 28, 2024. This snapshot includes a total of 926,240 unique applications,

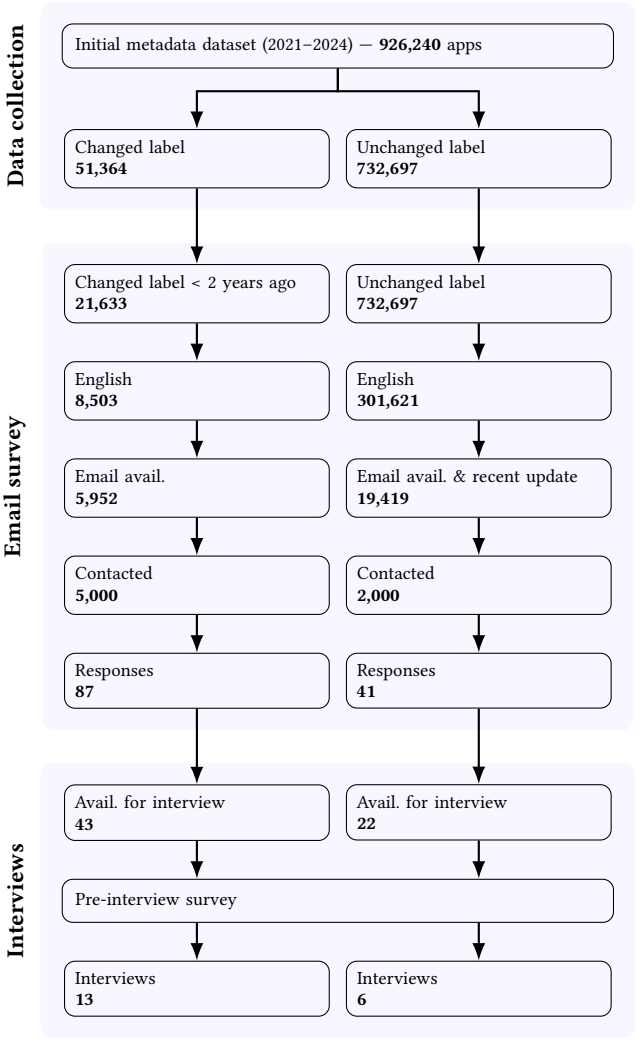

Figure 2: Study overview with separate flows for changed vs. unchanged privacy labels.

corresponding to the full set of applications available through Apple's public sitemap at the time of retrieval, with no app filtering applied. We report on changes through December, 2025 for the overview in Figure 5. We excluded apps no longer available in the app store, since their privacy policy and email contacts were inaccessible. For each application, we identified its initial privacy label (if any), corresponding to the first privacy label appearing in our dataset. We searched all future snapshots to identify updates. For each update, we recorded the time and exact changes.[2]

Analysis. We explore the dataset in four ways to understand (1) if privacy labels change, (2) how they change, (3) which apps change their privacy labels, and (4) if the changes improve alignment with the app's privacy policies. To get an overview about privacy label

[1] https://privlabels.gwusec.seas.gwu.edu/

[2] The dataset of applications with and without privacy label changes is publicly available on OSF: https://doi.org/10.17605/OSF.IO/BUXYJ

changes, we used descriptive statistics and visualizations. Our analysis of the dynamics of changes is based on the transition matrix (common in state-based models) for privacy type changes, and a longitudinal heatmap for changes to the data categories. To understand which app characteristics predict changes in privacy labels over time, we used negative binomial regression models. The dependent variable was the number of privacy label changes observed for each app. Negative binomial regression (NB2) was used to account for overdispersion in the count outcome. Models include controls for app age (in years since release), price (log-transformed), the presence of in-app purchases, user engagement indicators (rating count and average rating), and app categories. Results are reported as incidence rate ratios (IRRs) with Wald 95% confidence intervals. To understand if privacy label changes improve alignment with the apps' privacy policies, we use a BERT model from prior work [2] to suggest an appropriate privacy label, and then calculate the hamming distance to the app's privacy label before and after the change. We used a paired sign-test to make sure that the measured directional shift is not due to random variation.

3.2 Email Survey of Developers

Although privacy labels rarely change, this does not necessarily imply that they are inaccurate. To get a broad overview of reasons for privacy label changes and corresponding privacy policy changes, we reach out to app developers. To encourage replies, we kept these emails to developers short – limiting ourselves to three questions: (1) asking developers to explain why their privacy labels had changed or remained unchanged, (2) whether any corresponding updates had been made to their privacy policies and, if so, what those changes were, and (3) asking whether they would be willing to participate in a follow-up interview. Each email also included the lab's contact information, a consent form, and a link to the lab website. Section B in the Appendix shows an example email.

We customized these emails to each app, including a visualization of the changes to the app's privacy labels (see Figure 11 in the Appendix), along with a summary of changes generated with Gemini (exact version and prompt in Table 6). We only used publicly available information to generate these summaries. Figure 3 shows a sample of the generated summary included in the emails.

> **Data Linked to You:** The app now collects the user's Email Address, Name, Other User Content, and User ID which are linked to the user's identity for app functionality.
>
> **Data Not Linked to You:** The app now collects Performance Data under Diagnostics, for app functionality, which is not linked to the user's identity.

Figure 3: Example of privacy labels changes summary generated by Gemini.

Recruitment. Figure 2 shows the selection process for recruiting developers for this email survey. We focused on apps with recent privacy label updates (< 2 years) for which developers might still remember the reason for changes. Since we conducted the study in English, we excluded apps that did not have an English-language app names. For the remaining apps, we searched for the developers' email addresses in the publicly available privacy policy linked on the Apple App Store. We removed duplicate email addresses associated with multiple applications sharing the same privacy policy to avoid contacting the same developer multiple times. We also recruited developers who maintained consistent privacy labels, starting with the 285,463 apps with "Data Not Collected" label and the 447,234 apps with other consistent privacy labels. Using the same filtering as before, we found contact information for 4,972 "Data Not Collected" applications and 14,447 applications in the remaining categories. To ensure coverage across app types, we randomly sampled 1,000 apps from the "Data Not Collected" group and 1,000 apps from the remaining privacy label categories, making sure that apps from all App Store categories were included in each group. We stopped sending more mails after we noticed saturation (i.e., no new thematic insights) within the first 2000 sent emails. We did not collect demographic information from developers as the information about a single respondent would not necessarily be representative of the development team. In total, we received $N = 128$ responses as shown in Table 1. The assigned privacy labels in the sample are similar to the app store in general, only "Data Not Linked to You" is overrepresented in the sample (Table 2).

Table 1: Recruitment emails sent to apps with changed and unchanged privacy labels.

Mails	Changed	Unchanged
Sent	5,000	2,000
Bounced	380	136
Auto-Reply	498	195
Viable	87	41

Table 2: Distribution of privacy labels (Sample vs. App Store).

Privacy Label	Sample	App Store
Data Not Collected (DNC)	30%	30.97%
Data Used to Track You (T)*	30%	23.26%
Data Linked to You (L)*	37%	37.77%
Data Not Linked to You (NL)*	59%	43.99%

*Not mutually exclusive.

Analysis. Two authors conducted a conventional content analysis [22] of the email responses. First, they reviewed the responses independently to generate an initial set of codes, then met to discuss discrepancies and refine the codebook. The final codebook, shown in Appendix C, comprised 18 codes (12 for changed privacy labels and 6 for consistent privacy labels), covering topics such as reasons for changing or maintaining privacy labels, the update status of apps' privacy policies, the influence of monetization strategies on privacy labels, and the impact of third-party SDKs on label content. The coders independently coded all responses and then met to reconcile any differences.

3.3 Developer Interviews

Following up the email survey, we conducted a semi-structured interview study between July and December 2025. To gain an in-depth understanding of the socio-technical barriers for developers to maintain accurate APL, we interviewed $N = 19$ developers, 13 from apps that changed their privacy label, and 6 from apps that did not. Our interview protocol (see Appendix F) explored three main sections: (1) how decisions regarding privacy label content and changes are made, (2) developers' perceptions of the usefulness of privacy labels, and (3) developers' perspectives on the accuracy of privacy labels in representing their apps' data practices, including how privacy labels relate to and differ from apps' privacy policies. The interview questions were tailored to participant's specific context, exploring either their decision to update or reasons for not updating the labels. Interviews lasted between 40 and 60 minutes.

Recruitment. We recruited participants from a pool of email survey respondents who had expressed a willingness to participate in a follow-up interview. Specifically, a subset of 65 developers (43 associated with apps with changed privacy labels and 22 with unchanged privacy labels) who completed our email survey and were willing to participate in this follow-up interview. We sent out invitation emails to them. The invitation emails included a link to a pre-interview survey (included in Appendix E), asking participants for their background in app development, demographic information, and consent. The survey was administered using Qualtrics. We scheduled interviews via Calendly and conducted them via Zoom. All sessions were audio-recorded with participant permission and subsequently transcribed for analysis. The demographics of our participants are presented in Table 8. The distribution of interviewees' apps privacy labels varies, and all privacy labels types are represented in the sample. Apps with tracking (T) labels are 5/19, Linked labels (L) are 7/19, Not Linked (NL) are 8/19, and Data Not Collected (DNC) are 8/19.

Interview Process. The interview covered the label creation process, developers' perceptions of labels, and a reflection on the label accuracy. *(1) Label Creation Process:* We investigated the decision-making process regarding label creation and updates and where the responsibility of label management resides. We further asked the challenges and confusion encountered in this process and whether participants had any evaluation process to ensure accuracy. To situate APL in a broader context, we then inquired how participants synchronized disclosures across different apps and platforms and how they managed privacy policies. Finally, We asked participants to reflect on the impact of privacy labels on their actual data practices. *(2) Perception of Labels:* We first asked about perceived usefulness of APL from both the developer's perspective and their assessment of the user's perspective. We then inquired about the perceived necessity to communicate label update to users and the recommended methods to convey updates. *(3) Label Accuracy:* Prior research by Ali et al. [2] showed that privacy labels frequently diverge from the data practices described in apps' privacy policies. To explore this, we first inquired participants' subjective assessments of label accuracy. Then we showed them the results of Ali et al.'s policy analysis tool [2] (based on their app's privacy policy) along with the corresponding excerpts from the privacy policy. Figure 4 shows how we presented the privacy policy analysis results. We prompted them to reflect on the mismatches and queried whether they would find such automated analysis tools useful in their workflow. Finally, we solicited recommendations on privacy labels. Section F in the Appendix contains the full interview guideline.

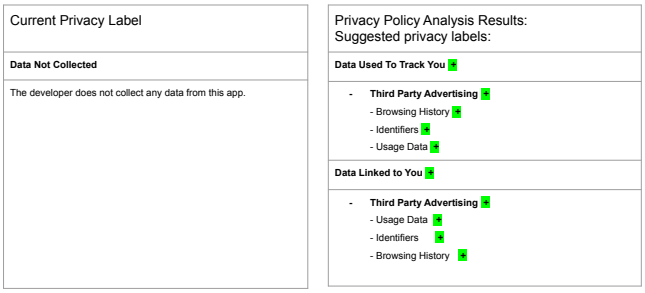

Figure 4: Example of privacy policy analysis results and suggested privacy labels shown to developers.

Analysis. We used inductive codebook-based thematic analysis [13] to analyze the interviews. We transcribed the interviews verbatim and two authors applied inductive coding facilitated by the software Taguette. In the first round, two researchers coded an initial subset of three interviews independently to develop two preliminary codebooks. They then met to resolve discrepancies, refine code definitions, and finalize a shared codebook. In the second round, using the refined codebook, the two researchers coded remaining interview transcripts and met regularly to iteratively update the codebook. When the coding was complete, the researchers organized the codes into topic summaries. Following McDonald et al.'s [37] recommendation, we did not calculate the inter-rater reliability since we aim to identify rich patterns rather than seek agreement. Section D in the Appendix includes the final codebook grouped a total of 58 topics.

3.4 Limitations

The main limitations stem from our large-scale email-based recruiting approach. The low email response rate of 1.8% (128 responses out of 7000 sent emails) could imply that only motivated respondents participated, limiting generalizability. However, the low response rate is similar to that in comparable studies that cold-emailed developers [1]. Most developers who responded were solo developers who worked on apps as a hobby or for smaller companies, which indicates that the findings may not be generalizable to larger companies. However, hobbyists and solo developers are an essential part of the developer community that impact the successful deployment of privacy mechanisms. In addition, for apps without privacy label changes, we randomly sampled 2,000 apps from the full set of eligible apps. We used stratified sampling based on the app categories to ensure diversity of all app types. This improves coverage for the qualitative study but comes at the expense of full representativeness. For this study, we focused on Apple's privacy labels, which relies on developer-reported information, including App Store privacy labels and privacy policies. These disclosures may not fully reflect an app's actual data collection and sharing

practices in production. Since we do not perform network traffic analysis or other forms of runtime measurement, we cannot verify the extent to which declared practices align with real-world behavior. We also acknowledge that Android has a similar system called the data safety section (DSS). However, data safety sections are structured differently (including data usage, handling, and security measures) and were also introduced a year after Apple's privacy labels, limiting comparability in a longitudinal study.

3.5 Ethical Considerations

Participation involved minimal risk, we used official university email accounts to reach out, clearly stating the study purpose and the research team's affiliation. Each outreach email included an informed consent document and stated that participation was voluntary. Consistent with the IRB determination, we treated replies as implied consent. To promote transparency, the lab's website posted the recruitment notice. Interview participants provided informed consent via a pre-interview survey, and confirmed verbally at the beginning of each interview, including consent for audio recording. We did not collect personally identifiable information and anonymized developers, applications, companies, and locations. Email correspondence and recordings were stored on university infrastructure. We deleted recordings after verifying the transcripts. To compensate survey respondents for their time, we offered a voluntary raffle for one of six $50 Amazon gift cards. All interview participants received a $20 Amazon gift card. Our institution's IRB approved this study and classified it as Exempt Human Subjects Research. Additionally, our dataset, obtained through weekly to quarterly scraping, was collected following standard ethical practices, including rate limiting to avoid undue server load and relying solely on publicly available records accessed through Apple's official API. No private records or application-specific APIs were accessed or used in the creation of the dataset.

4 RQ1: Changes in Privacy Labels

This section presents a longitudinal analysis of changes in Apple App Store privacy labels over a fixed analysis window spanning from July 21, 2021 to October 28, 2024. The longitudinal analysis describes the observed changes in privacy labels, how these privacy labels change privacy types and collected data categories, which kind of apps tend to change their privacy labels, and if these changes improve the alignment with the apps' privacy policies.

4.1 Do Privacy Labels Change?

When Apple introduced privacy labels, apps that already existed in the App Store did not have to implement them unless they released a new version. However, they could add privacy labels earlier if they wanted to. As Figure 5 shows, Apple's decision to tie privacy label compliance to new app versions leads to a steady growth of compliance from 42% in July 2021 to 86% in November 2025. We expect this rise to plateau in 2026 as it approaches the share of apps that are considered abandoned.

Most apps (76%), however, do not change their privacy labels. The number of apps that have changed their privacy label in the past rose slowly from 0.03% in July 2021 to 9.6% in November 2025 (Figure 5). These changes are not rising uniformly, though. In particular, between May 2023 and January 2024, the number of apps that modified their privacy labels jumped from 2.9% to 5.21% and continued to rise at an accelerated rate afterwards. These changes coincide with Apple's announcements on privacy manifests and the data practices of integrated SDKs in June 2023 [5], December 2023 [4], and February 2024 [6]. This sustained increase suggests that platform-level privacy announcements prompt developers to review and update their privacy labels over an extended period.

Of the apps that change their privacy label at all, almost all only change them once. While some change their privacy labels twice or three times, only few change them more often. Figure 6 shows the detailed distribution of the number of privacy label changes per app between July 2021 and October 2024. It suggests that most app developers treat privacy labels as largely static disclosures rather than frequently updated information.

Prior work [2] showed that 97% of apps with *Data Not Collected* label have a privacy label stating otherwise. By mid-2023, the number of apps using the *Data Not Collected* privacy label had increased to 30% of the dataset. Afterwards, the number plateaued and began to decrease. Since this privacy label does not require any more detailed information, we assume it might be the developers' first choice when they are forced to make a decision (because of an update) but are unsure about the actual data practices.

4.2 How do Privacy Labels Change?

We categorized privacy label updates into the four dimensions corresponding to their hierarchical structure: privacy type, purposes, data categories, and data types [3].

Privacy Types. We first consider the top level of privacy labels, the privacy types: *Data Used to Track You*, *Data Linked to You*, *Data Not Linked to You*, and *Data Not Collected* – where *Data Not Collected* is mutually exclusive with other labels. As Figure 7 shows, there is a general shift from the *Data Not Collected* privacy label to the other labels that indicate some type of personal data processing. The most often added privacy type is *Data Used to Track You*, indicating either an increase in tracking practices or more transparency in reporting them. The privacy types *Data Linked to You* and *Data Not Linked to You* show more internal changes, i.e., minor changes that maintain the high level privacy type, suggesting that data practices are hard to assign to these labels, requiring more follow-up corrections.

To better understand the dynamics of change, we investigated how apps transition between specific privacy type assignments. Figure 8 presents the transition matrix, summarizing of stability and change across privacy type assignments, analogous to transition matrices commonly used in state-based models. Most entries in the matrix are in the bottom right of the diagonal, meaning that the new privacy labels reported more expansive data collection. Three trends stand out: First, the cluster in the bottom right corner are all the apps that added *Data Used to Track You*, accounting for 41.9% of all transitions. Second, the line at the bottom shows all the apps that moved away from *Data Not Collected* to report some form of data use, amounting to 27.4% of all transitions. Third, the diagonal line from the bottom left to the upper right shows all label changes that did not change the privacy types (22.3%), i.e., a refinement of how collected data is reported in the privacy label. The highlighted cells next to the diagonal represent transitions between *Data Not*

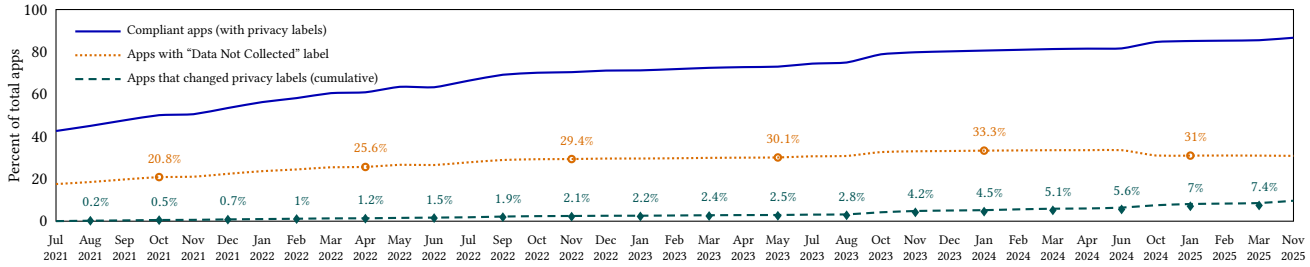

Figure 5: Longitudinal analysis of Apple App Store privacy labels (July 2021–November 2025).

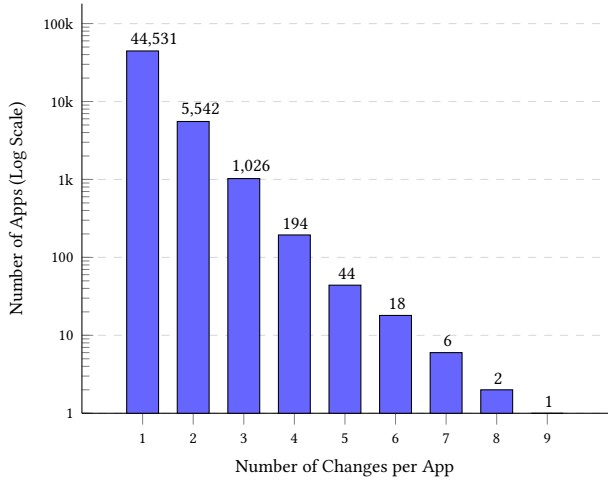

Figure 6: Distribution of privacy label changes across the N=51,364 apps that changed their privacy label.

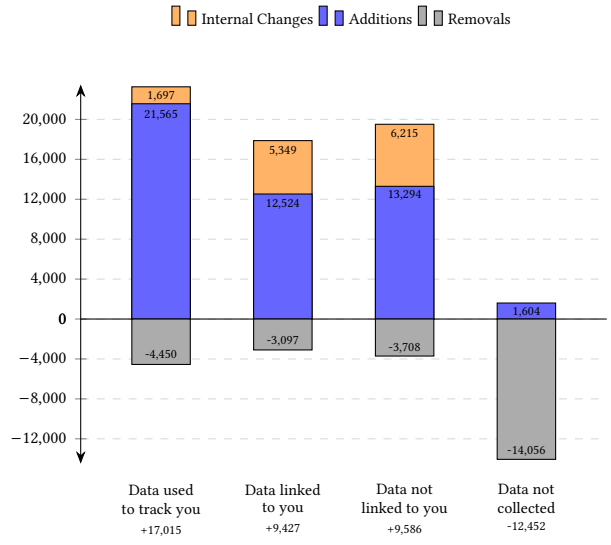

Figure 7: An overview of the number of apps that changed privacy labels broken down by the four privacy types.

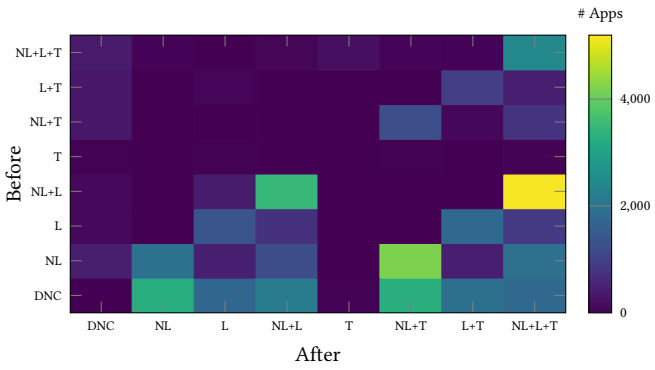

Figure 8: Transition matrix for privacy type assignments.

Notes: Cells show observed frequencies of reassignment between privacy types; the diagonal indicates unchanged assignments. DNC = Data not collected; NL = Data not linked to you; L = Data linked to you; T = Data used to track you.

Linked to You and *Data Linked to You* (10.7%). These adjacent shifts indicate that distinguishing between linked and non-linked data is a particularly unstable boundary in practice. Noteworthy, the row and column for *Data Used to Track You* is mostly empty, i.e, tracking rarely appears as a standalone classification but is typically reported alongside other forms of data use.

To gauge the effect of Apple's introduction of privacy manifests on privacy labels, we compared the privacy label transitions before and after the December 2023 announcement with a differential transition matrix (Figure 9). The transition shift is modest but noticeable, with the NL+L+T state contributing 46% of the total weighted KL divergence ($D_{\mathrm{KL}} = 0.062$). States that previously included tracking were less likely to remove the tracking and more likely to either add *Data Linked to You* or *Data Not Linked to You* when it wasn't there before. Apps with simpler privacy labels showed little practical change after the introduction of privacy manifests.

Purposes. Across apps that modified their privacy labels, reporting of data use purposes generally expanded. The most frequently added purposes were analytics (+3,090 apps), third-party advertising (+2,640), and developer advertising (+1,869), primarily under the *Data Linked to You* and *Data Not Linked to You* privacy types.

Data Categories. At the data category level, changes were more modest in magnitude but widespread across apps. Table 3 summarizes the top five data categories with the most changes within

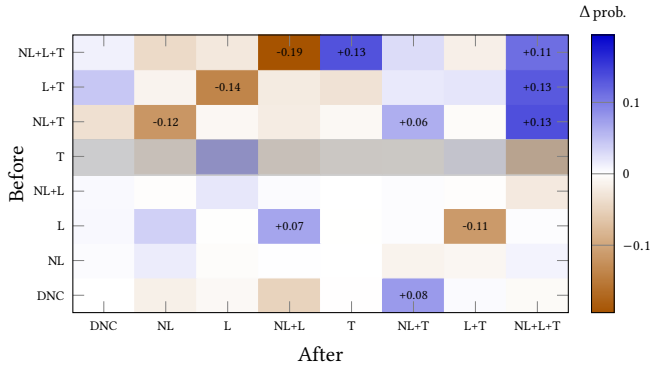

Figure 9: Difference in row-normalized transition probabilities (after December 2023 announcement of privacy manifests).

Notes: Blue = increased probability after December 2023; orange = decreased. Only cells with an absolute change above 0.05 are labeled. Grey row (T) is sparse and has fewer than 100 observed transitions.

Table 3: Most frequent additions and removals of app data categories within privacy types (PTs).

Category	Privacy Type	▲	▼	Net
Identifiers	Not Linked	2793	2189	+604
	Linked	2396	1450	+946
	Track	623	889	-266
Usage Data	Not Linked	2632	1205	+1427
	Linked	1983	1205	+778
	Track	615	931	-316
Diagnostics	Not Linked	2264	1224	+1040
	Linked	1126	903	+223
	Track	509	566	-57
Location	Not Linked	1801	1319	+482
	Linked	1617	1217	+400
	Track	793	812	-19
User Content	Not Linked	783	524	+259
	Linked	1345	768	+577
	Track	85	334	-249

these three privacy types, highlighting the type of change (addition or removal), the associated privacy type, and the number of affected apps. Most data categories involved in changes were Identifiers, Usage Data, Diagnostics, and Location, which experienced both additions and removals. The majority of these changes were linked with the Data Linked to You and Data Not Linked to You privacy types. The increase in disclosure of certain data categories in Figure 10 is particularly noticeable around November 2023. This period coincides with Apple's introduction of the Privacy Manifest and subsequent updates in December 2023 [5], indicating a platform-driven shift in developers' disclosure behavior.

Table 4: Negative binomial regression results for the frequency of privacy-label change runs.

Predictor	IRR (e^{β})	95% CI	p
App age (per year)	1.079	[1.076,1.082]	< .001
User rating count (per 10k)	1.001	[1.001,1.002]	< .001
User rating stars (per star)	1.199	[1.193,1.205]	< .001
Price (log-transformed)	0.513	[0.487,0.541]	< .001
In-app purchases (yes)	1.465	[1.428,1.504]	< .001
Genre fixed effects		Included	

Notes: Only shows predictors with a relative CI width less than 1. The Incidence Rate Ratio (IRR) was obtained by exponentiating the regression coefficient β.

Data Types. At the most granular level, data type changes were smaller in absolute numbers. Increases were observed for advertising data (+493), device identifiers (+121), and other usage data (+288), whereas decreases occurred for crash data (-802 apps), names (-331), and physical addresses (-303).

4.3 Which Apps Change Privacy Labels?

We used negative binomial regression to investigate whether the specific app characteristics drive label changes. The estimated dispersion parameter was $\theta = 0.29$, indicating substantial overdispersion relative to a Poisson model. The full model substantially improves fit relative to the intercept-only specification ($\Delta G^2 = 14{,}291$, likelihood-ratio test). McFadden's pseudo-R^2 is 0.042, indicating good explanatory power. Variance inflation factors (VIF) for all predictors were below 1.35, indicating no multicollinearity concerns. Table 4 reports the predictors with statistically significant effects and relative CI width < 1. The full results are in Appendix A.

App Age and Popularity. We considered that an app's popularity and age could increase the likelihood for updated privacy labels. We used the number of user ratings and the average rating of an app as proxies for popularity. We used the official App Store release date (excluding 809 apps for which it wasn't available) to determine the age. The release dates of apps with privacy labels changes in our dataset range from 2008 to 2023. Both app age and popularity are associated with a higher expected number of privacy label changes (Table 4): each additional year since an app's release is associated with a 7.9% increase, each 10,000-unit increase in user rating counts is associated with a 0.1% increase, and each additional average rating star is associated with a 19.9% increase in expected privacy label changes.

Monetization Strategy. Few apps in our dataset (1.2%) needed to be bought, but 35.3% of the apps offered in-app purchases. Our regression results (see Table 4), indicate that app developers' choice between these two types of monetization strategies had opposing associations. While up-front payment for apps was associated with a 48.7% (IRR = 0.513) lower expected number of privacy label changes per unit of log-transformed price, on-going monetization (through in-app purchases) was associated with a 46.5% (IRR = 1.465) increase in expected privacy label changes. This means a $1 paid app is associated with 37% fewer privacy label changes than a free app.

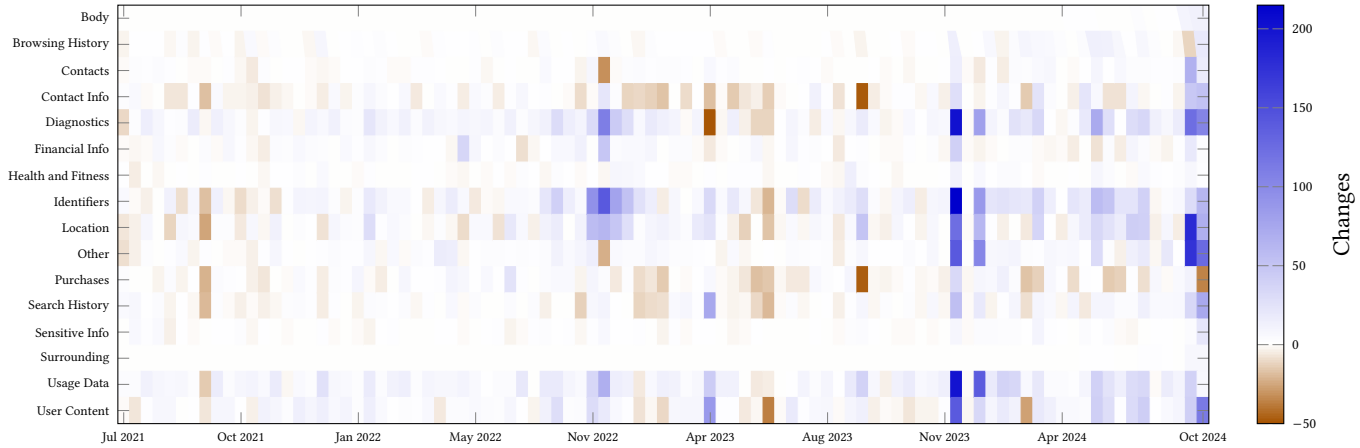

Figure 10: Heatmap of changes to collected data categories over time.

This suggests that a strong commitment to monetizing the app through up-front payments requires fewer privacy label changes while on-going monetization relies at least partially on personalized data collection, e.g., for advertisements, leading to more privacy label changes. In Section 5, we discuss how monetization was one of the factors that drove developers to change privacy labels.

App Category. We also analyzed differences in changes in privacy labels across app categories to understand whether updates are concentrated in particular types of apps. Our negative binomial regression model indicated statistically significant negative associations between some categories and the number of privacy label changes. However, the confidence interval of these associations is large (i.e., relative CI width above 1), so we did not include them in the main results of Table 4. Two of them, Games and Photo & Video, have a relative CI width below 2, making the estimates reasonably precise. They are associated with a 10.4–70.8% and 3.2–68.8% decrease in privacy label changes, respectively. See the full table in Appendix A for details.

4.4 Did The Privacy Labels Get More Accurate?

To understand how the updated labels align with the privacy policy of the app compared to previous labels, we used a BERT model from prior work [2]. For the 51,364 apps that changed privacy label, we identified 33,052 policies linked in the official app store. We looked at two levels of privacy labels, privacy types and data categories.

In total, 11,596 apps changed privacy types to better align with their stated privacy policies. The average hamming distance improved by 0.36, corresponding to apps changing one (0.25) or two (0.50) privacy types. In contrast, 9,942 apps aligned less with their privacy policies after the update. A two-sided sign test indicated that improvements were more frequent than expected under a 50–50 split ($p < .001$, 95% CI [0.532, 0.545]), corresponding to +3.84 percentage points below 0.5. Interestingly, 3,997 apps changed the privacy types without changing the (lack of) alignment with their policies. The majority of these (2,355 apps) made two privacy type changes. For data category disclosures, 9,637 apps improved alignment with the privacy policy, and 10,981 apps moved away from their privacy policy. A two-sided sign test showed that improvements were less frequent than expected under a 50–50 split ($p < .001$, 95% CI [0.461, 0.474]), i.e., −3.26 percentage points below 0.5. We also find that 4,836 apps improve their alignment to the policy by only refining the data categories but not the privacy types, indicating privacy label maturity.

Improvements are only marginally more (or less) likely than a coin flip, suggesting the absence of a strong directional correction mechanism. Privacy label changes may therefore reflect compliance formalities or developers' evolving interpretations rather than systematic efforts to improve alignment with the privacy policy. These findings motivated the deep-dive survey and interview questions, which we explore further in Section 5 and 6.3.

5 RQ2: Factors Leading to Change or Not

We conducted an initial email survey of developers to understand the factors (not) leading to change and the alignment with privacy polices. The 128 responses came from apps across diverse categories and popularity levels (see app characteristics in Table 7). Privacy label updates are mainly driven by app feature and functionality changes, while privacy labels remain unchanged primarily due to consistent data practices.

Why Developers Change Privacy Labels. We identified four topics explaining that label updates were driven by: *(1) App Features and Functionality Changes*, including adding or removing advertisements, modifying analytics, or updating third-party SDKs. For instance, P57 specified switching from anonymous ads to targeting ads to increase revenue: "We switched from anonymous to targeted ads [...] Targeted ads have a higher eCPM [~ revenue per 1,000 ads] compared to anonymous ads." *(2) Privacy Feature Changes*, related to integrating or removing App Tracking Transparency (ATT). *(3) Correction of Previous Labels*, which came up from an internal review of the privacy policy or due to initial confusion during the first label setup. P59 mentioned that Xcode's privacy reports helped them to correct their initial label, reducing uncertainty. *(4) Compliance with Apple Requirements and Regulations*, including legal regulations (e.g., GDPR) and third-party guidelines, as P28 explained: "only

reason we try to keep it up to date is because we don't want to get sued or rejected on app review."

Why Developers Do Not Change Privacy Labels. We also identified four topics explaining that unchanged labels were due to: *(1) No User Data Collected*, which could simplify the approval from Apple: "By not collecting any data, the approval process stays smooth and Apple approves app releases much faster." (P108), *(2) Unchanged Data Practices*, which could minimize liability, as P119 highlighted: "My belief has remained consistent that storing user data carries a significant liability risk, especially for a small business like ours.", *(3) No Perceived Need for Label Change* because "I didn't touch these and even don't know why I should to change these." (P103), and *(4) External Data Handling*, where data collection was outside of the developer's scope and relied on external entities.

Reasons for Misaligned Privacy Policies. In many cases, policy updates were driven by factors unrelated to data practices (e.g., language refinement). Some developers did not change their privacy policies because they intentionally used broad policies that include unimplemented features, covering multiple apps and platforms, as P88 explained: "the privacy policy was always much broader than the app store labels." In contrast, few developers who updated privacy labels without updating their policies acknowledged that their policies were outdated and planned to revise them. For example, P4 noted: "The policy still states that AdMob is used, and so is actually outdated now and should be updated." Finally, one developer (P25) reported that while they did not change their policy, the process of updating privacy labels increased their awareness of privacy practices. These findings confirm our policy analysis in Section 4.4 and help explain the gap between updated labels and apps' policies.

6 RQ3: Maintaining Privacy Labels

While the email survey results in Section 5 establish driving factors behind label updates (i.e., changes in app functionality and monetization strategies), these insights do not fully capture the internal workflows and challenges developers face. To address this gap, the following section presents our interview findings, which are organized around five topic summaries: *Privacy Label Maintenance Decision Process*, *Perceived Usefulness of Privacy Label Communication*, *Perceived Accuracy of Privacy Labels*, *Perceived Relationship Between Privacy Labels and Privacy Policies*, and *Challenges & Recommendations*.

To reflect the qualitative nature of our results and given our relatively small sample size, we prioritize emerging insights over generalization. Consistent with previous interview-based security and privacy studies [19, 45], we use the following terms to represent specific participant ranges: none (0%), a few (1–25%), some (26–45%), about half (46–55%), most (56–75%), almost all (76–99%), and all (100%).

6.1 Privacy Label Maintenance Decision Process

Overall, the responsibility of label maintenance depends on the development team size. For participants working individually or in a small team, they tend to interpret privacy labels by themselves; For those working in larger teams, the decision is collective and distributed across different people. In some cases, developers are not involved in the label decision.

Label Process and Maintenance Responsibility. When speaking of who made the decisions of maintaining the privacy labels, we observed that the responsibility heavily depends on the team size and lacks traceability. Specifically, for participants who worked as individuals or in small teams, they handled the entire life-cycle of the label: from initial creation to version updates. For a few participants working in larger teams, the responsibility was distributed across different members, departments, or even external companies. For example, P17 noted: "The apps I worked on professionally for work, I don't get to decide the privacy label. We have a legal team [...] So I don't really check the privacy label for my work." This distribution of labor can lead to challenges to trace the accountability, as P12 explained "everybody does little bits and pieces of it. And nobody really has everything." We observed similar patterns of organizational structures for labels' content decisions. Most participants referred to Apple's documentation and supplemental resources, such as third-party library guidelines for content decisions.

Label Maintenance Varies Across Apps and Platforms. Almost all participants had experience maintaining multiple apps. Among these, some reported maintaining the same decision-making process for different apps, while others indicated that their process varied by app popularity or different app data practices. For instance, P7 noted that they would pay more attention to popular apps, leaving others' privacy disclosures outdated: "[app name] is the app that has gotten the most attention and has the most users so and that I work on the most so there are probably out of date ones for my other apps." Most participants' apps were not available on Google Play Store. Among those whose apps were available on both stores, most reported managing each platform independently without cross-referencing their disclosures: "but I didn't compare it in any way. I wasn't consciously trying to make sure they both matched" (P2) Interestingly, P8 observed an increasing convergence between the Google and Apple ecosystem: "while they were very different at the very beginning [...] they're aligning more and more."

6.2 Perceived Usefulness of Privacy Label Communication

We observed a tension in developers' perception of privacy labels usefulness. Participants generally found privacy labels useful and favored communicating label updates to users, yet they perceive low user engagement due to *Privacy Fatigue*, *Peer Influenced Decision-Making*, and *No Meaningful Choices*.

Privacy Labels Are Useful to Developers and Users. Most participants find the labels useful for both developers and users. Interestingly, some indicated that labels build trust with users and help market their apps as privacy-friendly alternatives. For example, P7 stated: "I don't do nearly any tracking because I want to be a privacy-friendly alternative." For users, participants perceived labels as helpful to users for data transparency, app installation decision-making, and correcting potential misconceptions: "without the privacy labels they might think that we're collecting more than we actually are so that's why I think it's helpful for both of us." (P11) Consistently, they favored communicating privacy label

updates to users. Specifically, most of them proposed that such notifications should be delivered in-app and managed by Apple, as P4 noted: "the first time you open an app after the data labels have changed it should just alert you." In contrast, a few participants expressed concerns that privacy labels might mislead users: "usually all they're going to do is scare some people off because they feel like you're collecting too much data." (P15) This concern suggests while privacy labels enhance readability by simplifying complex details, they may simultaneously obscure nuances that are important for understanding data practices.

Users Do not Use Privacy Labels. Although participants generally found privacy labels useful, they perceived low user engagement with privacy labels. We identified three barriers to user interactions with privacy labels: *(1) Privacy Fatigue [14]*: some mentioned that users did not care about privacy: "people got used to apps being free and they got used to their data just constantly being collected so I think most people just don't care" (P7), *(2) Peer Influenced Decision-Making [17, 18]*: a few participants mentioned that users often install apps based on peer recommendations rather than store descriptions. In these cases, users tend to bypass privacy labels to trial the app directly: "they buy an app or download an app [...] because someone else told them to" (P18). *(3) No Meaningful Choices*: a few participants also pointed out that users lacked meaningful choices if they wanted apps with better privacy guarantee: "there's no real alternative [...] So then they don't care" (P6).

6.3 Perceived Accuracy of Privacy Labels

Although developers were generally confident that their labels reflected actual app behavior, they lacked a formal validation process. They were less certain about the alignment of privacy labels with their policies. Some privacy policies were overly broad or out-of-date, not accurately reflecting the practices.

Label Reflection of App Behavior and Policy Alignment. Almost all participants expressed confidence in the accuracy of their labels to reflect actual app behavior. Despite this general confidence, a few participants reported experiencing uncertainty involving third-party library behaviors. For example, P4 stated "when I added all the ones for the adverts I wasn't 100% sure that they were correct". However, around half of participants perceived their privacy labels to be well-aligned with the app's privacy policy while others perceived partial alignment or misalignment. A few participants with recent privacy labels changes, attributed misalignment to outdated privacy policies: "I think currently it might not even align at all [...] it's probably outdated" (P4). Interestingly, a few participants, such as P17, reported misalignment due to broad privacy policies: "It mentioned something we might use."

Lack of Validation Process. Most of developers have no formal process to evaluate the accuracy of their privacy labels. For instance, P11 stated: "after it's approved by Apple we don't audit them ourselves we just assume it's good to go," indicating that they relied on the platform's approval as implicit validation. P10 similarly expressed uncertainty about how to validate accuracy: "there's no way for me to say that this is 100% accurate." On the other hand, some participants described evaluation process, in which they self-validated privacy labels by following the platform and third parties' guidelines. For example, P3 noted: "It's really just following the guideline that Unity provides. So I just go through that checklist that they provide on what to add."

6.4 Perceived Relationship Between Privacy Labels and Privacy Policies

Participants generally perceived privacy labels and privacy policies to serve the same purpose but for different audiences: policies for legal professionals "the privacy policy is mostly like lawyer speak" (P1) and labels for users. Interestingly, P16 perceived the privacy labels as serving the platform itself "the privacy labels is something Apple wants to have to make it clearer for users". As articulated in Section 6.3, the accuracy of privacy labels was unclear. Similarly, privacy policies were also perceived as unreliable, due to factors such as outdated policies, the use of general-purpose policies, and errors introduced by automated policy generation tools.

Policy Maintenance Responsibility. We observe a similar pattern in responsibility structures for privacy policies and privacy labels, where for solo developers or in small teams, privacy policies are often assigned by the developers, while in larger organizations, policies are created and maintained by the legal team or other non-technical staff. This indicates that the scale of operations may impact how much policy and labels align.

Developers Agree with the Privacy Policy Analysis. After sharing with participants the privacy policy analysis results and the suggested privacy labels (see Figure 4), participants generally appreciated external and automated support in choosing accurate privacy labels. For example, P15 asked whether we could share the full analysis to discuss it with their team "because I think that it definitely brings up enough little questions for me that I feel like it'd be worth bouncing off, Some people to see [...]". In addition to acknowledging divergences, some participants also provided reasons for them. These included *outdated privacy policies:* "it's probably outdated actually at this point." (P4), *the use of general-purpose policies applied across multiple apps:* "but I use the exact same privacy policy for all of my games. [...] it's more simple. Otherwise, I'd have to have like, 20 different [policies]" (P3), and *errors introduced by automated policy generation tool:*"it is just the artifact from the generator" (P9)

6.5 Challenges & Recommendations

Developers offered several suggestions for improving privacy labels. We identified five reoccurring recommendations that tackle identified challenges:

Opaque Third-Party Practices and Compliance Require Better Guidance and Support. Some participants were uncertain about the data practices of integrated third-party libraries – "we're not the ones actually seeing the data" (P3) – complicating accurate reporting. This aligns with prior work calling for clear and standardized third-party guidelines to help developers comply with legal requirements [32, 39]. In addition, about half the participants reported facing compliance issues with the platform or legal standards. For example, when setting privacy labels, a few participants found "It was like as difficult as submitting taxes. [...] I have no idea if I did it right until months later." (P2) Others, like P14 faced challenges with

navigating global laws: "There's a lot of like the laws and maybe in EU and in America [...] I don't really want to deal with as a personal developer." Clearly, developers need more guidance from platforms. Besides improved documentation, a few developers also asked for efficient and granular feedback in Apple's review process.

Workflow support. P10 criticized that the current privacy label creation flow is disconnected from app development: "app privacy is just like done in the app web interface on the Apple side. So it's a very disconnected piece to [...] a regular app delivery." About half of participants expressed the need for more workflow support, particularly through automation of privacy labels management and support for teamwork. Furthermore, P13 proposed centralizing and automating third-party disclosures: "imagine some sort of central repository of privacy labels connected to each library, then this would just happen automatically." P10 proposed privacy labels version control for teamwork support, "so if something has changed, you can track changes of who may change as when," particularly since privacy is managed collectively across different roles in larger organizations.

Granular Privacy Information and User Agency. Some disclosed data practices apply only to specific features and a small subset of users. Suggestions included distinguishing between optional vs. required data collection and clarifying data use purposes: "like differentiate which data is used [and if] can user opt in or not" (P1). Others recommended separating disclosures by stakeholders, such as developers, third-party, and platforms: "there should be like a grayed out, Apple says we're taking financial info, [...] and it says Apple's handling this [...] but the developer isn't" (P12). Similarly, P3 noted "so anything with like these ads, [...] there would be nothing in the developer column and only anything in the third-party integrations." Developers (who did not collect data) also wanted to promoting privacy-preserving apps. For example, P19 suggested a search functionality to filter apps by privacy: "some advanced search that [...] they could search for, for apps that are not collecting any data". Design interventions such as separating disclosures by data handler or providing structured guidance of what developers need to disclose if no data handled by first-party, may reduce ambiguity and improve accuracy.

7 Discussion and Conclusion

7.1 What Do Changes in Privacy Labels Mean?

Changes in privacy labels are relatively infrequent. In the overwhelming amount of cases, updated privacy labels report an increase in data collection. A striking example in our dataset are the 14,056 apps that changed their label from *Data Not Collected* to different kinds of data practices. Does that mean these apps just started collecting more data? Not necessarily, since prior work [2] showed that 97% of these apps have a privacy policy stating otherwise. The pattern suggests that privacy labels may not fully reflect actual app data practices, consistent with Koch et al. [33], who compared traffic patterns with privacy labels. Underreporting of data practices can have several reasons: it could be out of convenience when the app store requires them, the actual data practices might not be easy to determine, the official guidelines could be hard to interpret in edge cases, or the developers might want to avoid unflattering privacy labels. These findings highlight that developers need more support and guidance to encourage accurate privacy labels disclosures that reflect the apps' data practices.

7.2 What Motivates Developers to Maintain Up-to-Date Labels?

All parts of the study, have suggested that Apple's announcements, requirements, and new tools, influenced developers to update their privacy labels. In particular, Privacy Manifests support developers in understanding third-party data practices, facilitated label revisions after its introduction. Developers frequently cited compliance with Apple's requirements and maintaining user trust as key motivation for updating labels. Prior work has documented inconsistencies and ambiguity in SDK privacy guidance [23], while our findings show how developers navigate this uncertainty in practice when determining what to disclose in privacy labels. Developers frequently rely on third-party documentation and platform-provided tools such as Xcode privacy reports to infer SDK data practices, suggesting that platform interventions can partially reduce the opacity of third-party ecosystems. However, developers still express uncertainty about disclosure responsibilities, particularly when data collection is performed by external third parties rather than by the app developers themselves. These findings highlight the need for clearer platform guidance and better differentiation between first-party and third-party data practices during the disclosure process. However, many are also convinced few users care about privacy labels at all. Without a structured process to choose and verify their privacy labels, developers rely on the platform to notify them when corrections are required. At the same time, it remains unclear if Apple verifies the accuracy of privacy labels. This suggests a potential mismatch between developers' assumptions and platform practices. To address this gap, we recommend introducing real-time feedback mechanisms during label submission. For example, the platform could compare Xcode's privacy reports with the information disclosed in privacy labels and flag inconsistencies. Then it could nudge developers towards a careful review to improve accuracy.

7.3 How Do Privacy Labels Relate to Policies?

Accuracy of privacy labels is hard to determine, prior studies have measured the accuracy of privacy labels using privacy policies as the baseline and identified discrepancies between labels and policies. These discrepancies have been documented at scale by Ali et al. [2] and found that many apps' indicated data collection is not reflected in the corresponding privacy policies. However, prior work has primarily focused on measuring inconsistencies, with less attention to the socio-technical mechanisms that produce them. Our study extends this line of work by exploring potential reasons for such discrepancies and how developers interpret the relationship between these two disclosure mechanisms. We find that privacy policies may not consistently serve as precise representations of actual app data practices. Developers often acknowledge differences between privacy labels and policies, but perceive privacy labels as more closely aligned with current app behavior, while framing privacy policies as legal documents designed for broad coverage rather than precise

behavioral descriptions. This creates a structural asymmetry between the two disclosure mechanisms. Privacy policies often serve multiple purposes beyond a single app, including reuse across multiple apps from the same developer or organization and compliance with broader legal requirements. As a result, they may be written in a generalized or shared form that is not always specific to individual apps. In contrast, privacy labels are defined at the per-app level and require app-specific declarations of data practices. Inconsistencies between privacy labels and policies may also arise from workflow separation , where labels and policies are often updated without coordination. As a result, observed discrepancies may not necessarily reflect intentional misrepresentation, but rather organizational and procedural constraints in maintaining synchronized disclosures.

7.4 Who Do Privacy Labels Actually Serve?

Across interviews, three entities emerged as the perceived beneficiaries of privacy labels: *developers*, *users*, and *Apple*. These perspectives raise a broader question, who are privacy labels ultimately designed to benefit?

Developers. Many developers perceived privacy labels as somewhat useful for them. They described labels as tools for communicating data practices, aiming for transparency and building trust. Several participants explicitly framed privacy labels as a a marketing tool, particularly for apps disclosing less data collection. Interestingly, this perceived marketing value coexisted with skepticism about user engagement. Developers generally believed that only a minority of users use privacy labels. Despite this, they assume that users' trust would increase for apps that collect less data. This aligns with prior work by Balash et al. [10], who found that users' risk perception depend on the privacy labels' content. However, whether privacy labels meaningfully influence real-world app selection remains unclear. In many practical context, users have few alternatives. For example, certain services (e.g., city-specific parking apps) may function as monopolistic or contract-bound platforms, leaving users with no choice regardless of privacy disclosures.

Users. Participants often doubted privacy labels' usefulness to users because of limited users attention and limited meaningful choices. Even as privacy labels may make users more aware of data practices, developers questioned users' control, particularly when alternative apps with similar functionality are unavailable. This highlights potential tension between transparency and control, privacy labels may increase visibility into data collection practices without necessarily expanding users' capacity to act on that knowledge. . To address the perceived gap in user usefulness, several developers recommended that Apple integrate privacy attributes into App Store search and filtering functionality. For example, privacy-related filters could allow users to search for alternatives such as 'Data Not Collected fitness apps'. Apple has recently implemented searchable accessibility attributes through its accessibility nutrition labels [8], suggesting that similar filtering mechanism for privacy disclosures may be technically feasible. Integrating privacy into search could increase labels visibility and create stronger competitive incentives for privacy-preserving apps.

Apple. One developer perceived privacy labels as serving Apple's institutional interests. To them, they allow Apple to market themselves as privacy-friendly, while shifting disclosure burdens to developers. Hence, privacy labels are only partially about user decisions but about platform governance and risk management.

Privacy labels are a step forward in communicating to users what happens to their data within app ecosystems. Carefully implemented, they could serve both users and developers. Privacy labels rarely change, and when they do, changes are typically influenced by compliance requirements or by updates to apps' data practices. Developers want to communicate their apps' data practices accurately but struggle to make it because of lacking feedback to self-interpreted guidelines. Most of these challenges stem from opaque third-party data practices and uncertainty about disclosure responsibilities when data is indirectly collected through external entities, such as SDK providers. The lack of user agency to restrict data practices or choose alternative apps also keeps developers from maintaining privacy labels. Hence, platforms should support developers in choosing appropriate privacy labels and strengthen users' control to increase the impact of privacy labels.

Acknowledgments

The authors used generative AI-based tools to revise the text, improve flow and correct any typos, grammatical errors, and awkward phrasing. We thank Masood Ali for his support with the BERT-based policy analysis and Jan Tolsdorf for his support in customizing the outreach emails to developers' apps at scale. We thank Daniel Zappala for his improvement suggestions. This work was supported by the National Science Foundation under Grant No. 2247952.

References

[1] Shubham Agarwal, Rafael Mrowczynski, Maria Hellenthal, and Ben Stock. 2025. "I have no idea how to make it safer:" Studying Security and Privacy Mindsets of Browser Extension Developers. In *34th USENIX Security Symposium (USENIX Security 25)*. 2927–2946. https://dl.acm.org/doi/10.5555/3766078.3766229

[2] Mir Masood Ali, David G Balash, Monica Kodwani, Chris Kanich, and Adam J Aviv. 2024. Honesty is the Best Policy: On the Accuracy of Apple Privacy Labels Compared to Apps' Privacy Policies. *Proceedings on Privacy Enhancing Technologies* 4 (2024), 142–166. https://petsymposium.org/popets/2024/popets-2024-0111.php

[3] Apple Developer. 2022. *App Privacy Labels Now Live on the App Store.* Retrieved Jan 10, 2026 from https://developer.apple.com/news/?id=3wann9gh

[4] Apple Inc. 2023. *Privacy updates for App Store submissions.* Retrieved Feb 1, 2026 from https://developer.apple.com/news/?id=r1henawx

[5] Apple Inc. 2023. *What's new in privacy on the App Store.* Retrieved Feb 1, 2026 from https://developer.apple.com/news/?id=av1nevon

[6] Apple Inc. 2024. *Privacy updates for App Store submissions.* Retrieved Feb 1, 2026 from https://developer.apple.com/news/?id=3d8a9yyh

[7] Apple Inc. 2025. *App privacy details on the App Store.* Apple Developer. Retrieved Dec 17, 2025 from https://developer.apple.com/app-store/app-privacy-details/

[8] Apple Inc. 2025. *Overview of Accessibility Nutrition Labels.* Retrieved Feb 2, 2026 from https://developer.apple.com/help/app-store-connect/manage-app-accessibility/overview-of-accessibility-nutrition-labels/

[9] Ioannis Arkalakis, Michalis Diamantaris, Serafeim Moustakas, Sotiris Ioannidis, Jason Polakis, and Panagiotis Ilia. 2024. Abandon All Hope Ye Who Enter Here: A Dynamic, Longitudinal Investigation of Android's Data Safety Section. In *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association, Philadelphia, PA, 5645–5662. https://www.usenix.org/conference/usenixsecurity24/presentation/arkalakis

[10] David G. Balash, Mir Masood Ali, Chris Kanich, and Adam J. Aviv. 2024. "I would not install an app with this label": Privacy Label Impact on Risk Perception and Willingness to Install iOS Apps. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*. USENIX Association, Philadelphia, PA, 413–432. https://www.usenix.org/conference/soups2024/presentation/balash

[11] David G. Balash, Mir Masood Ali, Monica Kodwani, Xiaoyuan Wu, Chris Kanich, and Adam J. Aviv. 2025. Longitudinal Analysis of Privacy Labels in the Apple App Store. arXiv:2206.02658 [cs.CR] https://arxiv.org/abs/2206.02658

[12] Reuben Binns, Ulrik Lyngs, Max Van Kleek, Jun Zhao, Timothy Libert, and Nigel Shadbolt. 2018. Third Party Tracking in the Mobile Ecosystem. In *Proceedings of the 10th ACM Conference on Web Science* (Amsterdam, Netherlands) *(WebSci '18)*. ACM, New York, NY, USA, 23–31. https://doi.org/10.1145/3201064.3201089

[13] Virginia Braun and Victoria Clarke. 2021. *Thematic Analysis: A Practical Guide*. Sage Publications.

[14] Hanbyul Choi, Jonghwa Park, and Yoonhyuk Jung. 2018. The role of privacy fatigue in online privacy behavior. *Computers in Human Behavior* 81 (2018), 42–51.

[15] Lorrie Faith Cranor. 2012. Necessary but Not Sufficient: Standardized Mechanisms for Privacy Notice and Choice. *Journal on Telecommunications and High Technology Law* 10 (2012), 273–308.

[16] Lorrie Faith Cranor, Candice Hoke, Pedro Leon, and Alyssa Au. 2014. Are they worth reading? An in-depth analysis of online advertising companies' privacy policies. In *2014 TPRC Conference Paper*. https://doi.org/10.2139/ssrn.2418590

[17] Sauvik Das, Adam D.I. Kramer, Laura A. Dabbish, and Jason I. Hong. 2014. Increasing Security Sensitivity With Social Proof: A Large-Scale Experimental Confirmation. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (Scottsdale, Arizona, USA) *(CCS '14)*. ACM, New York, NY, USA, 739–749. https://doi.org/10.1145/2660267.2660271

[18] Alexander De Luca, Sauvik Das, Martin Ortlieb, Iulia Ion, and Ben Laurie. 2016. Expert and Non-Expert Attitudes towards (Secure) Instant Messaging. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. 147–157.

[19] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. 2019. Exploring How Privacy and Security Factor into IoT Device Purchase Behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300764

[20] Alisa Frik, Subham Mitra, Priyasha Chatterjee, and Julia Bernd. 2026. Are Bite-Size Data Safety Details a Healthy Diet for Android Telehealth App Users? Impacts of Privacy Nutrition Labels on Users' Privacy Perceptions. *Proceedings on Privacy Enhancing Technologies* 1 (2026), 366–392. https://doi.org/10.56553/popets-2026-0019

[21] Google. 2020. *Provide information for Google Play's Data safety section - Play Console Help*. Retrieved Feb 3, 2026 from https://support.google.com/googleplay/android-developer/answer/10787469

[22] Hsiu-Fang Hsieh and Sarah E. Shannon. 2005. Three Approaches to Qualitative Content Analysis. *Qualitative Health Research* 15, 9 (2005), 1277–1288. https://doi.org/10.1177/1049732305276687 PMID: 16204405.

[23] Hiroki Inayoshi, Shohei Kakei, and Shoichi Saito. 2024. Detection of Inconsistencies between Guidance Pages and Actual Data Collection of Third-party SDKs in Android Apps. In *Proceedings of the IEEE/ACM 11th International Conference on Mobile Software Engineering and Systems* (Lisbon, Portugal) *(MOBILESoft '24)*. ACM, New York, NY, USA, 43–53. https://doi.org/10.1145/3647632.3647991

[24] Akshath Jain, David Rodriguez, Jose M. Del Alamo, and Norman Sadeh. 2023. ATLAS: Automatically Detecting Discrepancies Between Privacy Policies and Privacy Labels. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroSPW)*. 94–107. https://doi.org/10.1109/EuroSPW59978.2023.00016

[25] Carlos Jensen and Colin Potts. 2004. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. ACM, New York, NY, USA, 471–478. https://doi.org/10.1145/985692.985752

[26] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A "nutrition label"' for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (Mountain View, California, USA) *(SOUPS '09)*. ACM, New York, NY, USA, Article 4, 12 pages. https://doi.org/10.1145/1572532.1572538

[27] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. 2010. Standardizing privacy notices: an online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. ACM, New York, NY, USA, 1573–1582. https://doi.org/10.1145/1753326.1753561

[28] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. 2013. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 3393–3402. https://doi.org/10.1145/2470654.2466466

[29] Rishabh Khandelwal, Asmit Nayak, Paul Chung, and Kassem Fawaz. 2023. Comparing Privacy Labels of Applications in Android and iOS. In *Proceedings of the 22nd Workshop on Privacy in the Electronic Society*. 61–73. https://dl.acm.org/doi/10.1145/3603216.3624967

[30] Rishabh Khandelwal, Asmit Nayak, Paul Chung, and Kassem Fawaz. 2024. Unpacking Privacy Labels: A Measurement and Developer Perspective on Google's Data Safety Section. In *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association, Philadelphia, PA, 2831–2848. https://www.usenix.org/conference/usenixsecurity24/presentation/khandelwal

[31] Mugdha Khedkar, Ambuj Kumar Mondal, and Eric Bodden. 2024. Do Android App Developers Accurately Report Collection of Privacy-Related Data?. In *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering Workshops* (Sacramento, CA, USA) *(ASEW '24)*. ACM, New York, NY, USA, 176–186. https://doi.org/10.1145/3691621.3694949

[32] Simon Koch, Manuel Karl, Robin Kirchner, Malte Wessels, Anne Paschke, and Martin Johns. 2025. The Impact of Default Mobile SDK Usage on Privacy and Data Protection. *Proceedings on Privacy Enhancing Technologies* 2025, 1 (Jan. 2025), 808–823. https://doi.org/10.56553/popets-2025-0042

[33] Simon Koch, Malte Wessels, Benjamin Altpeter, Madita Olvermann, and Martin Johns. 2022. Keeping Privacy Labels Honest. *Proceedings on Privacy Enhancing Technologies* 2022 (2022), 486–506. Issue 4. https://doi.org/10.56553/popets-2022-0119

[34] Yucheng Li, Deyuan Chen, Tianshi Li, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I. Hong. 2022. Understanding iOS Privacy Nutrition Labels: An Exploratory Large-Scale Analysis of App Store Data. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. ACM, New York, NY, USA, Article 356, 7 pages. https://doi.org/10.1145/3491101.3519739

[35] Yanzi Lin, Jaideep Juneja, Eleanor Birrell, and Lorrie Faith Cranor. 2024. Data Safety vs. App Privacy: Comparing the Usability of Android and iOS Privacy Labels. arXiv:2312.03918 [cs.HC] https://arxiv.org/abs/2312.03918

[36] Aleecia M McDonald and Lorrie Faith Cranor. 2008. The Cost of Reading Privacy Policies. *I/S: A Journal of Law and Policy for the Information Society* 4 (2008), 543–568. Issue 3.

[37] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (Nov. 2019), 23 pages. https://doi.org/10.1145/3359174

[38] Abraham H. Mhaidli, Yixin Zou, and Florian Schaub. 2019. "We Can't Live Without Them!" App Developers' Adoption of Ad Networks and Their Considerations of Consumer Risks. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA, 225–244. https://www.usenix.org/conference/soups2019/presentation/mhaidli

[39] Trung Tin Nguyen, Michael Backes, and Ben Stock. 2022. Freely Given Consent? Studying Consent Notice of Third-Party Tracking and Its Violations of GDPR in Android Apps. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security* (Los Angeles, CA, USA) *(CCS '22)*. ACM, New York, NY, USA, 2369–2383. https://doi.org/10.1145/3548606.3560564

[40] Federica Paci, Jacopo Pizzoli, and Nicola Zannone. 2023. A Comprehensive Study on Third-Party User Tracking in Mobile Applications. In *Proceedings of the 18th International Conference on Availability, Reliability and Security* (Benevento, Italy) *(ARES '23)*. ACM, New York, NY, USA, Article 97, 8 pages. https://doi.org/10.1145/3600160.3605079

[41] Abbas Razaghpanah, Rishab Nithyanand, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Mark Allman, Christian Kreibich, and Phillipa Gill. 2018. Apps, Trackers, Privacy, and Regulators: A Global Study of the Mobile Tracking Ecosystem. In *Proceedings 2018 Network and Distributed System Security Symposium* (San Diego, CA, USA). https://doi.org/10.14722/ndss.2018.23353

[42] David Rodriguez, Akshath Jain, Jose M. Del Alamo, and Norman Sadeh. 2023. Comparing Privacy Label Disclosures of Apps Published in both the App Store and Google Play Stores. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroSPW)*. 150–157. https://doi.org/10.1109/EuroSPW59978.2023.00022

[43] Gian Luca Scoccia, Marco Autili, Giovanni Stilo, and Paola Inverardi. 2022. An empirical study of privacy labels on the Apple iOS mobile app store. In *Proceedings of the 9th IEEE/ACM International Conference on Mobile Software Engineering and Systems* (Pittsburgh, Pennsylvania) *(MOBILESoft '22)*. ACM, New York, NY, USA, 114–124. https://doi.org/10.1145/3524613.3527813

[44] Yue Xiao, Zhengyi Li, Yue Qin, Xiaolong Bai, Jiale Guan, Xiaojing Liao, and Luyi Xing. 2023. Lalaine: Measuring and Characterizing Non-Compliance of Apple Privacy Labels. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA, 1091–1108. https://www.usenix.org/conference/usenixsecurity23/presentation/xiao-yue

[45] Yaman Yu, Tanusree Sharma, Melinda Hu, Justin Wang, and Yang Wang. 2025. Exploring Parent-Child Perceptions on Safety in Generative AI: Concerns, Mitigation Strategies, and Design Implications. In *2025 IEEE Symposium on Security and Privacy (SP)*. 2735–2752. https://doi.org/10.1109/SP61157.2025.00090

[46] Shikun Zhang, Yuanyuan Feng, Yaxing Yao, Lorrie Faith Cranor, and Norman Sadeh. 2022. How Usable Are iOS App Privacy Labels? *Proceedings on Privacy Enhancing Technologies* 2022 (2022), 204–228. Issue 4. https://doi.org/10.56553/popets-2022-0106

[47] Shikun Zhang and Norman Sadeh. 2023. Do Privacy Labels Answer Users' Privacy Questions?. In *Symposium on Usable Security and Privacy (USEC)* (San Diego, CA, USA). https://doi.org/10.14722/usec.2023.232482

Table 5: Negative Binomial Regression Predicting Privacy Label Changes

Predictor	IRR	95% CI	p-value	Rel. CI width
Intercept	0.026	[0.015, 0.046]	$< 0.001^{***}$	1.19
App genre (reference category omitted)				
Adventure	0.340	[0.137, 0.840]	0.020^{*}	2.07
Board	0.326	[0.124, 0.857]	0.023^{*}	2.25
Books	0.634	[0.358, 1.122]	0.118	1.21
Business	0.714	[0.407, 1.252]	0.239	1.18
Card	0.625	[0.262, 1.488]	0.288	1.96
Casino	0.518	[0.237, 1.136]	0.101	1.74
Casual	0.512	[0.148, 1.774]	0.291	3.18
Developer Tools	0.693	[0.370, 1.295]	0.250	1.34
Education	0.709	[0.405, 1.244]	0.231	1.18
Entertainment	0.713	[0.406, 1.252]	0.239	1.19
Family	0.585	[0.265, 1.293]	0.185	1.76
Finance	0.933	[0.532, 1.636]	0.808	1.18
Food & Drink	1.343	[0.766, 2.356]	0.304	1.18
Games	0.511	[0.292, 0.896]	0.019^{*}	1.18
Graphics & Design	0.586	[0.328, 1.047]	0.071	1.23
Health & Fitness	0.814	[0.464, 1.428]	0.474	1.18
Lifestyle	0.793	[0.452, 1.390]	0.418	1.18
Magazines & Newspapers	0.986	[0.547, 1.777]	0.962	1.25
Medical	0.818	[0.465, 1.439]	0.486	1.19
Music	0.866	[0.493, 1.521]	0.616	1.19
Navigation	0.878	[0.498, 1.549]	0.654	1.20
News	1.278	[0.727, 2.245]	0.394	1.19
Photo & Video	0.550	[0.312, 0.968]	0.038^{*}	1.19
Productivity	0.621	[0.354, 1.091]	0.097	1.19
Puzzle	0.854	[0.356, 2.051]	0.725	1.99
Racing	0.823	[0.398, 1.702]	0.599	1.58
Reference	0.647	[0.367, 1.141]	0.133	1.20
Roleplaying	12.543	[4.132, 38.074]	$< 0.001^{***}$	2.71
Shopping	1.559	[0.889, 2.734]	0.121	1.18
Simulation	0.491	[0.161, 1.501]	0.212	2.73
Social Networking	0.859	[0.488, 1.510]	0.597	1.19
Sports	1.198	[0.682, 2.102]	0.530	1.19
Stickers	0.012	[0.002, 0.090]	$< 0.001^{***}$	7.33
Strategy	0.420	[0.170, 1.042]	0.061	2.08
Travel	1.129	[0.643, 1.982]	0.672	1.19
Trivia	0.710	[0.333, 1.511]	0.374	1.66
Utilities	0.595	[0.339, 1.044]	0.070	1.19
Weather	0.728	[0.407, 1.300]	0.283	1.23
Word	0.493	[0.212, 1.149]	0.101	1.90
App characteristics				
App age (years)	1.079	[1.076, 1.082]	$< 0.001^{***}$	0.07
Price†	0.513	[0.487, 0.541]	$< 0.001^{***}$	0.11
In-app purchases	1.465	[1.428, 1.504]	$< 0.001^{***}$	0.05
User rating stars	1.199	[1.193, 1.205]	$< 0.001^{***}$	0.01
User rating count (per 10k)	1.001	[1.001, 1.002]	$< 0.001^{***}$	0.00

Notes: Incidence Rate Ratios (IRR) reported with Wald 95% confidence intervals. App age was rescaled from days to years. Relative CI width is defined as $(CI_{\text{high}} - CI_{\text{low}})/IRR$. Several genre categories perfectly predict zero counts and yield IRRs of zero with unbounded confidence intervals; these are omitted from interpretation. †To improve numerical stability and reflect the diminishing marginal effect of higher prices, we model price as $log(1 + price)$. Significance denoted as $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

A Full Negative Binomial Regression Model Results

Table 5 contains the full results for all predictors we checked in the negative binomial regression model.

B Example Email to App Developers

Below is the simplified text of the outreach email sent to developers. We used the prompt shown in Table 6 to describe the differences between privacy labels in the app store and the ones suggested by our BERT model. Figure 3 shows an example of these privacy label changes summaries. The first question was dynamically adjusted based on whether the app has changes in the privacy labels.

> **Subject: Research Study Invite: Why Have Your App's Privacy Labels Changed?**
> Dear [Developer Name],
> We are researchers from the [redacted] University. Your app, **[App Name]** was part of our study analyzing how privacy labels evolve over time.
> We would value your insights to help us better understand our findings. By replying to the questions below, you can enter a raffle for one of three $50 gift cards.
> **Findings for [App Name]:**
> [a Gemini-generated summary was included here.]
> **Questions:**
> 1. **(For apps with changes):** What were the reasons for changing the privacy labels of this app?
> 2. **(For apps without changes):** Can you share the reasons why the privacy labels for this app have remained consistent over the past updates?
> 3. Were there any changes to your privacy policy during this time, and if so, in what way?
> 4. Would you be willing to participate in a brief follow-up interview to help us understand developers' practices with privacy labels?
>
> Thank you for your time.
> [A visualization of label changes and the consent form was attached]
> Best regards,

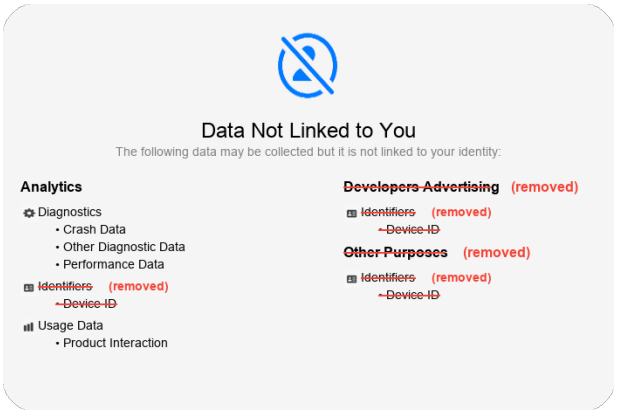

Figure 11: Example of the privacy label changes visualization attached in developers outreach emails. This specific example highlights removals (red/strikethrough) for an app's "Data Not Linked to You" category.

Table 6: LLM Configuration and Prompt Specification

Component	Specification
Model	Gemini 2.0 Flash (`gemini-2.0-flash`)
Temperature	1.0
Top-p / Top-k	0.95 / 40
Prompt	*"I'm going to paste a json object which provides a full summary of the changes to a iOS privacy label. Please write three sentences which summarize the differences within any of the categories that have changes mentioned in the json file (possible categories: data not collected, data linked to you, data not linked to you, and data used to track you). The sentences should be no more than 40 words. These sentences should include brief necessary explanations so that a non-expert can understand their meaning. (replace 'you' and 'your' in the explanation to 'user' because we are contacting developers about their apps; but keep them in the original Data Types names). I want the format to be like: Data Linked to You: summary explanation, Data Used to Track You: summary explanation. [...] If categories was added or removed, please include it as well."*

C Codebook for the Email Responses

Apps with Changed Labels:

Features & Functionality Changes (74)
Added Advertisement (19), Removed Advertisement (17), Tracking User Interaction (1), Added Location (3), App Functionality Changes (2), Removed Location (1), User Account (2), Added Analytics (20), Not Using this SDK Anymore (3), Removed Analytics (6)
Privacy Feature Changes (2)
App Tracking Transparency (2)
Correction of Previous Privacy Labels (12)
Incorrect Previous Labels (7), Re-evaluation of Data Practices (3), Confusion about Privacy Labels (2)
Compliance with Apple Requirement & Other Regulations (18)
Compliance with SDK Data Practices (7), Compliance with App Stores Changes and Requirements (11)
Motivation for Changes (Users) (13)
User Centered Decision (13)
Motivation for Changes (App Improvements) (8)
App Improvement (3), App Redesign (5)
Motivation for Changes (Changes in Data Collection Practices) (11)
Collect Less Data (9), Collect More Data (2)
Motivation for Changes (Monetization Strategy) (20)
Policy Changed (45)
Policy Updated for Platform Compliance (7), None Data practices Change (4), Reflect Data Practice Change in Either App or the Label (34)
No Policy Changes (29)
Broad Policy Language (7), Planned Update (3), Insignificance Data Collection (3), Privacy Policy Did not change (15), Consequence of PL Change (1)

New Policy Created (2)
New/Initial Privacy Policy Created (2)
Uncertain about Policy Change (1)

Apps with Unchanged Labels:

No User Data Collected (25)
Data Processed Locally (2), No Business or Functional Need for User Data (14), Lack of Resources to Manage Data (2), No Data Collection (7)
No Perceived Need for Label Change (4)
Label Reflect Current Practices (4)
Unchanged Data Practices (2)
Unchanged App Functionality (4), Unchanged Data Handling (8), Unchanged Third-party Integration (3), Minimal Data Processing (5)
External Data Handling (4)
Data Collected by Apple (1), Data Collected by External Integration via login at run-time (1), Data Collected by External Integration via SDK in Code-Base (2)
Policy Changed (11)
App features changes (1), Non data practices changes (8), Compliance based changes (1), App architecture changes (1)
No Policy Changes (26)
Unchanged Privacy Policy (24), No Personal Data Collected (1), Periodic Policy Review (1)

D Codebook for Interviews

Apps with Changed Labels:

App Influenced Changes
App feature update, Data tracking updates, Business Model Change, Removal of third-part integration
Developer Internal decision making
Simultaneous label change, Developer's due diligence, Developers' experience, Unused Privacy Labels
Influence by external factors
Platform requirements (Apple), Region driven privacy compliance, Reactive compliance
User Influenced decision
User-Initiated Label Update
Independent responsibility
Independent developer manages PL
Distributed responsibility
Founder manages PL, CFO, Lack of Internal Traceability, Other team members
Privacy Conscious Decisions
Be trustworthy, Data minimization, Privacy-Driven Decisions, Reflecting Actual App Behavior
Internal decisions
Internal discussions, Cross-departmental work, Self-Guided Label Interpretation
App Marketing
Marketing strategy
Decisions Based on External Resources
Apple Documentation, Other External Resources, Third-party policies & guidelines, Do not review third-party documents
Compliance
Balancing Compliance and User Perception, Minimum Required Disclosure
Experience Some Difficulties
Medium difficulty, Frustrating process, Label Decision Effort, Time consuming
Easy Process
Across Apps Process Synchronizations
Same process across apps, Variation by app popularity, Variation by app complexity or data practices
Across Platform Synchronizations
Available in other platforms (google) , Cross-Platform Labeling Difference, Independent Platform Labeling, Effort Toward Cross-Platform Alignment, App not available in other platforms
Useful for Users
Contextual based usefulness , Useful in Correcting Misconceptions, Useful in transparency, Useful in trustworthiness, Useful in apps selection
Lack of Usefulness for Users
Not useful for users , Privacy Resignation, Potential to Mislead Users, Meaningless Privacy Indicator, Uncertain usefulness for customers
Useful for Developers
Useful in marketing
Lack of Usefulness for Developers
Not useful for developers, No technical usability, Uncertain usefulness
Label Impact on Users
Potential misrepresentation of app, Minor impact, Contextual-Based Impact (inaccuracy, Extensive Data Collection)
Users Engagement with Labels
Low User Engagement, Users Download Before Evaluating Information, User engagement when uncertain, User disengagement for popular apps, Low priority from users & developers, Low engagement with app information, Limited Choices reduces user engagement
Support of Communicating Changes
OS notification by Apple, In-App Notification, Communication if changes have large impact, Communicating only if additions, Potential usefulness for communicating updates, Opposed to communicating updates, Communication Lacks Meaningful Choice For Users
Label-Policy Alignment
Misalignment, Outdated Privacy Policy, Privacy policy as afterthought, Label-Policy Well Aligned, Partial Alignment, Broader policy scope
Labels Review
No validation process, Ad-hoc Process
Label Accuracy Confidence
Confidence, Uncertain
Action About Uncertainty
Platform Feedback Reliance, Following platform guidelines, Manual rechecking
Policy Maintenance Responsibility
Independent developer responsibility, Distributed responsibility

Label-Policy Purposes
Same purposes, Different purposes, Policies serves developers, labels serves platforms, Labels are concise version of policies, Labels are consumer-centric
Label-Policy Differences Agreement
Disagreement of differences, Agreement on difference, Uncertain about the difference
Reflection on Label-Policy Differences
Wrong privacy policy link, Potential policy update, General-use privacy policy
Policy Analysis Tool
Potential interest in policy analysis tool, Use with reservations
Challenges
Compliance, Apple Requirements, Usability , Third-parties Data Practice Blackbox
Label Impact on Developers
Mindful data collection, App Feature Removal
Recommendations
Information presentation, New UI/UX , workflow support, further support/guidance

Apps with Unchanged Labels:

Reasons of Labels Consistency
Privacy-Driven Decisions, Risk avoidance, User-Driven Simplification of Privacy Labels, Compliance-driven labeling , Reflecting Actual App Behavior, User-driven data minimization
Independent Responsibility
Independent developer manages PL
Distributed Responsibility
Legal team, Other team members (e.g.,CTO)
Third-Party Integration
Third party Integrated, External Resources, Not aware of third-party data practices, Personal Interpretation of third-party practices, No third-party integration
Lack of Label Review
Static reliance on initial label , Minimal new data collection, Reactive Compliance, Functional Primacy
Variation in Practices
Variation by app data practices , Variation by app purpose
Across Platform Synchronizations
Available in other platforms (google) , Same Minimal Data Collection, One-time setup, App not available in other platforms
Not Useful for Developers
Potential to Mislead Users , Privacy-Usability Trade-Off, Increase labor
Useful for Developers
Useful in trustworthiness , Useful in communicating in data practices, Useful for marketing
Useful for Users
Useful in apps selection , Useful for user satisfaction
Not Useful for Users
Lack control
Low User Engagement
Lack of Choice, Privacy Fatigue, PL as a Tech Savvy Thing, Social-First Decision Making, PL Low-Hierarchy Placement
Labels Impact on Users
Contextual-Based Impact, User Skepticism of Privacy Label, Increase User Trust (Data Minimization)
Support of Communicating Changes
OS notification by Apple, In-App Notification, By email from developers, Communicate based on privacy implications, Avoid privacy fatigue
Label-Policy Alignment
Label-Policy Well Aligned, Privacy Policy Shaped by Privacy Labels, Partial Alignment, AI-generated Policy, Template-based Policy, Confidence in label-policy Accuracy
Policy Maintenance Responsibility
Independent developer responsibility, Distributed responsibility
Perceived Label-Policy Purposes
Same purposes, Different purposes, Policies serves developers, labels serves platforms, Labels are concise version of policies, Labels are consumer-centric
Labels Review
One-time setup, App Change-triggered privacy policy review, Legal Change-triggered privacy policy review
Label-Policy Differences Agreement
Disagreement of differences, Partial Agreement on difference, Precautionary Over-Disclosure
Policy Analysis Tool
Potential interest in policy analysis tool, Opposed to use
Challenges
Unfamiliarity with foreign privacy regulations, Understanding existing data flows, Interdisciplinary Collaboration Friction
No Challenges
Reliance on Platform Review, Clear process
Behavior Changes
Increased Privacy Awareness, Privacy Feature Marketing, User Driven Privacy Review
Recommendations
Information presentation, Privacy advocacy, workflow support

E Pre-Interview Survey

The following questionnaire was published to participants via Qualtrics prior to interviews. Dynamic fields marked as `[AppName]` were populated based on the specific app associated with the developer's recruitment link.

Part 1: Consent and Recording

Q1. Do you consent to participate in this study?
- ◯ Yes, I consent to participate
- ◯ No, I decline to participate

Q2. Do you agree to being audio recorded in the interview?
- ◯ Yes
- ◯ No

Part 2: App and Experience

Q3. Is the app [AppName] available on other platforms (e.g., Google Play Store)?
- ◯ Yes
- ◯ No

◯ I don't know

Q4. Approximately how many years of app development experience do you have?
◯ I don't have any development experience
◯ less than a year
◯ 1-2 years
◯ 2-3 years
◯ 3-4 years
◯ 4-5 years
◯ Other: _____

Q5. What best describes your role in app development? (Select all that apply)
◯ I work full-time in app development
◯ I work part-time in app development
◯ I develop apps as a hobby or side project
◯ I develop apps as part of academic or research work
◯ Other: _____

Q6. What is your role in the development of the app [AppName]?
◯ I worked independently as a solo developer
◯ I was part of a team of developers
◯ I hired others / worked with contractors or freelancers
◯ Other: _____

Q7. What is your employment status with the company that distribute the app [AppName]? (Select all that apply)
◯ I am a full-time employee
◯ I am a freelancer / contractor
◯ I run my own company or am self-employed
◯ I am a student
◯ I am not employed (e.g., hobbyist, independent)
◯ Other: _____

Q8. Approximately how many people are in the company that developed [AppName]?
◯ 1-4
◯ 5-9
◯ 10-19
◯ 20-49
◯ 50-99
◯ 100-249
◯ 250-499
◯ 500-999
◯ 1,000 or more

Part 3: Demographics

Q9. What is your gender?
◯ Male
◯ Female
◯ Non-binary
◯ Prefer not to answer
◯ Prefer to self-describe: _____

Q10. What is your age group?
◯ 18-24
◯ 25-34
◯ 35-44
◯ 45-54
◯ 55-64
◯ 65 or older
◯ Prefer not to answer

Q11. What is the highest degree or level of school you have completed?
◯ No schooling completed
◯ Some high school, no diploma
◯ High school graduate, diploma, or equivalent (e.g., GED, Abitur, baccalaureat)
◯ Some college credit, no degree
◯ Trade / technical / vocational training
◯ Associate's degree
◯ Bachelor's degree
◯ Master's degree
◯ Professional degree
◯ Doctorate degree
◯ Prefer not to disclose
◯ Other (please specify)

Q12. Which of the following best describes your educational background or job field?
◯ I have an education in, or work in, the field of computer science, computer engineering or IT.
◯ I do not have an education in, nor do I work in, the field of computer science, computer engineering or IT.
◯ Prefer not to disclose

Q13. In which country do you currently reside?
Select Country ▼

Q14. Have you scheduled an interview?
◯ Yes, I have scheduled an interview via the provided link
◯ No, I will reach out via email to ask for additional dates

F Interview Guideline

Below, we list the questions we used for our semi-structured interviews.

(1) How long have you been working in app development.
(2) What were your motivations for developing this app.
(3) How many apps have you developed and/or contributed to.
(4) What was your role in the development of this app [appname].
(5) How many people were involved in the development and distribution of this app.
(6) Was this app developed for your company/organization or for a client.
(7) Describe the decision process that led to changing/choosing the privacy labels for your app.
(8) Who is responsible for making these privacy labels changes in this app.
(9) How do you decide what to include in the privacy labels and what not to.
(10) How would you describe the process of changing/setting the privacy labels.
(11) What challenges did you face when updating/setting your app's privacy labels.
(12) Is there a process you follow to evaluate the accuracy of your app's privacy labels.
(13) Do you have the same process for updating/setting privacy labels across all apps you work on, or does it vary.

(14) Is this app available on other platforms.
(15) If yes, Do you try to keep the privacy labels and data safety sections consistent across platforms? Like the Apple Store and Google Play.
(16) Who usually handles updating/setting the privacy policy of this app.
(17) How well does your app's privacy policy align with the privacy label.
(18) How often do you refer to your app's privacy policy when updating/setting the privacy labels.
(19) Do you think privacy labels and the privacy policies have the same purposes.
(20) Did changing the privacy labels have any impact on your data practices.
(21) Looking back, is there anything you would do differently when updating your app's privacy label.
(22) Do you find privacy labels useful for you as a developer or app distributor.
(23) Do you find privacy labels useful for your customers.
(24) Do you think users use privacy labels.
(25) How do you think privacy labels affect how users perceive your app.
(26) Do you think privacy labels updates should be communicated to current users.
(27) Do you think your app's privacy labels actually reflect the app's data practices.
(28) Have you ever been unsure whether your privacy label was accurate.
(29) How confident are you that your app's current privacy label is accurate.
(30) What are your thoughts on this label [show the analysis of a text segment of the app's privacy policy] that was suggested by our analysis of your app's privacy policy?
(31) Do you think the interpretation of the policy here is accurate.
(32) If you had this full analysis of the privacy policy would you use it.
(33) In your opinion, what could make privacy labels more helpful for users or developers.
(34) What recommendations do you have for improving privacy labels.

Table 7: Metadata characteristics of the 128 apps that responded to our emails (Changed Privacy Labels and Unchanged Privacy Labels)

Metadata		Changed	Unchanged
App Category	Games	13	2
	Education	10	2
	Health & Fitness	8	6
	Productivity	8	1
	Utilities	6	2
	Weather	4	1
	Entertainment	4	3
	Other categories	34	24
Release Year	2009–2012	4	2
	2013–2016	10	6
	2017–2020	25	17
	2021–2022	24	7
	2023–2024	24	9
Popularity (Ratings)	0	11	6
	1–100	52	25
	101–1,000	14	6
	1,001–10,000	6	3
	10,000+	4	1
Monetization	Free apps	80	32
	Paid apps	7	9
Total		87	41

Table 8: Interview Participants' Demographics.

ID	Dev. Exp.	Location	App Cat.	# Rating	Purpose	Team Size	Role(s) in Team	App Privacy Label
P1	7-8 years	Poland	Weather	10k-20k	Hobby	3	Developer	NL
P2	3-4 years	USA	Lifestyle	< 100	Hobby	1	Solo developer	NL, L, T
P3	4-5 years	USA	Games	< 100	Hobby	1	Solo developer	NL, T
P4	3-4 years	UK	Education	< 100	Hobby	1	Solo developer	DNC
P5	1-2 years	France	Food & Drink	< 100	Hobby	1	Solo developer	L
P6	10 years	Germany	Utilities	< 100	Hobby	1	Solo developer	DNC
P7	4-5 years	Australia	Shopping	100-500	Hobby	1	Solo developer	NL
P8	11 years	Austria	Productivity	100-500	Job	6	Founder	L
P9	> 7 years	Georgia	Finance	< 100	Hobby	1	Solo developer	DNC
P10	> 10 years	USA	Utilities	< 100	Hobby	1	Solo developer	NL
P11	4-5 years	USA	Reference	< 100	Hobby	15	Founder, Designer	NL, L, T
P12	0 years	USA	Health & Fitness	1k-10k	Job	> 2 (outsourced dev. team)	CFO	NL, L
P13	20 years	Norway	Games	< 100	Hobby	1	Solo developer	NL, L, T
P14*	6 years	Japan	Photo & Video	< 100	Hobby	1	Solo developer	DNC
P15*	10 years	USA	Health & Fitness	100-500	Job	> 20	Director of technology	L, T
P16*	> 15 years	Japan	Photo & Video	< 100	Job	1	Solo developer	DNC
P17*	14 years	USA	Photo & Video	< 100	Hobby	1	Solo developer	DNC
P18*	12 years	USA	Photo & Video	1k-10k	Hobby	1	Solo developer	DNC
P19*	5 years	Poland	Health & Fitness	< 100	Hobby	1	Solo developer	DNC

The asterisk (*) denotes participants who did not change their apps' privacy labels.
Column descriptions: participant's development experience (Dev. Exp.), participant's geographic location (Location), app categories (App Cat.), app rating count (Rating Count), app's profit motive (Purpose), app development team size (Team Size), and participant's role(s) in the team (Role(s) In Team).
App Privacy Label indicates the current disclosed data collection practices for each app. DNC = Data Not Collected; L = Data Linked to You; NL = Data Not Linked to You; T = Data Used to Track You. DNC is mutually exclusive with all other categories, whereas L, NL, and T are non-exclusive and may co-occur.