The Overview of Privacy Labels and their Compatibility with Privacy Policies

Rishabh Khandelwal[*]**, Asmit Nayak**[*]**, Paul Chung, Kassem Fawaz**

University of Wisconsin – Madison

1 Introduction

Privacy policies have traditionally been the primary means of conveying the privacy practices of a service to users. However, studies have shown that privacy policies are often ineffective due to readability and reachability issues, as users tend to avoid reading them due to their length and vagueness [12, 21]. Introduced by Kelly et al [25], the concept of privacy labels has gained traction in the tech industry, with Google introducing Data Safety Sections (DSS) and Apple introducing Apple Privacy Labels (APL) for all new and updated apps on the App Store.

Researchers have shown the benefit of privacy labels for users, making privacy practices more accessible [50]. However, prior work has also shown that inaccurate labels can exist due to the developer's knowledge gaps or resource limitations [31]. Incorrect privacy labels can cause confusion and harm users by creating a false sense of security. Furthermore, inaccurate privacy labels can mislead users into downloading and using insecure apps, increasing their privacy risks. Therefore, it is crucial to investigate the accuracy and compliance of privacy labels in real-world scenarios, in order to determine how well they align with the actual data practices of apps.

Xiao et al [48] proposed a methodology to check for consistency of privacy labels by comparing practices in labels with privacy practices inferred by analyzing the dataflow using dynamic analysis. One major limitation of flow-to-consistency analysis is that it requires dynamic analysis of apps which is hard to scale, as pointed out by Xiao et al [48].

In this work, we check the consistency of privacy labels using different approaches – by comparing privacy practices reported in privacy labels with those present in privacy policies. We also compare privacy labels of the same apps across different platforms to gain an understanding of how developers report their apps' privacy practices. A major advantage of our approach is that using automated analysis, it can scale to a large number of apps. Thus, this paper aims to provide a comprehensive analysis of the current state of privacy labels and identify areas for improvement by asking the following research questions:

- What practices are developers reporting in privacy labels? How do these practices evolve over time?

- How do the privacy practices present in privacy labels compare with the privacy policies?

- Do apps have different practices across platforms?

---

[*]Equal Contribution

To answer the above questions, we conduct a large-scale analysis of the privacy labels of apps listed on the Google Play Store and Apple App Store. We also conduct a developer study with android developers to understand their data safety section and highlight the challenges faced by them while working with privacy labels. Our analysis includes a comparison of the privacy practices mentioned in the privacy labels with those present in the privacy policies, as well as a comparison of the privacy practices across apps cross-listed on both platforms.

We first start by developing a scraper for the Google Play Store and Apple App Store to collect metadata for over 2.5M apps on the Play Store and 1.3M apps on the App Store. We also periodically collected metadata for apps on the play store to track any changes made to an app's description, privacy policies, and data safety section. In addition to collecting each app's metadata, we also scraped privacy policies for apps on both platforms. Next, we automatically analyzed the privacy policies to extract privacy practices by developing a privacy label-centric taxonomy by adapting an existing privacy policy taxonomy. Specifically, we added missing elements and added more annotations for the new taxonomy. We then compare these extracted practices with those present in the privacy label to perform a consistency analysis. Finally, we curate a dataset with apps cross-listed on both platforms and compare the privacy labels to understand how consistent developers are in disclosing their practices via privacy labels.

With this work, we make the following contributions:

- We perform large-scale measurements of privacy practices reported in privacy labels across two major platforms - App store (n=1.38M) and Google Play Store(n=2.4M). We filter out apps with less than 1000 downloads for Google Play Store. This limits the number of apps on the Google Play Store to 1.14M. We find that only 50.2% of the apps provide privacy labels on the Google Play Store, whereas on the App Store, only 69.2% of the apps contain privacy labels.

- We perform a longitudinal analysis for privacy labels on the Google play store and study the evolution of Data Safety Forms before and after the hard deadline imposed by Google. We find that app developers have changed data safety forms frequently.

- We compare the data practices mentioned in the privacy policy with privacy labels for apps in app store and google play store and find that on play store, at least 40% of the apps have inconsistencies.

- We also identify 165K apps cross listed on both the platforms and compare how the practices are reported. Surprisingly, we find that privacy labels for 51.5% of the apps are not consistent across the different platforms.

- We provide first large scale datasets for privacy labels for Android (n=1.14M) and iOS (n=1.3M). Further, we also release the new dataset for the newly formed privacy centric taxonomy. Finally, we release a large policy dataset annotated with the privacy centric taxonomy. The datasets will be available after publication.

2   Background and Related Works

Privacy Nutrition Labels. Originally introduced by Kelley et al. [25,26], privacy nutrition labels aim to summarize the privacy practices of websites in a nutrition label format for better visual comprehension. They later designed the "Privacy Facts" display to allow the users to consider privacy while installing apps [27]. More recently, researchers proposed an Internet of Things (IoT)
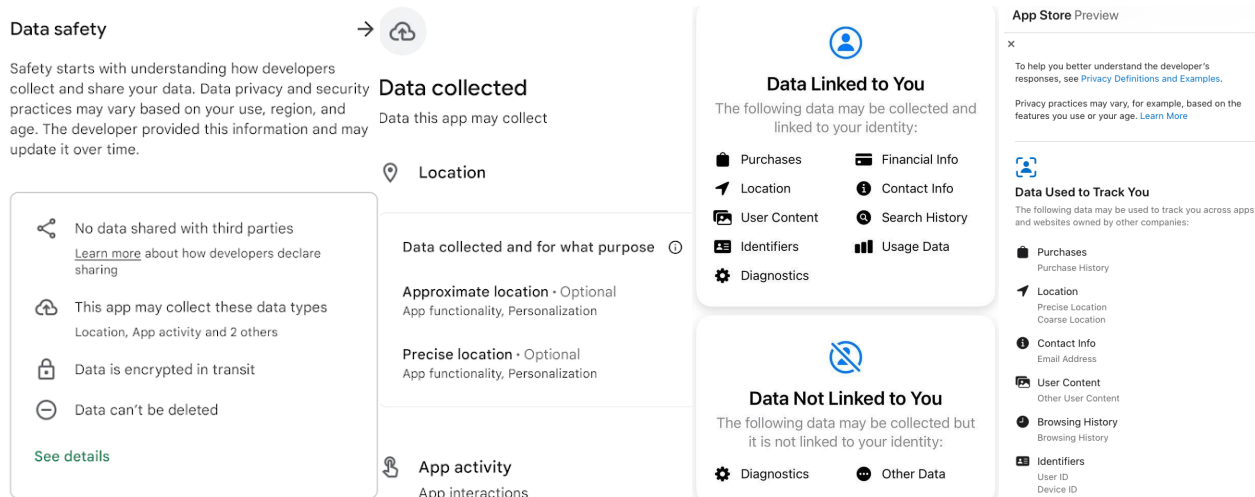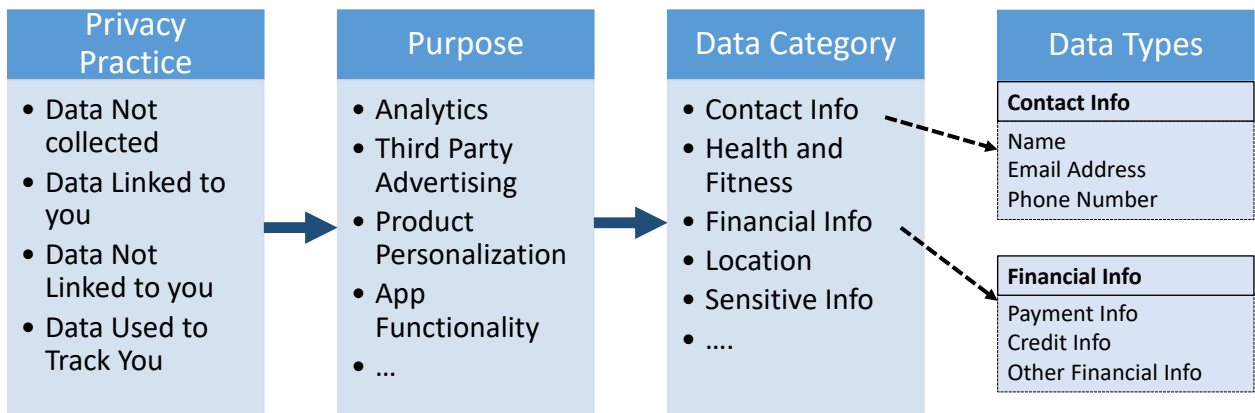
Figure 1: Illustrative Example of nutrition labels

Figure 2: The hierarchy of Apple Privacy Labels

security and privacy label [16,17] to surface privacy and security information related to IoT devices to the users. Researchers have also studied the design and evaluation of privacy notices and labels [10, 13, 14, 19, 25–28, 35, 41].

In December 2020, Apple adopted the privacy nutrition labels for the app store and mandated that app developers provide their apps' privacy information in the form of the Apple Privacy Label (APL). More recently, Google also required app developers to add a Data Safety Section (DSS) on the Google Play Store.

Apple Privacy Label. The Apple Privacy Label (APL) is a four-level hierarchy (as shown in Fig. 2). The top level consists of four high-level privacy practices, known as *Privacy Types*. The second level of the label discusses the purpose for data usage, while the third and fourth level describes high-level *Data Categories* and fine-grained *Datatypes*, respectively. In the top level, *No Data Collected* denotes that the app does not collect any data from the users.

Among the other three categories, *Data used to Track you* covers the practices when user data is linked with third-party data for targeted advertising, Ad measurement, or sharing with a data broker. Notably, tracking does not apply when the data is never sent off the device in *a way that can identify the user or device*, or if the data is used for fraud detection. *Data linked to you* covers the personal information and data that is linked to the user's identity as opposed to *Data not linked*
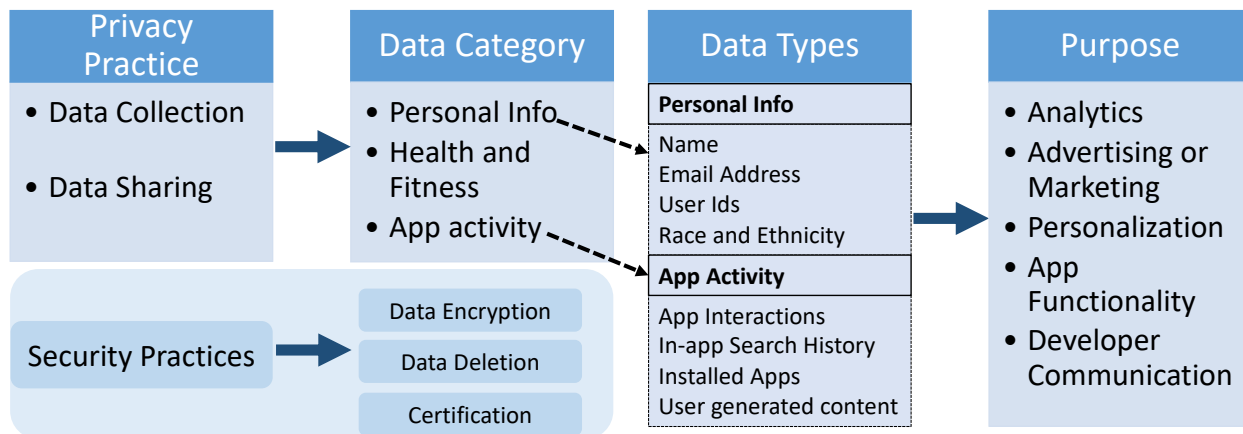
3

Figure 3: Google Data Safety Section

to you.

The next level describes the purposes for which data collected in *Data linked to you* and *Data not linked to you* may be used. Apple defines five main purposes: *Third party advertising and marketing*, *Developers' advertising and marketing*, *Analytics*, *Product Personalization*, *App Functionality* and *Other Purposes*. It is important to note that *Data Used to Track you* does not get a purpose level as its purpose is to track the users. In the *Data Categories* level, Apple defines 14 categories of data such as *Contact Info* (consisting of personal information), *Health and Fitness*, *Financial Info* etc. *Data Categories* consists of the final level - *DataTypes* which consists of 32 fine-grained datatypes that the developers can use, such as *App Interactions, Precise Location, Contacts, Phone* etc. An illustrative example of APL is shown in Fig. 1.

Google Data Safety Section The Data Safety Section (DSS) also consists of four levels, where the first is high level *Privacy Practices*. The second and third levels consist of *Data Categories* and *Data Types*, and the fourth level consists of *Purpose*.

The first level includes three practices: *Data Collection*, which covers the details about the data that is collected and its intended use; *Data Sharing*, where the developers disclose what data is shared with third parties; and *Security Practices* that covers the data practices related to user choice and data security. *Security Practices* include three tags: *Encrypted in Transit, Data Deletion Option*, and *Review against Global Security Standards.*

In the second level, *Data Categories* includes 14 categories such as App Info and Performance and App Activity. Each *Data Category* can also have *Data Types*, which provide fine-grained information about the data used by the app. For example, *App Activity* includes *App Interactions and Installed App*, as shown in Fig. 3. The final level of the Data Safety Section consists of *Purposes* that describe the reasons for collecting or sharing the data.

We note that even though the two privacy labels (APL and DSS) have some overlap at the lowest level, they cover different high-level practices. For instance, APL focuses on surfacing tracking practices and the linkability of the data. DSS focuses on data-centric practices, including collection, sharing, encryption, and deletion. In the rest of the paper, we will use APL and DSS to denote privacy labels for iOS apps and android apps, respectively. Further, we use the term *Privacy Labels* to refer to both APL and DSS collectively.

Usability of Privacy Labels. Researchers have studied the usability of APLs from both users' [50] and developers' [31] perspectives. Zhang et al. [50] studied 24 iPhone users to understand their experiences, understanding, and perceptions of privacy labels on the app store. They uncovered that users find the labels confusing with unfamiliar terms. From the developers' perspective, Li

et al. [31] interviewed 12 iOS developers and reported that the sources of errors by developers in privacy labels included both under-reporting and over-reporting data collection. They further concluded that the label design is generally confusing for the developers either due to known factors (lack of resources, improper documentation) or unknown factors (preconceptions, knowledge gaps). More recently, researchers also built and evaluated a tool [20] that helps iOS developers generate privacy labels by identifying data flows through code analysis. While these works focus on the usability evaluation of APL, our work compares the privacy practices present in privacy policies and labels.

Studies on Privacy Labels. Similar to our work, Xiao et al. [48] characterize non-compliance of apple privacy labels by studying data flow to label consistency of 5K iOS apps. They also provide insights for improving label design. This work is complementary to ours as we measure the consistency of privacy labels with the data practices mentioned in the apps' privacy policies.

The works most similar to ours perform longitudinal measurement of privacy labels to understand the adoption and evolution of apple privacy labels over time [8, 31, 42]. In particular, Scoccia et al. [42] conducted an empirical study of 17K apps to characterize how sensitive data is collected and shared for iOS apps. They found that free apps collect more sensitive data for tracking purposes. Li et al. [31] and Balash et al. [8] collected weekly snapshots of apple privacy labels and characterized the privacy practices mentioned in privacy labels for 573k apps. Balash et al. [8] also perform additional correlation analysis with app meta-data like user rating, content rating, and app size.

Our work is different in two ways. First, we provide complimentary analysis by analyzing privacy labels from Apple and Google to provide a comprehensive understanding of practices mentioned in APL and DSS. In doing so, we also verify their findings on how sensitive data is being collected and used. Second, we perform a consistency analysis of privacy labels with privacy policies. We also create a dataset with cross-listed apps on both platforms to understand how developers disclose their practices on different platforms. To the best of our knowledge, ours is the first work performing this analysis.

Automated Privacy Policy Analysis. In 2016, Wilson et al. [47] introduced a privacy policy taxonomy along with an annotated dataset (OPP-115). The taxonomy covers privacy practices mentioned in the privacy policies of the websites. In the past few years, several works have trained classifiers using the taxonomy for automated policy analysis [22, 36, 43, 45, 47]. Researchers have also used automated policy analysis to check for consistency within the policy [5, 6], as well as consistency with the code [51, 52]. Finally, automated analysis has also been used to study the impact of law and regulations on privacy policy [33, 49]. In this work, we extend the OPP-115 taxonomy to cover practices from nutrition labels and develop new classifiers to extract relevant privacy practices from the policies.

3 Data Collection Pipeline

We show an overview of the data collection pipeline in Fig. 4. We begin by scraping the metadata and privacy labels for the apps from Google Play and Apple App Store (Section 3.1). We then design a classification pipeline to automatically annotate the policies of the mobile apps (Section 3.2). Finally, we identify cross-listed apps between Google Play and Apple App Store (Section 3.3).

3.1 Privacy Labels

First, we describe the collection method for our privacy labels (both DSS and APL) datasets.
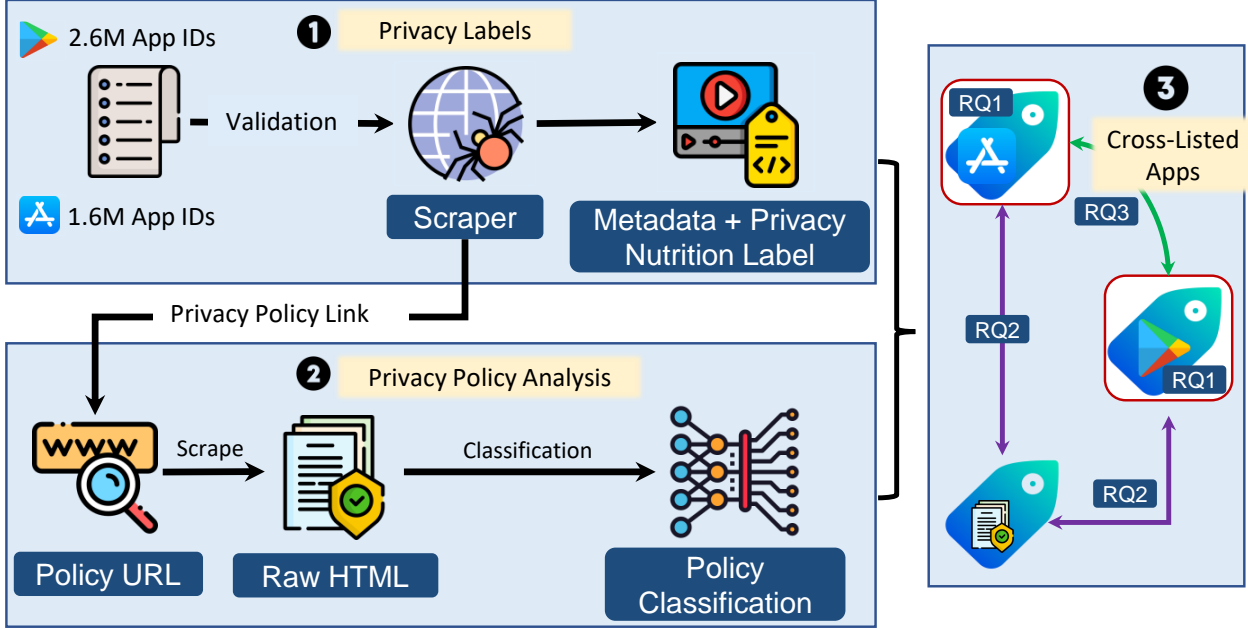
Figure 4: Overview of the data collection pipeline. RQs here refer to the *Research Questions* introduced in Section 1

Google Data Safety Section. We collected 10 snapshots of the Data Safety Sections for 2.6M apps present on the play store between June 20, 2022, and Nov 25, 2022. Google required app developers to complete the data safety section by July 20, 2022. By collecting data before and after this date, we are able to capture how the app developers responded to Google's requirement for adding a data safety section to their apps. Note that we captured weekly snapshots from June 20 to Aug 1, which includes the date set by Google. The remaining two snapshots were taken on Sept 9 and Nov 25.

To collect the data safety section, we start with the apk list provided by Androzoo [4]. This daily updated list consists of up-to-date Android app ids from various sources, including those from the Google Play store. Using the app ids and a customized version of publicly available google play store scraper library `google-play-scraper` [2], we capture the metadata of each app, including its data safety sections and the link to the privacy policy. We used four local machines to perform the scraping. The total time to retrieve data for 2.6M apps, from Google Play, is between 24 to 48 hours. We note that this set also includes apps with very low download counts. To ensure that our statistical analysis is not skewed by these apps, we filter out apps that have fewer than 1000 downloads resulting in a total of 1.14M apps with 573k having privacy labels. We refer to this dataset as **DSS Dataset**.

Performing the longitudinal analysis, we find that during the period of June 20, 2022, to November 25, 2022, the number of apps with DSS increased from 28.76% to 47.71%. The largest change was observed between July 13, 2022, and July 26, 2022, when the percentage of apps went from 35% to 37.4% in 13 days. Interestingly, we find that 5% of apps removed their DSS over the course of our data collection. Specifically, 27K updated their play store page to remove the privacy label. Off these apps, 8.6K apps were deleted, including 1.3K apps which had over 100K downloads. For example, *Sport Prediction* with 1M app download had a DSS as of August 1, 2022, but did not have it by Nov 25, 2022. Furthermore, we find that 1.3K apps updated their DSS to reflect a change in their practices *i.e.* they updated DSS to add or remove privacy practices. We investigate the factors responsible for these changes in Section 4.3.

Apple Privacy Labels. In this work, we do not perform a longitudinal analysis for APLs as Balash et al [8] performed a similar study earlier in 2022. Instead, we collect a single snapshot of the Apple Privacy Labels (APLs) on November 13, 2022, for consistency. To curate the dataset, we begin by parsing the XML site map for the app store.[1] Using the URLs from the sitemap, we use the Apple Store Catalogue API to extract the metadata for each app, including the privacy nutrition label and a link to the privacy policy. We performed the crawl using 11 instances of google cloud functions to scrape 1.6M apps in 15 hours.

We extracted information for 1.38M apps out of the 1.6M apps available, filtering out those with non-English content. As a result, we obtained 955K (69.2%) with APLs. In comparison, Balash et al [8] in March 2022 found that 60.5% of apps had Apple Privacy Labels. The higher percentage in our study suggests that new APLs are still being added to apps. We refer to this dataset as **APL Dataset**.

3.2 Privacy Policy Analysis

We build a privacy policy analysis pipeline to automatically annotate the privacy policies of the apps (the second component in Fig. 4). This annotation allows us to analyze the consistency between the privacy labels and privacy policies of each app at a scale (Section 5).

Text Extraction and Cleaning. Starting with the privacy policy URL, we crawl the corresponding webpage, clean the HTML by removing the headers and footers, and extract the text using the `BeautifulSoup` library [39]. We use the `PyPDF2` [18] library to extract text from privacy policies that are PDF documents. Following prior works [22], we apply three exclusion criteria. First, we filter out the instances where the text length is less than 100 words. Second, we filter out the instances where the policy is stored in non-standard formats, such as images. Finally, as our analysis pipeline relies on the English language, we filter out non-English policies (330K for Play Store, 230K for App Store) using the language detection library `polyglot` [3].

Policy Classification. To extract the practices from privacy policies, we follow the pipelines from existing works [22,33] that use OPP-115 taxonomy [47]. However, in our case, not all practices mentioned in privacy labels can be extracted as the underlying taxonomy either lacks these classes or the datasets do not have sufficient samples. For example, a data category used in APL, *Sensitive Info* is missing from the taxonomy. To overcome this limitation, we extend the existing taxonomy by adding missing classes. Further, two of the authors perform annotation for the added classes. The annotators had an overlapping set of 200 segments to measure the inter-annotator agreement. We find that the annotators showed high agreement with a Cohen's Kappa value of ($\kappa = .85$). We provide more details about the taxonomy and the annotation in Appendix A.1.

Following prior works [43,45], we use DistilBERT [40] to train the classification models. We then use these models to extract privacy practices from policies. This approach allows us to accurately extract the practices from privacy policies, even for those practices that were not covered by the existing taxonomy. Table 1 shows the performance of our classifier on a held-out test set. We note that our classifiers outperform previous classifiers, primarily because we added new annotations.

3.3 Identifying Cross-Listed Apps

We next describe the process to identify cross-listed apps across the two platforms ((3) in Fig. 4). We use the resulting dataset to compare the privacy labels of apps across the two platforms (Section 6). Identifying two versions of the same app across platforms is challenging due to the lack of unique identifiers [24].

---

[1]We used the `ultimate-sitemap-parser` library.

Category	CNN [37]	BERT [15]	Ours
First-party-collection-share	82	91	**98**
Third-party-sharing-collection	81	90	**96**
Identifiability	77	91	**97**
Does-does-not	86	93	**96**
Encryption-in-transit	N/A	N/A	**99**
Data Deletion Option	N/A	N/A	**91**

Table 1: Selected Classifiers' performance on the test set. For the performance of all classifiers, please see Table 3 in Appendix A.1.

To uniquely map apps across platforms, we develop a heuristic based on combinations of pseudo-identifiers, such as the app name, developer name, privacy policy, and developer website. We start with the apps which have the same name across both platforms (n=220K). Next, if the privacy policy of the apps matches, then we treat them as a unique match (n=85K). In some cases, like the *NTLC Catalog* app on the Play Store and App Store, the app developers can include platform-specific identifiers in the URLs for privacy policies. To capture these instances, we match the first level domain of the privacy policy URLs and identify them as unique matches (n=54K). Finally, while providing privacy policy links is highly encouraged in both platforms, some apps do not contain the link to the privacy policy. To further increase the coverage, we also match the first-level domain of the developer website, which is present on both platforms. Using these criteria, we are further able to get 25K matches. This way, we obtain a total of 165K apps that have instances in both Apple play store and Google play store.

Manual Verification: To assess whether our heuristic results in false positive matches, two of the authors manually verified 150 app pairs identified using each of the three heuristics and found that no app from Google Play Store was matched to an incorrect app from App Store. It is worth noting that for our analysis, having an accurately mapped set is more important than capturing all instances of cross-listed apps.

Cross-listed Apps Dataset Using the method described above, we find a total of 165K cross-listed apps. Among these apps, we find that 5% have privacy nutrition labels only on the Google Play Store, 20.2% have the label only on the Apple App store, 60.8% have labels on both the platforms and 13.9% do not have a privacy nutrition label on either platform. The higher rate of privacy labels for the App store can be understood as Apple enforced nutrition labels on their platform earlier than Google, giving more time for developers to add the details in the APL.

3.4 Ethical Considerations

We collected data only from publicly available web pages and APIs. While our data collection scripts might load Google and Apple's servers, we were careful to not abuse these resources. In particular, we added back-off strategies in case of errors and waited for sufficient time before retrying for the failed cases. Furthermore, for privacy policy extraction, we were respectful of robots.txt and only extracted HTML when the website allowed us to.

3.5 Observations

Our measurement pipeline results in two initial observations. First, app developers have been slow to add privacy labels to their apps, even after the hard deadlines have passed. Privacy labels are present only for 69% of the apps on the Apple app store and 50.2% of the apps on the Google

play store (as of November 2022). Second, our measurement pipeline produced large-scale datasets for Apple Privacy Label (n=955K), Google Data Safety Section (n=573K), and privacy policies (n=598K) corresponding to the apps. In addition, we generated a new *Privacy Label Taxonomy* by adding missing elements to OPP-115 taxonomy. We also supplement the existing privacy policy datasets by adding annotations for the new categories.

4 Data Practices in Privacy Labels

4.1 Google Data Safety Section

In this section, we analyze the DSS dataset (Section 3.1) comprising 573K apps. We first discuss the practices present in DSS and then examine how these practices vary with an age rating, price, and popularity.

Data Collection and Sharing: Among the apps having DSS, we saw 42.3% collecting at least one type of data, and 35.8% sharing at least one data type (purple bars in the top plot for Fig. 5). This suggests that the majority of the apps on the play store report do not collect or share data. This is in contrast with the findings from prior work [46] that found that the majority of the apps use at least one third-party application, which has been shown to collect sensitive information [11,32]. One possible explanation for this is that developers find it hard to understand the collection and sharing practices of third-party libraries. This is also supported by prior research [9,31]. As such, when inquired about change in DSS, one developer also alluded to lack of transparency by third parties:

"We don't collect or share any user data. But we use Meta (former Facebook) audience network for monetizing non-paying users with ads. Unfortunately, the details provided by Meta are very vague.."

We also note that among the apps not collecting any data around 23% are sharing data. This is because *Data Collection* is defined as the instance when the developers retrieve the data from the device using the app [7], whereas *Data Sharing* is defined as when the data is transferred from the device to a third party. This way, the developers can share data without collecting it if the application uses third-party libraries which directly send data to their servers.

Security Practices: We find that 23% of the apps do not provide any details of their security practices. 65% of the apps encrypt data that they collect or share while it's in transit, and 42% allow the users to request that their data be deleted or automatically anonymize data within 90 days. Notably, we find that 17.4% of the apps state that they do not collect or share data, but encrypt the data in transit. We explore this behavior further in Section 4.3. As apps need network permissions to transmit data, we cross-verified encryption practices with apps' network permission requests and find that 10.5% apps do request network permission but do not encrypt data, potentially exposing user data in plain text. Additionally, 2.2% of apps do not request network permissions, yet state that they encrypt data in transit, suggesting that some developers might be over-reporting their practices, consistent with prior research [31].

Category and Purpose Level Practices: In Fig. 6, we present the top-5 data categories for *Data Collection* and *Data Sharing* by apps in play store. A full plot including all data categories can be found in Fig. 13 (Appendix). Our findings indicate that the data categories *Personal Information* and *App activity* are among the most frequently collected, and are primarily used for *App functionality* and *Analytics*. However, *Location* and *Device Ids* are more commonly shared for the purpose of *Advertising or Marketing*. We emphasize that this flow poses serious privacy risks and allows for tracking by third parties. We also observe that sensitive data types such as
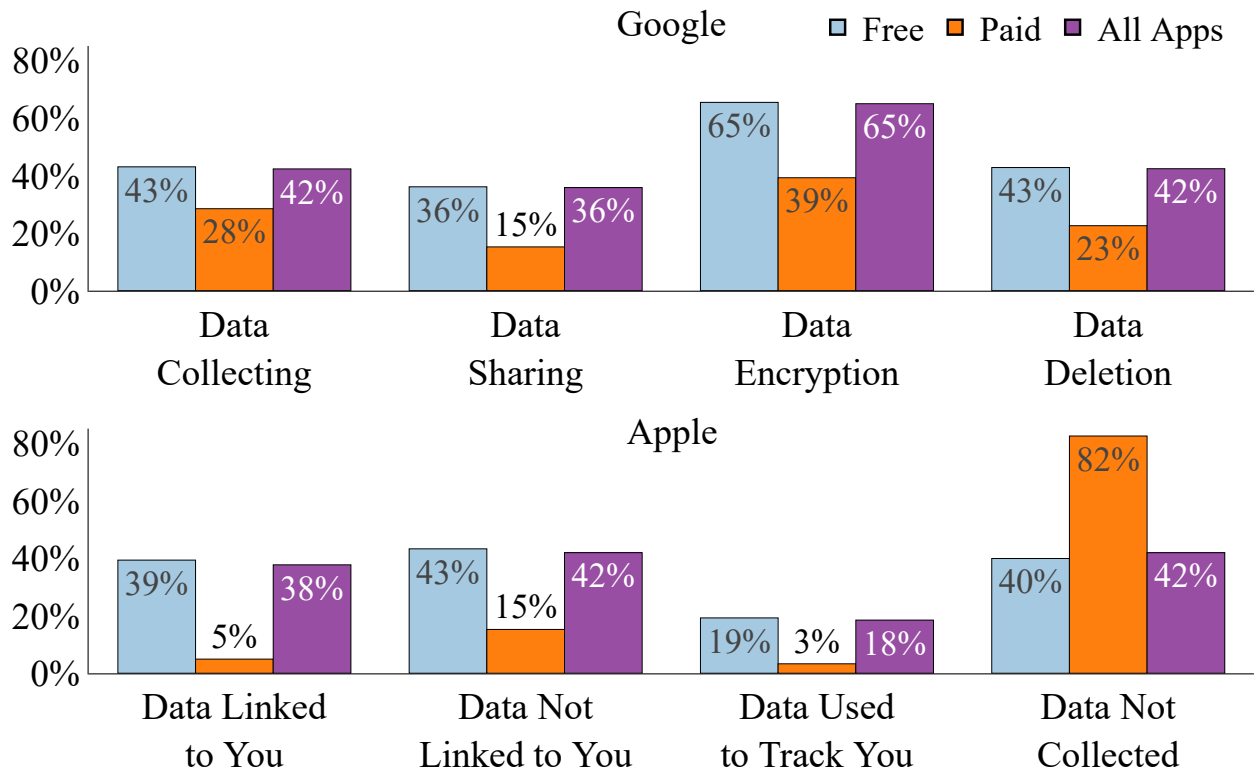
9

Figure 5: Distribution of privacy types in Google Data Safety Sections and Apple Privacy Labels. The normalization is done by the total number of apps with privacy labels.
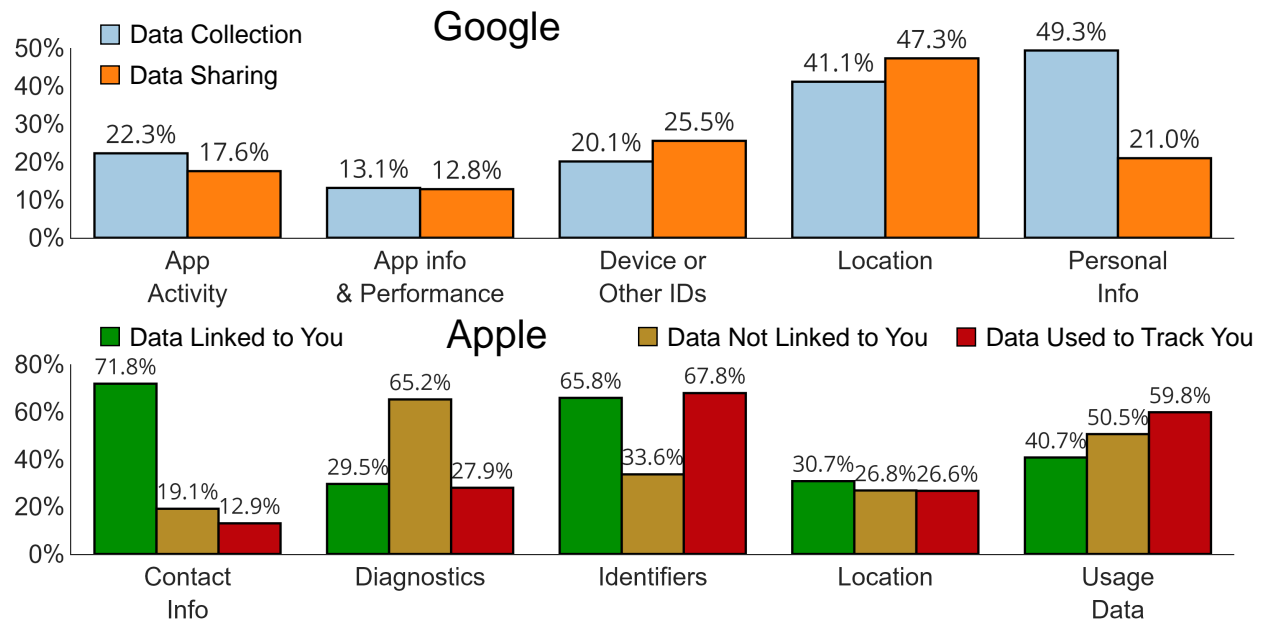

Figure 6: Distribution of Top-5 data categories for high-level practices for apps in Play Store (top) and App store (bottom). The normalization is done by the total number of apps with privacy labels. For plots with data categories, see Fig. 13 in the Appendix.
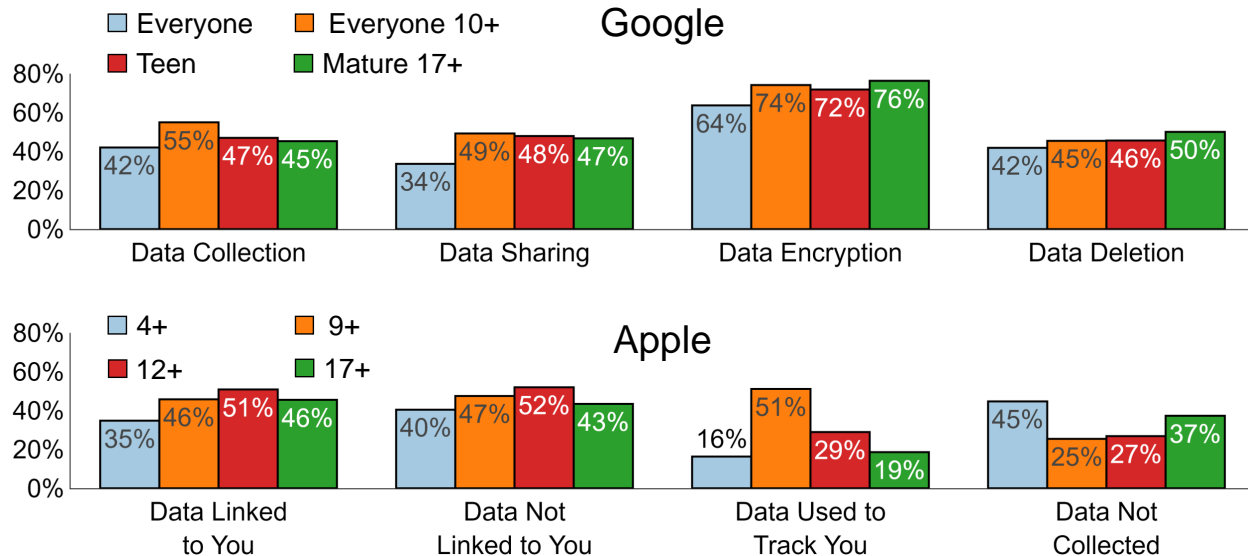
Figure 7: Distribution of privacy types based on age rating for DSS and APLs. The normalization is done by the total number of apps with privacy labels.

Audio, *Files and Docs*, and *Health and Fitness* are collected less frequently, with the most common purpose being *App functionality*. Furthermore, we note that out of the 7 possible purposes for collecting data there are over 4K apps that list 6 or more purposes for the data they collect, which may indicate that app developers list all purposes out of convenience. For example, *Workplace from Meta* with over 15M+ downloads, lists the same 6 purposes for all the data they collect like access to *Installed Apps*, *SMS or MMS*, *Music Files*. This is consistent with the findings of Li et al. [31], who suggest that developers may over-report in cases of ambiguity.

Variation of Practices with Popularity: We first investigate the relationship between privacy practices and app popularity. We classify apps into three categories based on their number of downloads: extremely popular (greater than 1M download, n=56K), semi-popular (more than 10K downloads, n=524K), and low-popular (less than 10K downloads, n=621K). Our findings reveal that 1) the fraction of apps displaying Data Safety Sections (DSS) increases with the popularity of the apps (42% for low-popular, 51% for semi-popular and 76% for extremely popular) and 2) the fraction of apps collecting and sharing data is less for popular apps (41% for low-popular, 46% for semi-popular and 12% for extremely popular). These results suggest that developers from popular apps tend to report more privacy-friendly practices.

Variation of Practices with Age Rating: Next, we examine how the practices of apps differ based on their age rating as determined by the Google Play Store. The Play Store assigns five different age ratings: Everyone, Teen, Mature 17+, and Everyone 10+[2]. We acknowledge the importance of this distinction, as apps that are accessible to children and teens (falling in the Everyone and Teen categories) are expected to have higher transparency and collect less data. However, our analysis of the dataset reveals that 59% of apps with the *Mature 17+* rating have a Data Safety Section (DSS), while the fraction of apps with a DSS in the other age ratings ranges from 47% (Everyone) to 55% (Everyone 10+). The data practices for different age ratings are shown in Fig. 7. We find that the fraction of apps having *Data Collection* and *Data Sharing* is lowest for apps rated for *Everyone*, whereas apps targeting *Mature 17+* have the highest encryption

---

[2]Google also has Adults 18+ rating, but we found less than 200 apps in this category and decided to filter it out for this analysis

rate.

Variation of Practices with Price: Finally, we study the difference in practices based on whether the app is available for free, free with in-app purchases, or paid. We find that 68% of the paid apps have DSS whereas, for free apps, only 46% have DSS. Fig. 5 shows the distribution of high-level practices with free and paid apps. We note that for paid apps, a fraction of apps collecting and sharing data is lower. Furthermore, apps with *Data Encryption* and *Data Deletion* are lower because the apps are collecting and sharing fewer data. This suggests that paid apps tend to have better data practices.

4.2   Apple Privacy Labels

Next, we examine the Apple Privacy Label (APL) dataset (Section 3.1) consisting of privacy labels from 955K apps. We first discuss the practices present in APL and then dive into variations of practices with an age rating and price. Finally, we conclude by comparing the low-level practices mentioned in APL and DSS

High-Level Practices: In our dataset, 42% of apps collected data from users that were not linked back to the user (Data Not Linked to You), whereas 37% of apps did collect data that is linked to the user (Fig. 5. Note that apps could collect multiple types of data some of which may be linked to the users while others may not. Furthermore, around 18% of the apps reported collecting data that was used to track the users. Note that this reflects the status of the APLs after the *Apple Tracking Transparency* policy was implemented, which requires developers to obtain consent from users before tracking. We also find that 42% of apps report that they do not collect any data from users. Recent works [29, 30] analyzing iOS apps have found that at least 80% of the apps still use tracking libraries in the apps. Further, these libraries have been shown to collect user data [11, 32]. Similar to the case of android developers, this discrepancy can be explained by the lack of transparency of privacy practices by the third-party libraries, resulting in confusion for the developers.

For *Data Used to Track You*, we find that *Usage Data* and *Identifiers* are most commonly used. We note here that Apple defines *Tracking* as when data collected is linked with third-party data for targeted advertising, as well as when the data is shared with a data broker. Additionally, we observe that 25% of the apps collecting *Location* information also use it for tracking. This poses severe privacy risks to the users as entities can track the physical location of the users which can reveal sensitive details about users' habits and routines.

Data Category and Purpose Level Practices: In Fig. 6 (bottom), we show the top-6 data categories mentioned in the high-level practices in the APL dataset. We find that for *Data Linked to You*, *Contact Information* and *Identifiers* are collected most frequently, whereas for *Data Not Linked to You*, *Diagnostics* and *Usage Data* are collected most frequently. Apple defines *Contact Information* as name, email, phone number, and physical address, whereas *Usage Data* refers to product interactions and advertising data such as information about the ads that the user has viewed. Analyzing purposes for these data categories, we find that nearly 60% of the apps use these data categories for *App functionality* and *Analytics*. It is also worthwhile to note that *Contact Information* is used for *Advertisements* in only 8% of the apps that collect this information, indicating that apps generally do not use personal information for advertisements. We also note that *Identifiers*, commonly used for tracking users for targeted advertising is used for *Advertisement or Marketing* in more than 20% of the apps that collect *Identifiers*. Interestingly, *Location*, under *Data Linked to You* is also used for *Advertisement or Marketing* by 20% of the apps that collect *Location*.

Variation of Practices with Age Rating Next, we investigate the correlation between the privacy practices described in the Android Permission List (APL) and the age rating and price of apps. The App Store assigns four different age ratings: 4+, 9+, 12+, and 17+ (which roughly align with the rating system used by the Google Play Store). Our analysis reveals that the fraction of apps with an age rating of 17+ is highest at 76%. However, we note that the high-level data practices, shown in Figure 1, are consistently more privacy-friendly for apps with lower age ratings. For instance, only 13% of the apps with an age rating of 4+ track users. Similarly, data collection for these apps is also consistently lower than that of other categories.

Variation of Practices with Price: Finally, we categorize the dataset into free and paid apps and examine the differences in privacy labels. Recall that for the play store, we observed that paid apps contained more DSS than free apps. For APL, we find the reverse trend with 70% of the free apps having APL as compared to 52% paid apps. On the other hand, the high-level practices are decidedly better for the paid apps, as shown in Fig. 5 (bottom chart). For instance, 82% of the paid apps reported not collecting any data, while only 3% of paid apps mentioned using data to track the user. This indicates that the paid apps on iOS platforms are more friendly than the free apps.

Comparison Between DSS and APL: As discussed in Section 2, DSS and APL provide different information to the users, and cannot be directly compared based on high-level practices. However, since the underlying data collected is the same, we can compare the practices shown in Fig. 6. We observe that the fraction of apps requesting similar datatypes is much smaller for apps on the play store than that of the app store (with a notable exception of Location). This can be attributed to the fact that developers have had a longer to work with the APL framework, while the DSS framework is still relatively new. In our communication with app developers, one app developer mentioned that they try different answers on the data safety form. We also received communication indicating that some developers updated their DSS based on the questions that we had asked. This indicates that the developers are unclear on the process involved in the data safety forms, which might result in some inaccuracies in the DSSs. This is also supported by the study conducted by Li. et al [32] where they find that app developers find it difficult and challenging to fill out the privacy labels, especially, because the frameworks for Apple and Google are starkly different and can create confusion.

4.3 Developer Study

In Section 3.1 and Section 4.1, we identified three trends in Data Safety Sections: (A) apps stating that they encrypt data without collecting or sharing data, (B) apps changing their practice from not collecting/sharing data to collecting/sharing data, and (C) apps changing their practices from collecting/sharing data to not collecting/sharing data. To gain a deeper understanding of these trends, we reached out to developers via email and asked them one general question about their Data Safety Sections and one specific question about the type of trend we observed in their app. We contacted 30K developers from the Play Store. It is worth noting that, since Apple does not provide email addresses for developers, we only conducted this study with Android developers.

In our initial email, we clearly identified ourselves as researchers and stated that we were studying their application and wanted information regarding their data safety section (Appendix C). Additionally, we do not collect any personally identifiable information from the developers and only use their publicly available contact information from the Play Store to contact them. As such, the study has been approved by the IRB at our institute.

Findings: Based on our initial emails, we received 2500 responses. After filtering out the auto-

mated replies using keyword filtering, we were left with 889 responses where the app developers describe the challenges they face while working with the privacy labels, as well as provided information about their data safety section. We further manually examined each response and curated a set of 307 replies. This manual filtering removed replies that included non-relevant replies. Next, one of the authors manually coded the responses to identify the factors for the trends, as well as general challenges described by the developers. Another author independently verified the findings by coding a subset of 50 responses independently. Specifically, we first present the major contributing factors for the different types of trends mentioned above. We then discuss the top challenges that developers face while working with the Data Safety Section.

Type A: Apps Stating that they encrypt data without collecting or sharing data: For this trend, we obtained responses from 165 developers. Of these, 56% mentioned data is collected by third-party services like ads or Google Firebase but were not sure if it should be added to DSS while another 36% were not sure what data was collected, 3% of the developers were confused regarding encryption and added the option thinking of SSL encryption for communications between the server and the app, without collecting/sharing data. For example, one of the developers said the following: *"I use Google's own libraries for this. In the Google Play Console, Policies section, I had to guess that Google is sending data and I rely on Google to encrypt that data. Because Google says that the developer is responsible for the libraries they use. That's why you find the contradictory result."*

Type B: Apps changing their practice from not collecting/ sharing to collecting/sharing: For this trend, we received responses from 130 developers, 12% of whom did not understand the process and selected any option that was accepted whereas 74% changed DSS after realizing that third party libraries are collecting data. 12% of them had an app update while 2% changed DSS to ensure that they were up to date with the regulations like GDPR. For example, one developer said, *"… Admob SDK I am integrating with the app might collect information […] And According to Google policy, if I am using the latest version of their Admob SDK, I have to specify that the app is collecting or sharing data …"*

Type C: Apps changing their practice from collecting/sharing to not collecting/sharing: We only obtained 12 responses for this trend, 58% of which stated that their app was updated, but the DSS reflects an older version, 25% mentioned that data was collected by ad libraries that have since been removed, and 9% mentioned regulations as a factor for the change in DSS. For example, one developer said, *"…we have changed the data safety section of our application because we […] removed any data collecting libraries such as Firebase […] Admob for monetization…"*

Challenges for Developers: We find that the developers are generally confused about how to fill the Data Safety Section. The source of confusion varies from *Not understanding the Process* to *Not understanding if data collected by the third party should be reflected in DSS*. For example, one developer stated *"What […] keeps changing every few months is Google's privacy policies. They are difficult to understand and they shift like sand… I don't really understand half of them and so we just keep submitting answers in hopes it's what they are looking for…"* indicating that the process is very unclear, while another mentioned *"the reason for the change was because google play forced me to put that information"*. These confusions are problematic as they may result in inaccurate privacy labels. They can also under-represent the privacy practices in the privacy labels which can give a false sense of security to the users, increasing their privacy risks.

We note that in an earlier qualitative study, Li et al. [31] found that "Developers felt unconcerned about privacy and that it was not their responsibility". In our study, we found that developers cared about user privacy, but did not have enough means (either lack of resources or lack of transparency with third-party libraries' privacy practices) to create accurate labels, causing some frustration on

their part. For example, one developer said *"... we use Meta (former Facebook) audience network for monetizing non-paying users with ads. Unfortunately, the details provided by Meta are very vague, but definitely are considered as collecting and sharing data. If possible we would love to switch to an ad provider that offers proper non-personalized ads with zero/minimal data collection, but it seems impossible to find such a provider.".*

4.4 Takeaways

The analysis presented here results in three main takeaways: 1) Privacy practices reported in the privacy nutrition labels differ from the privacy practices derived using app analysis by prior works [46]. Specifically, prior works have shown that third-party libraries are used in the majority of the apps and that these libraries collect sensitive information from the users. This is inconsistent with what we find in the privacy labels. This inconsistency can be explained by the fact that privacy practices of third-party libraries are often vague and create confusion among the developers (consistent with findings from literature [31]). 2) We also show that paid apps, and apps that are open to all age groups, including children, are more privacy-friendly. As shown in Fig. 5 and Fig. 7, these apps are less likely to engage in tracking, data collection, and data sharing. 3) Fig. 6 also shows that location data is often used for advertising, marketing, and tracking. This poses severe privacy risks, as location data can reveal sensitive information about an individual's habits and routines. Our research suggests that further attention should be paid to the use of location data in mobile apps, and the potential risks it poses to user privacy.

5 Practices Present in Privacy Policies

The next research question that we answer is: *How do the privacy practices mentioned in the privacy labels of the apps compare with the privacy practices described in their privacy policies?* We perform this comparison by training machine learning classifiers to automatically extract privacy practices mentioned in the privacy labels, as described in Section 3.2. For Google's DSS we have 346K apps with valid policies, whereas for Apple, we have 343K apps. As described in Section 3.2, we filter out the policies which are not in English. Note that to obtain presence of a particular practice in privacy policy, we require that there exists at least one segment which classifies the segment for that practice. For example, for a policy to have *Data Encryption* practice, we require presence of at least one segment where our classifier tags it as positive for *Data Encryption*. A complete mapping from classifiers to practices in APL/DSS is described in Table 2.

There are two types of inconsistencies that can arise: 1) *In Label*, where a given privacy practice is mentioned in the privacy label but is absent from the privacy policy, and 2) *In Policy*, where a practice is found in privacy policy but is missing from privacy label. As privacy policies can potentially cover multiple applications, websites and products, *In Policy* inconsistency does not necessarily mean that policy is inconsistent with the privacy label. For example, the Google app *Clock* reports that it does not collect or share any *Location* information. However, since Google has one policy to cover all the products, the policy states that they can collect *Location* (applicable in Google Maps). In such cases, it is inaccurate to say that privacy label are inconsistent with the privacy policy without further analyses. However, if an app mentions collection of data and the policy does not mention it, then we can conclusively say that the privacy policy and the app are inconsistent. Thus, in this work, we will focus primarily on *In Label* inconsistencies, except when there is a negative practice is involved (*Data Not Collected* or *Data Not Linked to You*). This is because if the policy says that data is not collected, then no app corresponding to that policy should collect any data, and in this case we focus on *In Policy* consistency.
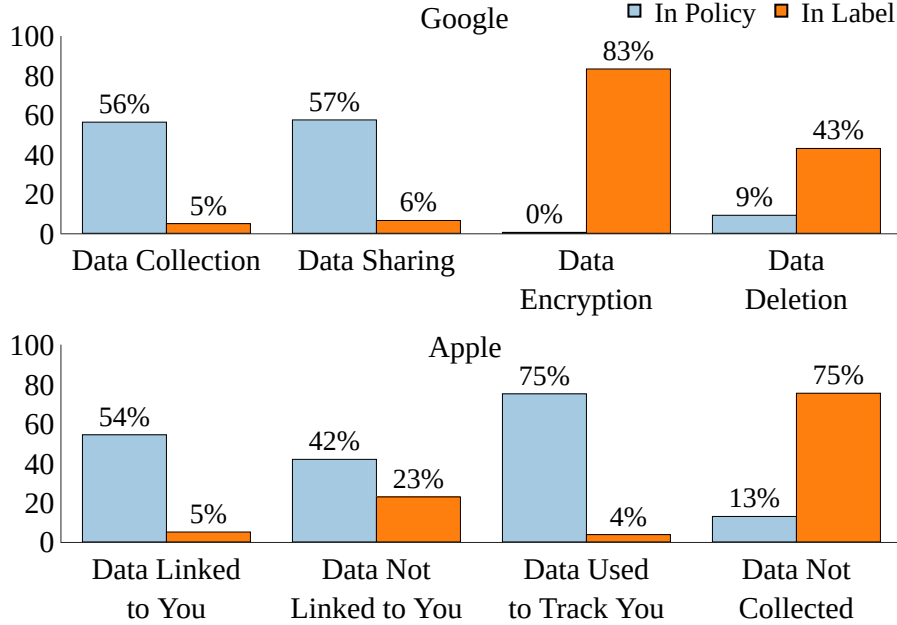
Figure 8: Inconsistencies between privacy policies and DSS and APL. The normalization is based on the total number of apps with privacy labels and classified policies.

5.1 Google Data Safety Sections

Fig. 8 shows the *In Policy* and *In Label* inconsistency for high level practices for apps on the play store. We note that only 5% (6%) of the apps with DSS that collect (share) data have *In Label* inconsistency with their privacy policies. We also find that *In Policy* inconsistencies for these categories are more than 55%, but as discussed above, these could be due to privacy policy covering multiple apps and websites. To understand the extent to which this happens due to multiple apps, we analyze DSS for apps from the same developers. There are 15,380 developers who have 3 or more apps. These developers have an average of 13 apps and a median of 7 apps per developer. We find that 68%(10,420) of these developers have duplicate data safety section for their apps. For example, the app developer *Premium Software* has over 9 apps across 6 genres but with only 2 unique DSS. This also highlights that developers might be duplicating their DSS across their apps, even though the apps can span multiple genres and have different features.

Analyzing the inconsistencies for *Data Encryption* and *Data Deletion*, we find that the majority of apps declare them in their privacy labels but there is no mention of such practices in their privacy policies. For example, *Snapchat* mentions in their DSS that data is encrypted in transit but no corresponding practice is present in their privacy policy. Similarly, *Kik — Messaging & Chat App* state in their DSS that *Data can't be deleted* yet their privacy policy states that users can ask them to delete their information. It is worth noting that from a privacy and regulation standpoint, these two practices are extremely important. *Data Deletion* option gives the users the right to either delete their data or ensure that it stays in anonymized form, which has roots in several regulations such as the GDPR [33] and the CCPA [1]. *Data Encryption* on the other hand, is crucial to prevent data snooping attacks which aim to get unlawful access to the data while the data is in transit.

Analyzing practices at the category level, we find that there are significant inconsistencies between Privacy Labels and policies for data sharing and data collection. Specifically, we find that for data sharing, 89% of the apps have inconsistent (*In Label*) Privacy Labels for *Location*, 82% of the apps for *Device IDs*, and 74% of the apps for *Health and Fitness*. For example, *Myntra - Fashion*

16

Shopping App states that they collect location, health info, contact list, and much more, yet its privacy policy doesn't mention the collection or sharing of such data types. Similarly, *Tripadvisor: Plan & Book Trips* states that they collect and share location data yet there is no mention of such practices in their privacy policy. This suggests that developers report more precise data-sharing practices in Privacy Labels, and can inform users allowing them to make better choices.

5.2 Apple Privacy Label

Next, we analyze the apps on the App store and compare privacy practices mentioned in the apps' privacy policies and their Apple Privacy Labels. The bottom plot in Fig. 8 shows inconsistencies for the high-level categories present in APL. We find that for *Data Linked to You* and *Data Used to Track You*, the *In Label* inconsistency is 5% and 4% respectively. As the other two of the high-level practices, *Data Not Linked To You* and *Data Not Collected* are negations, we consider the *In Policy* inconsistency (see Section 5). We find that 42% of the apps have policies that state that they do not link data whereas the privacy label indicates otherwise. Furthermore, 13% of the apps have policies that do not have data collection or sharing, but the privacy label indicates otherwise. For example. *Superior Vision* app on App Store states that they collect *Health & Fitness* data yet their policy doesn't state that.

In Fig. 8, we observe that *In Policy* inconsistency for *Data Used to Track you* is very high. This implies that privacy policies include tracking practices while privacy labels do not. This can potentially be due to the presence of segments related to cookies in the privacy policy, for example. *Netflix*'s privacy policy talks about using cookies to track users on their site but not about tracking via their app.

We next examine the consistency at the data category level for *Data Linked to You* and find that, similar to DSS, *Location* (39%), *Identifiers (51%)* and *Health and Fitness* (60%) had the largest *In Label* inconsistencies. For *Data Not Linked to You*, we find inconsistencies primarily in the same data categories. For example, *Jetpack Joyride* states that they collect Location data but their privacy policy states that they collect Location based on IP Address and not the GPS location.

5.3 Takeaways

In this section, we find that at least 40% of the apps with DSS, and APL are inconsistent with their privacy policy. Additionally, we note that DSSs contain more information about *Security practices* than privacy policies, and thus can provide useful information to the users. We also note that sensitive datatypes such as *Location*, *Identifiers* and *Health and Fitness* had the largest *In Label* inconsistencies, indicating that the developers disclose collection/sharing of these fine-grained datatypes in the privacy labels.

6 Data Practices Across Platforms

Next, we compare the self-reported privacy practices in Privacy Labels of cross-listed apps across the Google play store and Apple app store using the cross-listed dataset described in Section 3.3.

6.1 Mapping DSS categories to APL categories

As previously discussed in Sec. 2, the privacy labels for android and iOS platforms cover different aspects of data practices. APL emphasizes on tracking and linkability of the collected data without
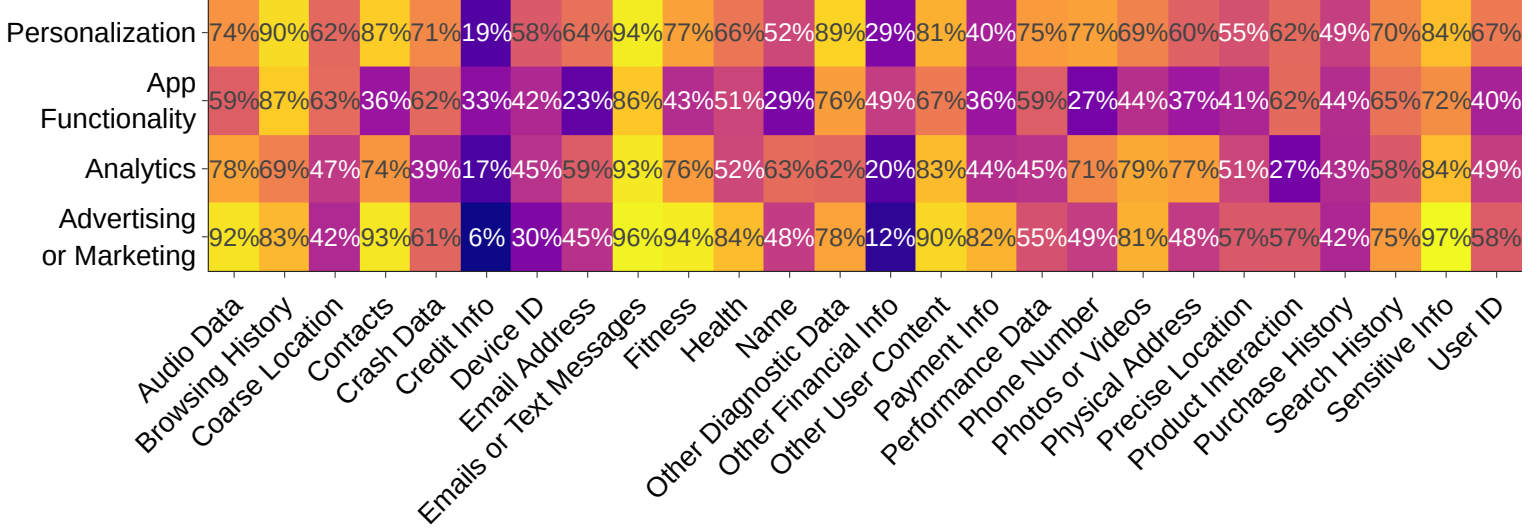
Figure 9: Normalized Heatmap showing the inconsistencies in the datatype-purpose pair. Normalization is done for each cell block in the heatmap, *i.e.*, for each datatype-purpose pair, we normalize with the total number of apps which have that datatype-purpose pair.

distinguishing between collected and shared data, while DSS focuses on security practices and whether the data is collected or shared with a third party. However, despite covering disjoint high-level practices, the lower-level attributes in the privacy labels - namely the `datatype` and `purpose` - have a large overlap. Thus, to compare the disclosure practices of app developers across platforms, we first find the common datatypes and purposes in the two labels, and then compare 1) the datatypes, and 2) datatype-purpose pairs.

It is worth noting that the datatype and purpose tags used in the two labels can be used to denote different concepts. For example, in APL, `App functionality` also includes fraud prevention and implementing security measures, whereas the data safety section has separate tags for app functionality and fraud prevention and security measures. For the purposes of this analysis, we combine these two purposes into *App functionality* to create a common map. Since APL does not have any tags for *Account Management* and *Dev. Communicaitons*, we removed them from DSS for this comparison. After taking the intersection of the available datatypes and purposes, we end up with 4 purpose categories and 26 datatypes. A complete mapping for classes from DSS to APL is shown in Appendix E.

6.2 Findings

We compare the self-reported privacy practices of the 100K apps that are cross-listed on both the platforms and have privacy labels. Specifically, we ask the following questions: a) How does the high-level practice of data collection compare between the two labels? and b) Is the purpose for using datatypes consistent between the two labels?

6.2.1 Comparison of Data Collection

To compare how many apps do not collect data, we rely on the *Data Not Collected* tag for the iOS platform and *Data Shared* and *Data Collected* tags for the android platform. We find that a
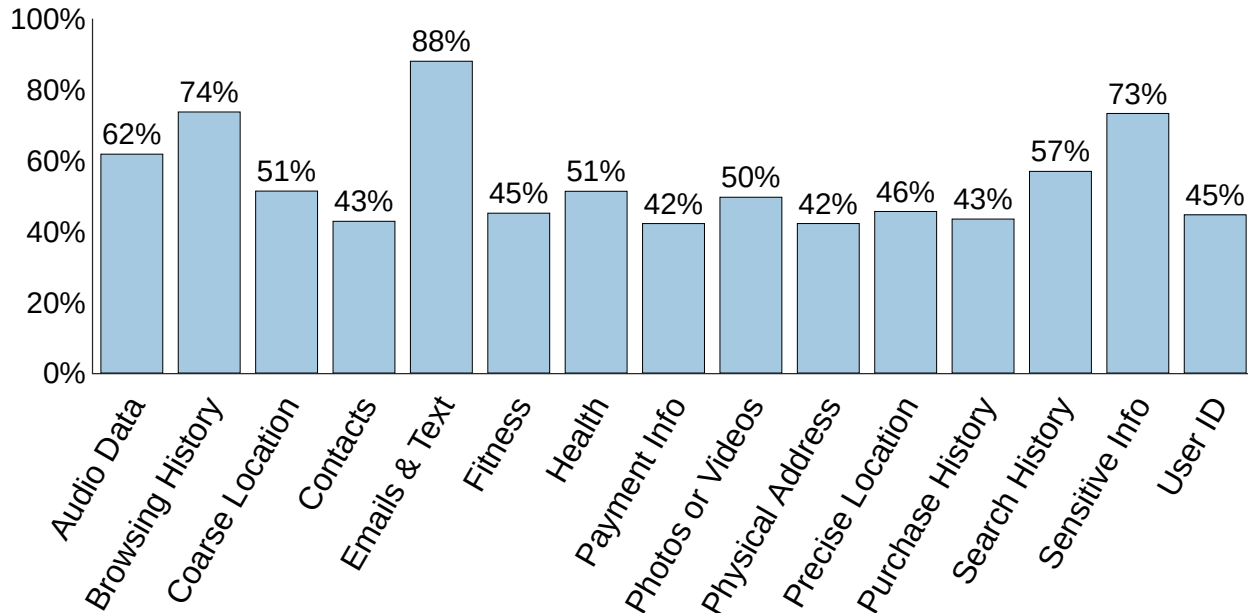
18

Figure 10: Distribution of inconsistent apps with datatypes. Each datatype is normalized with the number of apps using that particular datatype on either platform. Note that we have omitted some datatypes here for brevity. The full distribution can be found in the Fig. 13 in the Appendix.

total of 22K(22%) apps report different data collection practices on the two platforms. Of these apps, 42% of the apps report collecting data on android while 58% of the apps report collecting data on the iOS platform. Examining these apps further, we find that 18% of these apps have more than 100k downloads, and 5% has over 1M downloads indicating that even popular apps have this inconsistency. For example, *KineMaster - Video Editor* a video editing app with over 400M+ downloads on Google Play Store states that they do not collect any data in the Play store but states in App Store that they do collect sensitive data such as *Location* and *Identifiers*.

The inconsistency in self-reported data collection practices as indicated by the inaccuracies in privacy labels undermines the credibility of the Privacy Label framework. This poses a significant concern for users, as they may base their decisions on inaccurate information, thereby increasing their privacy and security risks.

6.2.2 Comparison of Fine-grained Practice

We compare fine-grained practices along two dimensions: 1) DataType where we check whether the privacy labels report collecting/sharing the same datatypes; and 2) DataType-Purpose pairs where we compare the common datatype-purpose pair in the two labels. If there is at least one datatype-purpose pair that is not present in both sets, we treat the app as inconsistent for that datatype-purpose pair. We also tag instances as inconsistent where the datatype-purpose pair is present in one of the labels and is missing from the other. For example, if the DSS of an app has *(Location - Personalization)*, while the APL has *(Location - App Functionality)*, then we treat the app as inconsistent. Similarly, if the DSS of an app has *(Location - Personalization)*, while the APL has *(Location - App Functionality)* and *(Location - Personalization)*, we still treat the app as inconsistent as the tags *(Location - App Functionality)* is not common in both.

Across the cross-listed apps that have privacy labels, we find that at least 60% of the apps have at least one inconsistency. For example, in *Tiktok*, DSS states that they collect the contact

list of the users for 'Advertising and Marketing' purposes, but APL states that the app does not collect a contact list. Fig. 10 shows the inconsistency in datatypes across the two platforms. We find that *Sensitive Information*, *Browsing History*, and *Emails or Text Messages* have the highest inconsistencies across the two platforms. From Fig. 10, we observe that DeviceID and Product Interactions are the data counts with the highest inconsistencies. We also observe that *Precise Location* and *Coarse Location* is inconsistent with *Advertising* implying that at least in one of the labels, location is used for advertising, raising privacy concerns for the users.

To analyze datatype-purpose inconsistencies, we show the normalized heatmap of inconsistent apps in Fig. 9. Normalization is done for each cell block in the heatmap, *i.e.*, for each datatype-purpose pair, we normalize with the total number of apps that have that datatype-purpose pair. We find that *Fitness* and *Sensitive Information* when used for *Advertising or Marketing* are frequently inconsistent. The plot shows that even though *Sensitive Information* and *Fitness* data are not collected very often (Fig. 10), when they are collected, they are often inconsistent in privacy labels across two platforms. On the other hand, *Credit Information* and *Financial Information* have the least number of inconsistencies, which is encouraging considering the sensitive nature of this information.

6.3 Takeaway

In this section, we analyzed the consistency of privacy labels for the same apps across the two platforms. We find that 60% of the cross-listed apps had at least one inconsistency between APL and DSS. We further find that inconsistencies are highest for *Sensitive Information*, *Browsing History*, and *Emails or Text Messages* datatypes. Through a detailed analysis of datatype-purpose inconsistencies, we find that *Emails and Text Messages* when used for Advertising results in inconsistencies 96% of the time, indicating a concerning problem with disclosure of practices in privacy labels.

7 Discussion

In this paper, we investigated the consistency of privacy labels with privacy policies and labels on other platforms. Our findings suggest that there is a significant degree of inconsistency in privacy labels. Overall, there is a need for greater consistency in the way that privacy practices are disclosed to users, both within and between platforms. In this section, we discuss the implications of our findings and suggest potential solutions for improving the transparency and consistency of privacy practices. We also discuss the limitations of our study.

Comparison between the two labels. We analyzed both the Data Safety Sections and the Apple Privacy Labels and find that the two labels cover different aspects of data practices. While both labels provide information about the types of data that apps collect, Apple's privacy label does not distinguish between data collection and data sharing. Apple's privacy label is more explicit about certain aspects of data practices, such as linkability, third-party advertising, and tracking, whereas data safety sections lack these details, but does inform the users about the safety of their data (*Data Encryption*) and the choices that they have with developers (*Data Deletion Option*). These practices may be of particular interest to the users in light of the GDPR [33], which requires companies to provide a clear and explicit purpose for the collection and use of personal data. The regulations like the GDPR and the CCPA also provide the right to delete the data to the users, which is covered in Data safety forms but not in the Apple privacy labels.

The comparison between the two labels highlights the importance of considering multiple sources of information when evaluating the data practices of apps. By combining the information provided

by both labels, users can make more informed decisions about their privacy and the apps they choose to use.

Usability of Privacy Policies For Apps. Privacy policies have been used as a default framework for notice and choice to users. Our analysis reveals that many developers have several products, including websites, Internet of Things (IoT) devices, and applications. However, it is common for these products from the same developer to have the same privacy policy, even if they collect data in different ways. This provides inaccurate information, as the privacy policy itself may not accurately convey the privacy practices of a specific product. This can be addressed by having separate privacy policies for each product or by clearly identifying the specific practices that apply to each product within a single privacy policy. Failing to do so may lead to misunderstandings and mistrust among users, and may also violate privacy regulations.

Inconsistencies in disclosed practices across platforms. Our findings indicate that there are inconsistencies between the privacy labels in the Apple Privacy Labels and the Google Data Safety Sections for the same apps. One possible reason for these inconsistencies is the confusing framework for privacy labels. While previous research [31] has shown that privacy labels are useful for both developers and users, it also highlighted that filling privacy labels is perceived as challenging extra work. On top of that, developers are also unclear about definitions which can result in confusion and ultimately, inaccurate privacy labels. This confusion can be compounded by the fact that different platforms may use different terminology to describe similar practices. For example, in Apple's privacy label, the term *tracking* is used when data collected is linked with third-party data for advertising purposes or when data is shared with a third party, which can be confusing to the developers, even when they are asked to pay close attention [31].

Another possible reason for the inconsistencies we observed is the casual attitude of some developers toward disclosing their data practices. Some developers may not fully understand the data practices of their own apps, or may not prioritize accurately disclosing this information to users. Finally, the platforms lack consistency checks to ensure that the information provided in the privacy labels is accurate. Without these checks, it is possible for developers to provide misleading or incomplete information about their data practices, just to meet the requirements.

We note that these inconsistencies can have serious consequences for users, as they may be confused about the privacy practices of the apps they use. If the practices disclosed in the privacy labels are inaccurate, it can reduce the efficacy of these labels as a tool for helping users make informed decisions about their privacy. Even worse, it could induce a false sense of security in users, who may assume that their data is being handled in a certain way when it is not.

Usability of Privacy Labels. Even though our analysis finds inconsistencies between privacy labels and privacy practices, evidence suggests that privacy labels generally carry more specific information about the practices. They include information about the types of data that an app collects, how the data is used, and whether it is shared with third parties. This information can be very useful for users who are concerned about their privacy and want to ensure that they are only using apps that respect their personal data.

However, the accuracy of privacy labels is not guaranteed. While developers are required to disclose their data practices in order to obtain a privacy label, there is no guarantee that the information they provide is accurate or complete. As such, it is important for platforms to recognize that developers may not always be honest about their data practices. Therefore, it is necessary to have systems in place to verify the accuracy of privacy labels and to hold developers accountable for any discrepancies. This is particularly important because the false labels can create a false sense of security among the users.

One potential model for regulating privacy labels is a system similar to the one used for food

nutrition labels, which are regulated by the Food and Drug Administration (FDA). A regulatory body could be established to oversee privacy labels and ensure that they are accurate and consistent. This could help to build trust among users and encourage developers to be more transparent about their data practices.

Limitations. Extracting privacy practices using automated analysis comes with several limitations. First, the framework used here treats privacy policies as segmented text, missing out of relations between different segments. This can potentially result in internal contradictions, as shown in [5]. Second, the classifiers used to extract privacy practices can introduce errors, which can then propagate through the pipeline and induce uncertainty in the inconsistency rates. We do however note that the error rate of our classifiers is significantly less than the inconsistency rate obtained, indicating that the results presented in the paper are valid.

8 Conclusion

In conclusion, our large-scale measurements of Privacy Labels have provided valuable insights into the privacy practices of apps. By analyzing Data Safety Sections for 2.5M apps and Apple Privacy Labels for 1.38M apps, we provided a comprehensive picture of the privacy practices of the applications. On one hand, privacy labels provide users with more specific information about the data practices of apps than traditional privacy policies. However, our analysis showed that there is often a discrepancy between the information disclosed in privacy labels and the information contained in privacy policies. This can be confusing for users and may make it difficult for them to make informed decisions about which apps to use based on their privacy concerns. Furthermore, our comparison of Privacy Labels for cross-listed apps in the Play store and Apple store showed differences in the practices disclosed, indicating that developers are not consistently disclosing the same information on different platforms. Overall, these findings highlight the importance of carefully reviewing Privacy Labels and other sources of information when evaluating the privacy practices of apps. They also suggest that there is a need for improved transparency and accountability in the app industry, as developers may not always be accurately disclosing their data collection and use practices. Having a more transparent system will allow the consumers to be aware of the data collection and use practices of the apps and make informed decisions about their privacy.

References

[1] California consumer privacy act (ccpa), Mar 2022.

[2] JoMingyu/google-play-scraper: Google play scraper for python, 2022.

[3] R. Al-Rfou. Polyglot, 2020.

[4] K. Allix, T. F. Bissyandé, J. Klein, and Y. Le Traon. Androzoo: Collecting millions of android apps for the research community. In *Proceedings of the 13th International Conference on Mining Software Repositories*, MSR '16, pages 468–471, New York, NY, USA, 2016. ACM.

[5] B. Andow, S. Y. Mahmud, W. Wang, J. Whitaker, W. Enck, B. Reaves, K. Singh, and T. Xie. {PolicyLint}: Investigating internal privacy policy contradictions on google play. In *28th USENIX security symposium (USENIX security 19)*, pages 585–602, 2019.

[6] B. Andow, S. Y. Mahmud, J. Whitaker, W. Enck, B. Reaves, K. Singh, and S. Egelman. Actions speak louder than words:{Entity-Sensitive} privacy policy and data flow analysis with

{PoliCheck}. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 985–1002, 2020.

[7] U. app privacy & security practices with Google Play's Data safety section Computer Google Play Help. 2022.

[8] D. G. Balash, M. M. Ali, X. Wu, C. Kanich, and A. J. Aviv. Longitudinal analysis of privacy labels in the apple app store. *arXiv preprint arXiv:2206.02658*, 2022.

[9] R. Balebako, A. Marsh, J. Lin, J. I. Hong, and L. F. Cranor. The privacy and security behaviors of smartphone app developers. 2014.

[10] R. Balebako, F. Schaub, I. Adjerid, A. Acquisti, and L. Cranor. The impact of timing on the salience of smartphone app privacy notices. In *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 63–74, 2015.

[11] T. Book, A. Pridgen, and D. S. Wallach. Longitudinal analysis of android ad library permissions. *arXiv preprint arXiv:1303.0857*, 2013.

[12] F. H. Cate. The limits of notice and choice. *IEEE Security & Privacy*, 8(2):59–62, 2010.

[13] L. F. Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *J. on Telecomm. & High Tech. L.*, 10:273, 2012.

[14] L. F. Cranor. Mobile-app privacy nutrition labels missing key ingredients for success. *Communications of the ACM*, 65(11):26–28, 2022.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[16] P. Emami-Naeini, Y. Agarwal, L. F. Cranor, and H. Hibshi. Ask the experts: What should be on an iot privacy and security label? In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 447–464. IEEE, 2020.

[17] P. Emami-Naeini, J. Dheenadhayalan, Y. Agarwal, and L. F. Cranor. Which privacy and security attributes most impact consumers' risk perception and willingness to purchase iot devices? In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 519–536. IEEE, 2021.

[18] M. Fenniak, M. Stamy, pubpub zz, M. Thoma, M. Peveler, exiledkingcc, and PyPDF2 Contributors. The PyPDF2 library, 2022. See https://pypdf2.readthedocs.io/en/latest/meta/CONTRIBUTORS.html for all contributors.

[19] G. Fox, C. Tonge, T. Lynn, and J. Mooney. Communicating compliance: developing a gdpr privacy label. 2018.

[20] J. Gardner, Y. Feng, K. Reiman, Z. Lin, A. Jain, and N. Sadeh. Helping mobile application developers create accurate privacy labels. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 212–230. IEEE, 2022.

[21] J. Gluck, F. Schaub, A. Friedman, H. Habib, N. Sadeh, L. F. Cranor, and Y. Agarwal. How short is too short? implications of length and framing on the effectiveness of privacy notices. In *Twelfth symposium on usable privacy and security (SOUPS 2016)*, pages 321–340, 2016.

[22] H. Harkous, K. Fawaz, R. Lebret, F. Schaub, K. G. Shin, and K. Aberer. Polisis: Automated analysis and presentation of privacy policies using deep learning. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 531–548, 2018.

[23] H. Harkous, S. T. Peddinti, R. Khandelwal, A. Srivastava, and N. Taft. Hark: A deep learning system for navigating privacy feedback at scale. 2022.

[24] A. Hooda, M. Wallace, K. Jhunjhunwalla, E. Fernandes, and K. Fawaz. Skillfence: A systems approach to practically mitigating voice-based confusion attacks. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(1):1–26, 2022.

[25] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, New York, NY, USA, 2009. Association for Computing Machinery.

[26] P. G. Kelley, L. Cesca, J. Bresee, and L. F. Cranor. Standardizing privacy notices: An online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, page 1573–1582, New York, NY, USA, 2010. Association for Computing Machinery.

[27] P. G. Kelley, L. F. Cranor, and N. Sadeh. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 3393–3402, 2013.

[28] P. G. Kelley, L. F. Cranor, and N. Sadeh. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, page 3393–3402, New York, NY, USA, 2013. Association for Computing Machinery.

[29] K. Kollnig, A. Shuba, R. Binns, M. Van Kleek, and N. Shadbolt. Are iphones really better for privacy? comparative study of ios and android apps. *arXiv preprint arXiv:2109.13722*, 2021.

[30] K. Kollnig, A. Shuba, M. Van Kleek, R. Binns, and N. Shadbolt. Goodbye tracking? impact of ios app tracking transparency and privacy labels. *arXiv preprint arXiv:2204.03556*, 2022.

[31] T. Li, K. Reiman, Y. Agarwal, L. F. Cranor, and J. I. Hong. Understanding challenges for developers to create accurate privacy nutrition labels. In *CHI Conference on Human Factors in Computing Systems*, pages 1–24, 2022.

[32] J. Lin. *Understanding and capturing people's mobile app privacy preferences*. PhD thesis, Carnegie Mellon University, 2013.

[33] T. Linden, R. Khandelwal, H. Harkous, and K. Fawaz. The privacy policy landscape after the gdpr. *arXiv preprint arXiv:1809.08396*, 2018.

[34] B. MacCartney. *Natural language inference*. Stanford University, 2009.

[35] A. M. McDonald, R. W. Reeder, P. G. Kelley, and L. F. Cranor. A comparative study of online privacy policies and formats. In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 37–55. Springer, 2009.

[36] N. Mousavi Nejad, P. Jabat, R. Nedelchev, S. Scerri, and D. Graux. Establishing a strong baseline for privacy policy classification. In *IFIP International Conference on ICT Systems Security and Privacy Protection*, pages 370–383. Springer, 2020.

[37] K. O'Shea and R. Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[38] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67, 2020.

[39] L. Richardson. Beautiful soup documentation. *April*, 2007.

[40] V. Sanh, L. Debut, J. Chaumond, and T. Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.

[41] F. Schaub, R. Balebako, A. L. Durity, and L. F. Cranor. A design space for effective privacy notices. In *Eleventh symposium on usable privacy and security (SOUPS 2015)*, pages 1–17, 2015.

[42] G. L. Scoccia, M. Autili, G. Stilo, and P. Inverardi. An empirical study of privacy labels on the apple ios mobile app store. 2022.

[43] M. Srinath, S. Wilson, and C. L. Giles. Privacy at scale: Introducing the privaseer corpus of web privacy policies. *arXiv preprint arXiv:2004.11131*, 2020.

[44] M. Tkachenko, M. Malyuk, A. Holmanyuk, and N. Liubimov. Label Studio: Data labeling software, 2020-2022. Open source software available from https://github.com/heartexlabs/label-studio.

[45] I. Wagner. Privacy policies across the ages: Content and readability of privacy policies 1996–2021. *arXiv preprint arXiv:2201.08739*, 2022.

[46] H. Wang, Y. Guo, Z. Ma, and X. Chen. Wukong: A scalable and accurate two-phase approach to android app clone detection. In *Proceedings of the 2015 International Symposium on Software Testing and Analysis*, pages 71–82, 2015.

[47] S. Wilson, F. Schaub, A. A. Dara, F. Liu, S. Cherivirala, P. G. Leon, M. S. Andersen, S. Zimmeck, K. M. Sathyendra, N. C. Russell, et al. The creation and analysis of a website privacy policy corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1330–1340, 2016.

[48] Y. Xiao, Z. Li, Y. Qin, J. Guan, X. Bai, X. Liao, and L. Xing. Lalaine: Measuring and characterizing non-compliance of apple privacy labels at scale. *arXiv preprint arXiv:2206.06274*, 2022.

[49] R. N. Zaeem and K. S. Barber. The effect of the gdpr on privacy policies: Recent progress and future promise. *ACM Transactions on Management Information Systems (TMIS)*, 12(1):1–20, 2020.

[50] S. Zhang, Y. Feng, Y. Yao, L. F. Cranor, and N. Sadeh. How usable are ios app privacy labels? *UMBC Faculty Collection*, 2022.

1st Party Collection — Collection Mode, Information Type, Purpose

3rd Party Collection — Action, Information Type, Purpose, Third Party

Choice, Control — Choice Type, Choice Scope

Specific Audiences — Audience group

Data Security — Security Measure

Access, Edit, Delete — Access Rights, Access Type

Data Retention — Retention Period, Purpose, Information Type

Policy Change — Notification Type

Other — Introductory, Contact Information, Practice not covered

How service provider collects data:
• unspecified
• Collect on website
• Collect in mobile app
• Collect on other websites
• Receive from unnamed third-party
• Receive from named third-party
• Collect on mobile website
• ...

Type of collected data:
• unspecified
• financial
• health
• contact
• location
• ...

Why the service provider collects data:
• unspecified
• Basic service
• Additional service
• Personalization
• Analytics/Research
• Marketing
• Advertising
• ...

How third part accesses user data:
• unspecified
• First party shares data
• Uses public information
• Track-on-first-party-website-app

Identity of third party receiving data:
• unspecified
• Unnamed third party
• Named third party
• Public
• Other users
• ...

Aspects of data collection/use the user can control:
• First party use
• First party collection
• Third party use
• Third party sharing/collection
• ...

How can user control data collection:
• Third-party privacy controls
• Opt-in
• First-party privacy controls
• Don't use service
• ...

Controls available to the user's stored data at the service provider:
• Edit-information
• Delete-account
• View
• Export

How long service provider keeps user data:
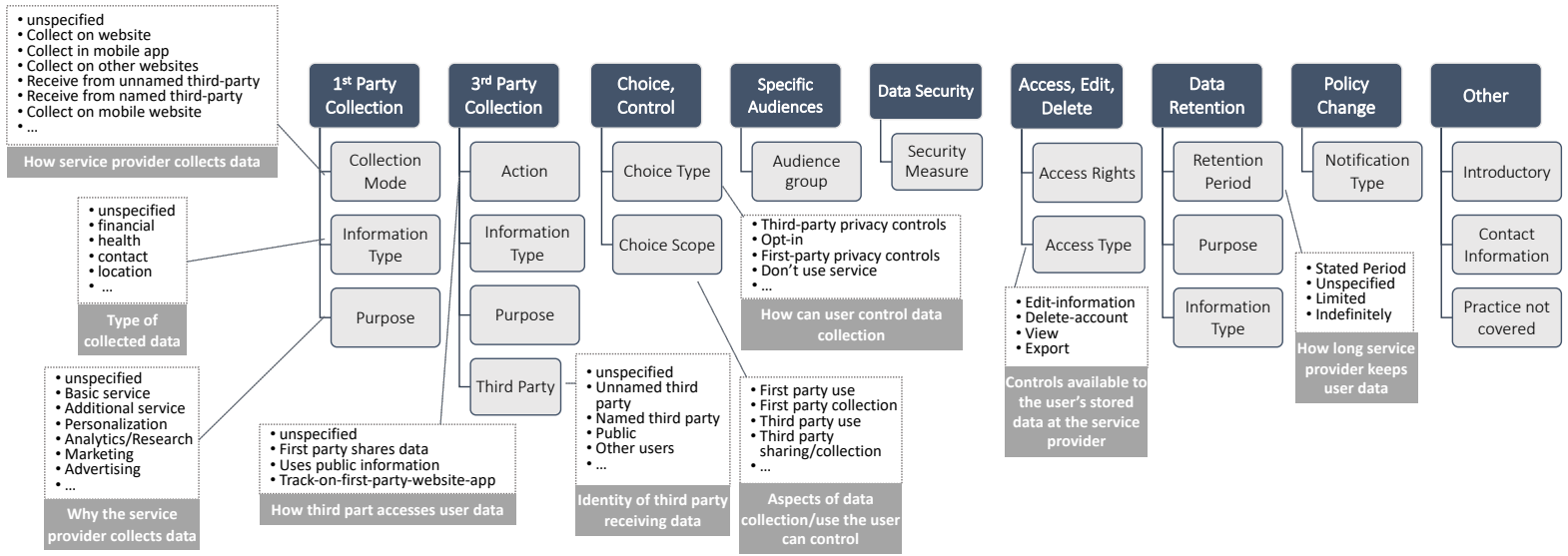• Stated Period
• Unspecified
• Limited
• Indefinitely

Figure 11: The privacy policy taxonomy by Wilson et al. [47]

[51] S. Zimmeck, P. Story, D. Smullen, A. Ravichander, Z. Wang, J. R. Reidenberg, N. C. Russell, and N. Sadeh. Maps: Scaling privacy compliance analysis to a million apps. *Proc. Priv. Enhancing Tech.*, 2019:66, 2019.

[52] S. Zimmeck, Z. Wang, L. Zou, R. Iyengar, B. Liu, F. Schaub, S. Wilson, N. Sadeh, S. Bellovin, and J. Reidenberg. Automated analysis of privacy requirements for mobile apps. In *2016 AAAI Fall Symposium Series*, 2016.

A   Privacy Policy Analysis

A.1   Privacy Policy Taxonomy

Limitations of the OPP-115 Taxonomy Figure. 11 shows the privacy taxonomy proposed by Wilson et al. [47]. The top-level defined high-level privacy categories whereas the lower level defined a set of privacy attributes that can take a particular set of values. Additionally, some examples of attribute-value pairs are shown such as Information Type and Purpose. Note that several lower-level attributes are shared across the high-level categories.

Prior works [22, 43] have used the OPP-115 taxonomy and the associated dataset to build machine-learning classifiers that tag segments of the policy with the labels from the taxonomy. However, there are two limitations to directly using the taxonomy (and existing frameworks such as Polisis [22]) to compare privacy practices between privacy labels and privacy policies. First, the OPP-115 taxonomy was developed for privacy policies of websites, which is vastly different than the ecosystem of applications (both Android and iOS). In particular, the applications have access to sensitive data types, which are present in the privacy labels. This taxonomy, while having some overlap with the APL and DSS privacy labels, does not cover such app-specific data types. For example, `app activity`, a data category covering users' interactions within the application, is not covered in the taxonomy. Second, the OPP-115 dataset has limited annotations for the lower-level attributes that overlap with the private labels. For example, *Encryption in Transit*, which is a separate practice covered in Data Safety Sections, only has less than 100 labeled instances in the OPP-115 dataset.
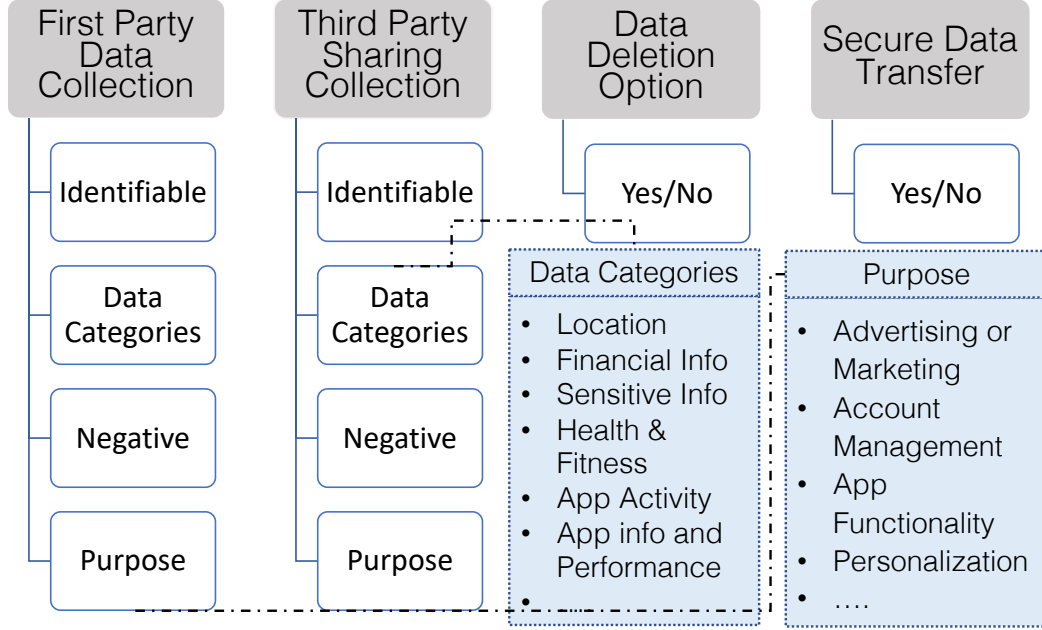
Figure 12: Privacy Label Taxonomy

We address these limitations by incorporating the missing labels to the existing OPP-115 taxonomy. We derive a *Privacy Label Taxonomy* (Fig. 12) as a union of a subset of the original OPP-115 taxonomy with the new labels from APL and DSS. To build the taxonomy, we first identify the categories from the taxonomy that are relevant to privacy labels, thus creating a subset of the original taxonomy. We then add the missing categories to get the new taxonomy.

Identifying relevant categories from OPP-115 As discussed in Sec. 2, privacy labels consist of high-level privacy practices, data categories, and purposes for the use of data. The high-level categories *First-party-data-collection* and *Third-party-sharing-collection* from the OPP-115 taxonomy are relevant as they map directly to Data Safety Sections' *Data Collection* and *Data Sharing* privacy types. Further, APL covers the first-party collection and sharing practices implicitly through *Data Linked to You* and *Data Not Linked to You*. Similarly, the attribute level categories *Purpose*, *Data Type*, and *Identifiable* are relevant.

For example, in Apple Privacy Label (APL), the *Data Used to Track You* privacy type includes the data that is linked with third-party data for targeted advertising. It also includes cases when the data is shared with a data broker. Note here that linking can be done by both the app developers (by using data obtained from a third party) or by sharing the data with a third party. Thus, this privacy practice can be represented with *Advertising or Marketing* purpose of *First-party-collection-use* and *Third-party-sharing*. At this stage, we drop the categories absent in the privacy labels. For example, *Policy Change* is a high-level category in OPP-115 which is not present in the *Privacy Label Taxonomy*.

Adding New Categories As indicated earlier, the OPP-115 taxonomy misses some of the lower-level data categories and purposes. We add these missing elements and adapt the OPP-115 taxonomy to *Privacy Label Taxonomy*.

Apart from the high level categories from the taxonomy, we also add two high level categories: *Data Deletion Option* and *Encryption in Transit*. Both the categories are part of *Security Practices* privacy type from DSS. *Data Deletion* corresponds to when the app "Provides a way for you to request that your data be deleted, or automatically deletes or anonymizes your data within 90

27

days". As there is no specific way to get this information from the taxonomy, we create a separate high level practice for Data Deletion. For *Secure Data Transfer*, there is low level element in the taxonomy that covers the practice, however, since the other categories from the taxonomy in the hierarchy are not related, we add *Secure Data Transfer* as a high level category. Also note that since there were less than 100 annotations for this category, we also perform additional annotations and increase the dataset size.
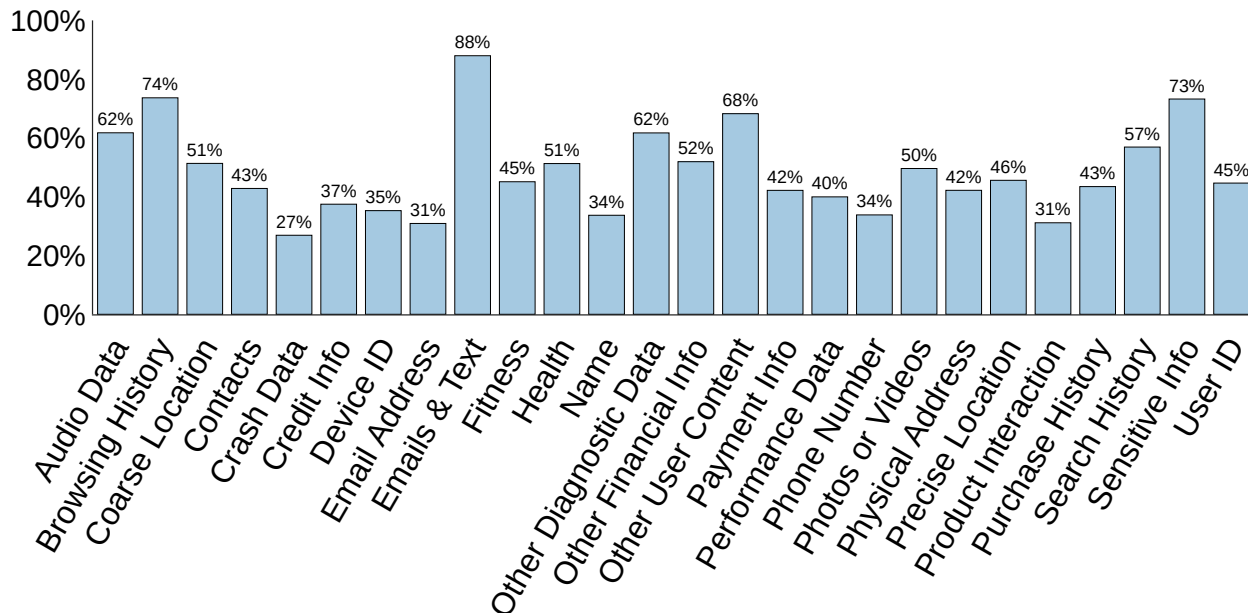
Figure 13

A.2 Annotation Setup

Creating Annotation Set For our *Privacy Label Taxonomy*, we were able to have the data missing from OPP-115 for 13 elements. Curating the candidate set for missing categories is a major challenge due to label imbalance. To address this issue, we follow the approach used by Harkous et al., [23] and use the task of *Natural Language Inference* (NLI) to curate the candidate set. The NLI tasks consist of a hypothesis and a premise, and the objective is to determine if the hypothesis is true (`entailment`), false (`contradiction`) or undetermined (`neutral`) given the premise [34]. For example, if the premise is: *"Your data is safely and completely removed from our servers or retained only in anonymized form."* and the hypothesis is *"Data deletion is being discussed"*, then this instance will receive an entailment. On the other hand, if the hypothesis were *"Policy change is being discussed"*, then the label would be neutral. This method of using NLI-based sampling to reduce the annotation effort has been shown to be effective by Harkous et al. [23].

We start by creating a hypothesis for each of the missing categories that we have. For example, for *Data Deletion Option*, we created two hypotheses: "Data deletion is being discussed" and "Data Anonymization is being discussed". For the NLI task, we used the T5-Large model checkpoint from `Huggingface`. This model is already trained on MultiNLI task [38] which consists of a multi-genre dataset covering a large variety of domains. Next, we run the NLI model and get weak labels for all the missing categories. Note that these are weak labels that are later manually annotated to create the training set.

28

Annotation Details Using the NLI sampling approach, we curated a candidate set with 2000 segments for each of the missing categories. These segments are roughly balanced based on the weak labels assigned by the NLI model. For each class, we then randomly sample 500 segments to annotate. Two of the authors annotated the segments and created the training set.

The annotation was performed using the label studio framework [44]. The framework supports not only simple natural language processing tasks but also sophisticated labels such as taxonomies and sentence highlightings. The framework also supports active learning with the capability of integrating a backend machine learning classifier of one's taste in order to facilitate annotation.

The label studio server was deployed in an internal network, where the two authors simultaneously worked on annotating and creating the training set. Figure 14 shows the annotation setup used by the authors.
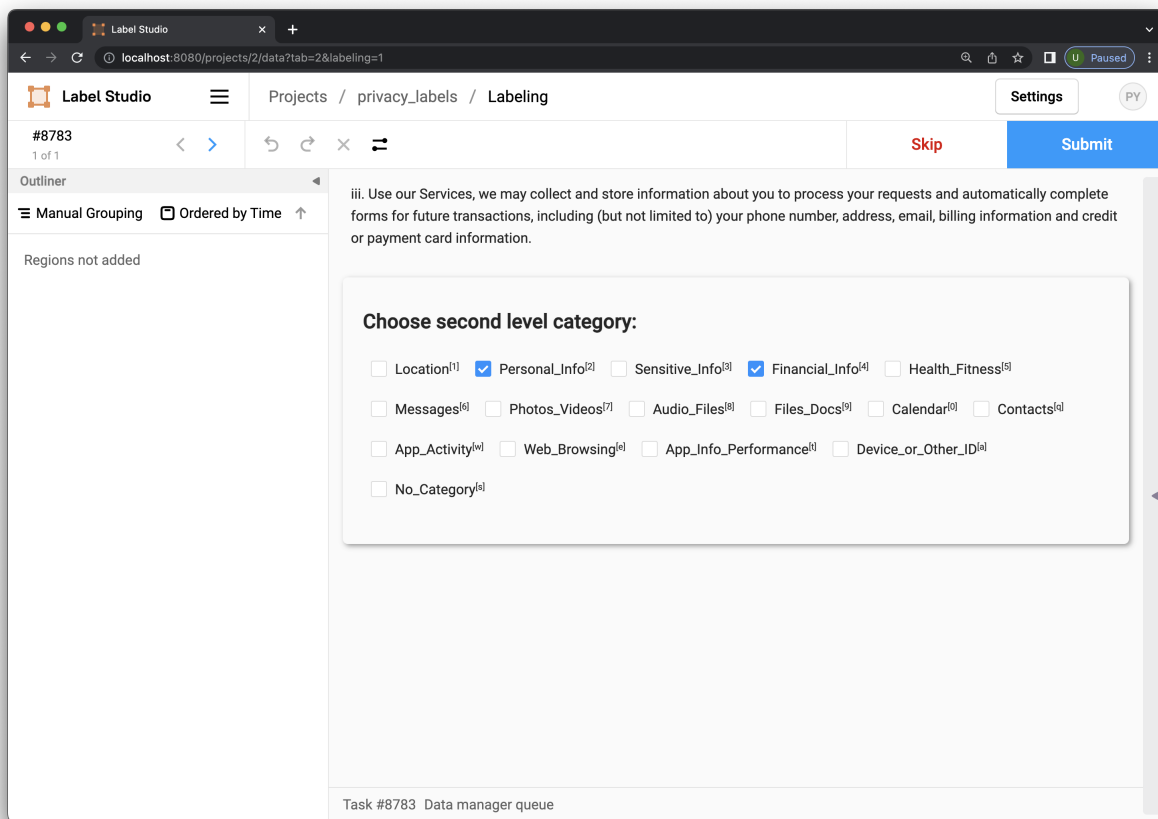

Figure 14: Label Studio Annotation Setup

After the annotation step, the dataset was converted into a data frame consisting of the text column and a binary indicator column for each of the categories, to prepare for the training.

A.3   Training Setup

Large language models like BERT [15], T5 [38] etc have shown remarkable performance using small training sets. Thus, for our purposes, we use the `distilbert-base-uncased` [40] model consisting of 67 million parameters. This model is the distilled version of BERT [40]. It has 40% fewer parameters, can run 60% faster and performs only slightly worse (~5%) than the original
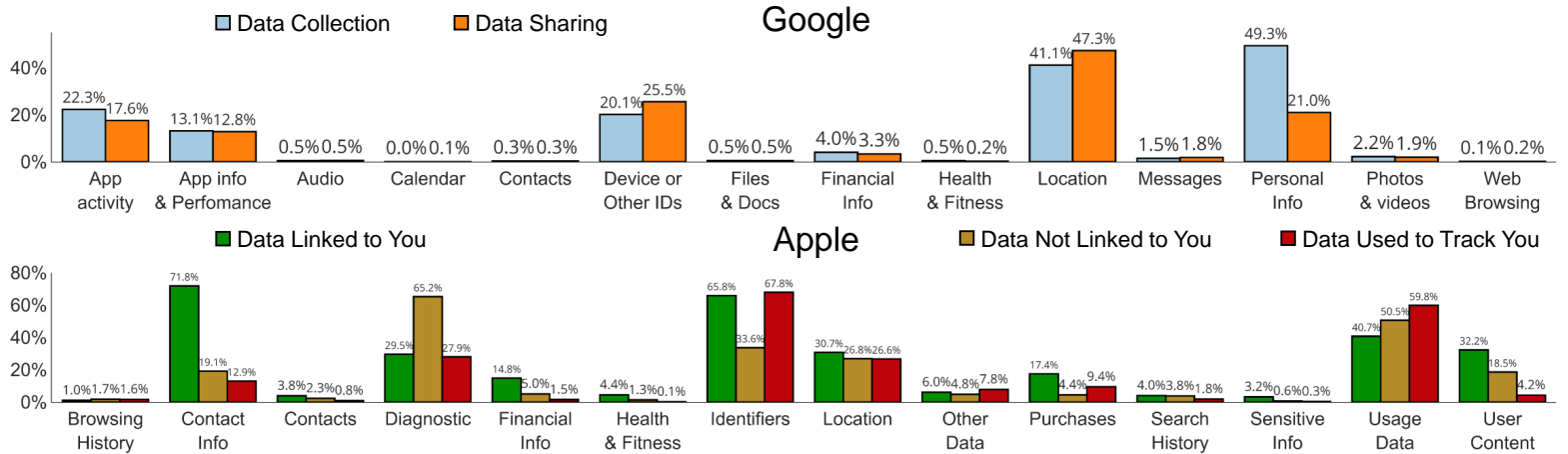
Figure 15: Distribution of data categories for high level practices for apps in Play Store (top) and App store (bottom).

`bert-base-uncased` model on several natural language tasks. Additionally, we also perform domain adaptation by pre-training the DistilBert model on privacy policy text with the Masked Language Model (MLM) task. In particular, we pre-trained the model with the default hyperparameters, with a batch size of 256 for 24800 steps on a single NVIDIA A100 GPU.

We then use the new pre-trained model to train the category classifiers for the Privacy Label Taxonomy. We use a classification head on the model after adding a linear classification. For classification, the data annotated by the authors is split into two parts: testing (20%) and training sets (80%).

B Data Practices in Privacy Labels

In this section, we look into the distribution of all the data types as shown in Fig. 13. From this figure, we observe that the six categories–App Activity, App Info & Performance, Device IDs, Financial Info, Location, Personal Info–are reported to be the most collected in Google, while the six categories–Contact Info, Diagnostics, Identifiers, Location, Purchases, Usage Data and User Content–are reported to be the most collected in Apple.

C Developer Study

For the Developer Study (Section 4.3) we sent emails to developers in 3 different categories: (A) apps stating that they encrypt data without collecting or sharing data, (B) apps changing their practice from not collecting/sharing data to collecting/sharing data, and (C) apps changing their practices from collecting/sharing data to not collecting/sharing data.

For category (A) we used the following template:

We hope this email finds you well. We are researchers at ¡LAB_NAME¿ and have been using your app, ¡APP_NAME¿, in our recent studies. We have noticed that in the data safety section of your app, it states that you encrypt data. However, we have also noticed that your app does not collect or share data.

We are reaching out to ask if you could clarify this for us. We are trying to better understand the data safety section implemented in your app. We appreciate any information you can provide.

Thank you for your time and we look forward to your response.

D   Comparing Privacy Policies with Privacy Labels

In this section, we provide details about how to obtain privacy practices present in the policies. We first note that *Data Category* and *Purpose* are the lowest levels in the taxonomy such that there are classifiers for each data category and purpose. So, extracting lower-level practices from policies is straightforward. Furthermore, we note that we have added classifiers for *Encryption in Transit* and *Data Deletion Option* separately. Additionally, we have two high-level classifiers, namely *First-Party-Collection* and *Third-party-collection* that capture segments that refer to data collection by first parties and data sharing to third parties, respectively.

To extract the remaining high-level practices, we follow the mapping shown in Table 2. Specifically, as shown in Table 2 we use multiple data category classifiers in conjunction with high-level classifier (First-party-collection-share/Third-party-collection-share) to indicate if the policy mentions data collection or data sharing. This result can be directly compared with the categories present in Google's Data Safety Section.

To match the policy to the Apple Privacy Label, we use an additional classifier: *Identifiability*. This represents if the data being collected is anonymous or not. If the data is anonymous then we equate it to the *Data Not Linked to You* label else the *Data Linked to You* label. For example, To obtain whether a policy is collecting *Location* under *Data Linked To You*, we check whether there are any segments where the lower level *Data Category* classifier tags the segments to contain *Location* information. Then we check whether these segments also have either *First-party-collection* or *Third-party-collection-share* tags in combination with *Identifiability-identifiable* tag.
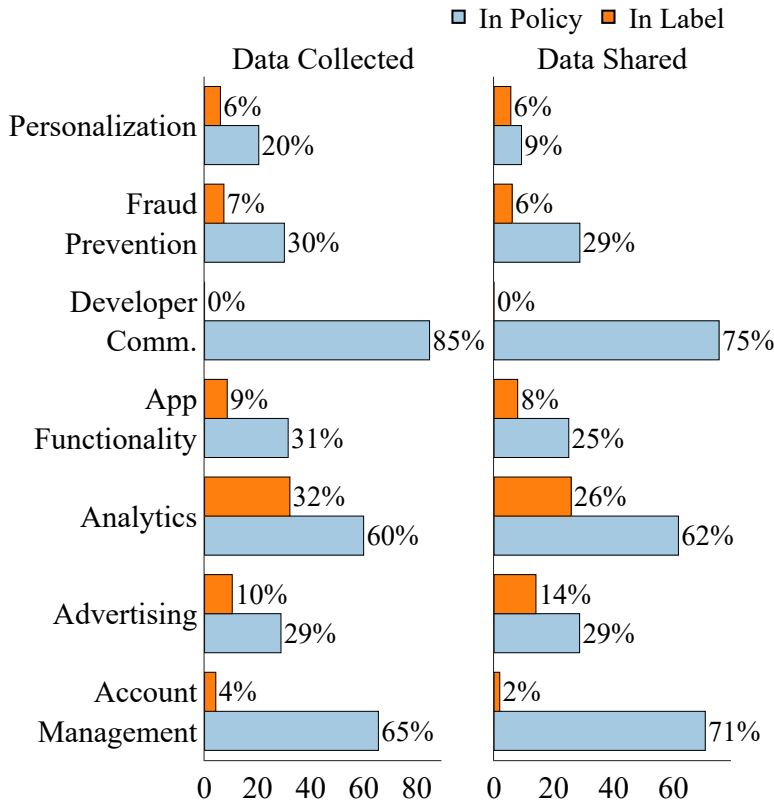

Figure 16: Inconsistencies between privacy policies and DSS, purpose based

E   Mapping from DSS to APL

In Table 4 and Table 5 we show how we convert the datatypes of DSS to those of APL. We do these conversions based on the definitions provided by Google and Apple respectively.
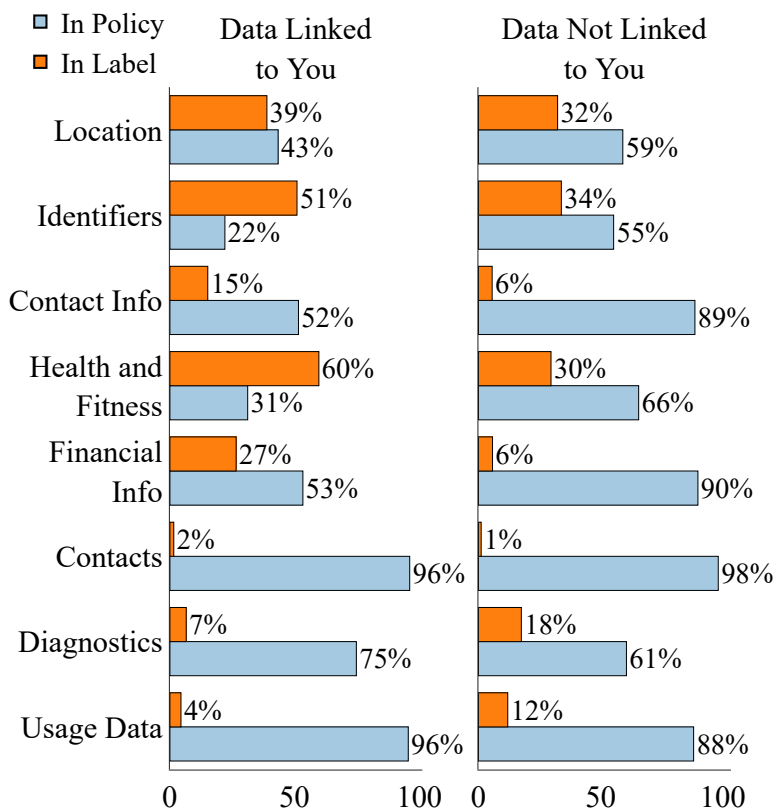

Figure 17: Inconsistencies between privacy policies and APL, datatypes based

Data Category Classifiers	High Level Classifier	Google Data Safety Section	Classifier	Apple Privacy Labels
App Activity, App Info and Performance, Sensitive Info, Location, Health and Fitness, ...	First-party-collection-share	Data Collection	Identifiability (Identifiable)	Data Linked to You
			Identifiability (Anonymous)	Data Not Linked to You
App Activity, App Info and Performance, Sensitive Info, Location, Health and Fitness, ...	Third-party-collection-share	Data Sharing	Identifiability (Identifiable)	Data Linked to You
			Identifiability (Anonymous)	Data Not Linked to You

Table 2: This table shows the bottoms-up approach to get the high-level classification for Google's Data Safety Section and Apple Privacy Labels

Category	CNN [37]	BERT [15]	Here
First-party-collection-share	82	91	98
Third-party-sharing-collection	82	90	96
Identifiability	77	91	97
Does-does-not	86	93	96
Encryption-in-transit	N/A	N/A	99
Data Deletion Option	N/A	N/A	91
(DC) App Activity	N/A	N/A	93
(DC) App Info and Performance	N/A	N/A	93
(DC) Sensitive Info	N/A	N/A	97
(DC) Location	N/A	N/A	99
(DC) Health and Fitness	N/A	N/A	97
(DC) Device or Other ID	N/A	N/A	94
(DC) Photos and Videos	N/A	N/A	96
(DC) Web Browsing	N/A	N/A	96
(DC) Contacts	N/A	N/A	87
(DC) Calendar	N/A	N/A	94
(Purp) Account Management	N/A	N/A	92
(Purp) Developer Communication	N/A	N/A	93
(Purp) Personalization	N/A	N/A	98

Table 3: Classifiers's performance on the test set. Training was done on one GPU with early stopping.

Google Purposes	→	Apple Purposes
Advertising or Marketing	→	Advertising or Marketing
Analytics	→	Analytics
App Functionality	→	App Functionality
Fraud prevention, Security, and Compliance	→	App Functionality
Personalization	→	Personalization
Account Management	→	N/A
Developer Communication	→	N/A

Table 4: Table showing the common mapping from Data Safety Card to Apple Privacy Label

Google DataType	→	Apple DataType
Approximate Location	→	Coarse Location
Precise Location	→	Precise Location
Name	→	Name
Email Address	→	Email Address
Address	→	Physical Address
Phone Number	→	Phone Number
Race And Ethnicity	→	Sensitive Info
Political Or Religious Belief	→	Sensitive Info
Sexual Orientation	→	Sensitive Info
User Ids	→	User Id
User Payment Info	→	Payment Info
Credit Score	→	Credit Info
Other Financial Info	→	Other Financial Info
Purchase History	→	Purchase History
Health Info	→	Health
Fitness Info	→	Fitness
Emails	→	Emails Or Text Messages
Sms Or Mms	→	Emails Or Text Messages
Other In-App Messages	→	N/A
Photos	→	Photos Or Videos
Videos	→	Photos Or Videos
Voice Or Sound Recordings	→	Audio Data
Music Files	→	N/A
Other Audio Files	→	N/A
Files And Docs	→	N/A
Calendar	→	N/A
Contacts	→	Contacts
App Interactions	→	Product Interaction
Other User-Generated Content	→	Other User Content
In-App Search History	→	Search History
Other Actions	→	N/A
Web Browsing History	→	Browsing History
Crash Logs	→	Crash Data
Diagnostics	→	Performance Data
Other App Performance Data	→	Other Diagnostic Data
Device Or Other Ids	→	Device Id
Other Info	→	N/A

Table 5: Table showing the common mapping from Data Safety Card to Apple Privacy Label
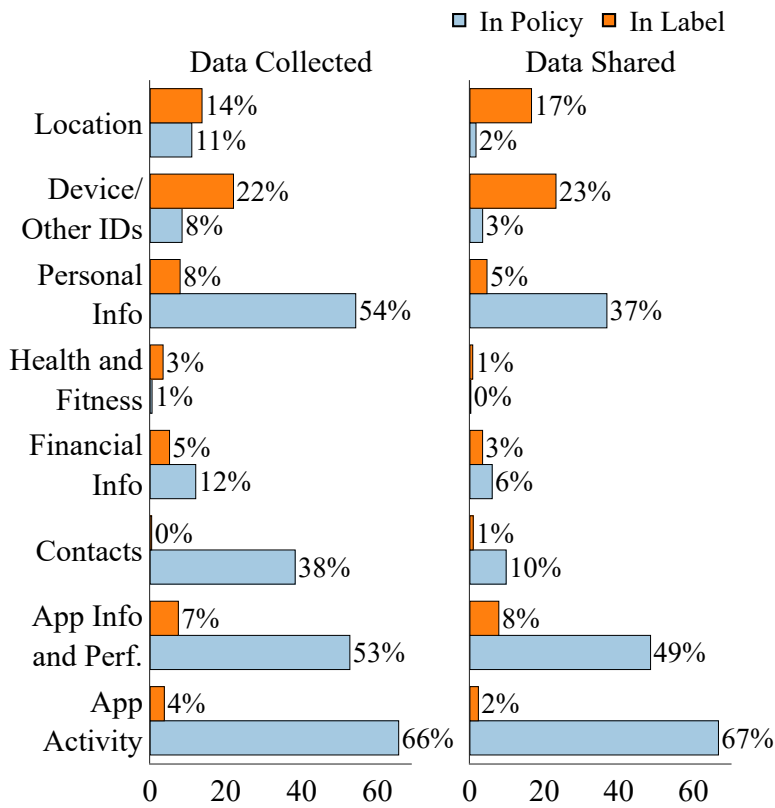
Figure 18: Inconsistencies between privacy policies and DSS, datatypes based
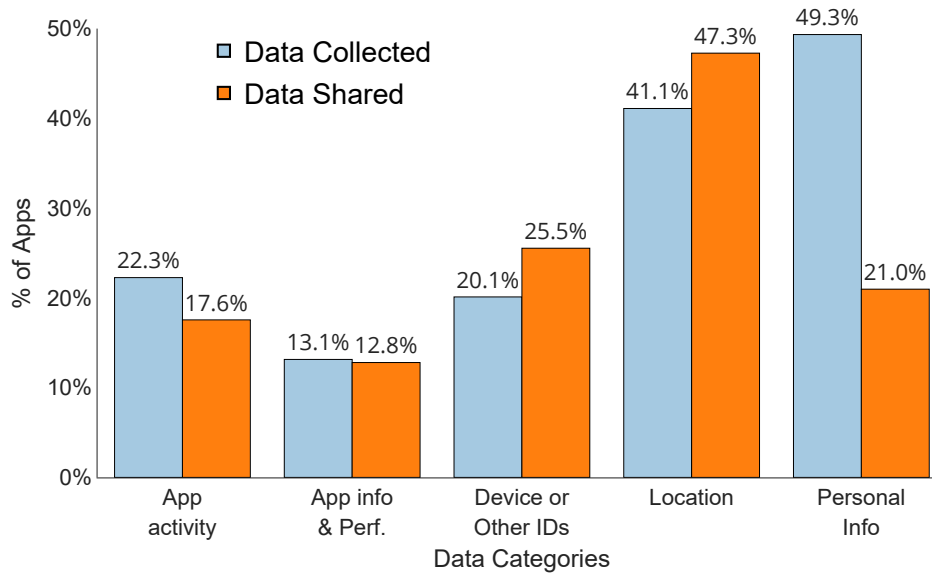

Figure 19: Plot showing what data category is collected or shared in Google Data Safety cards
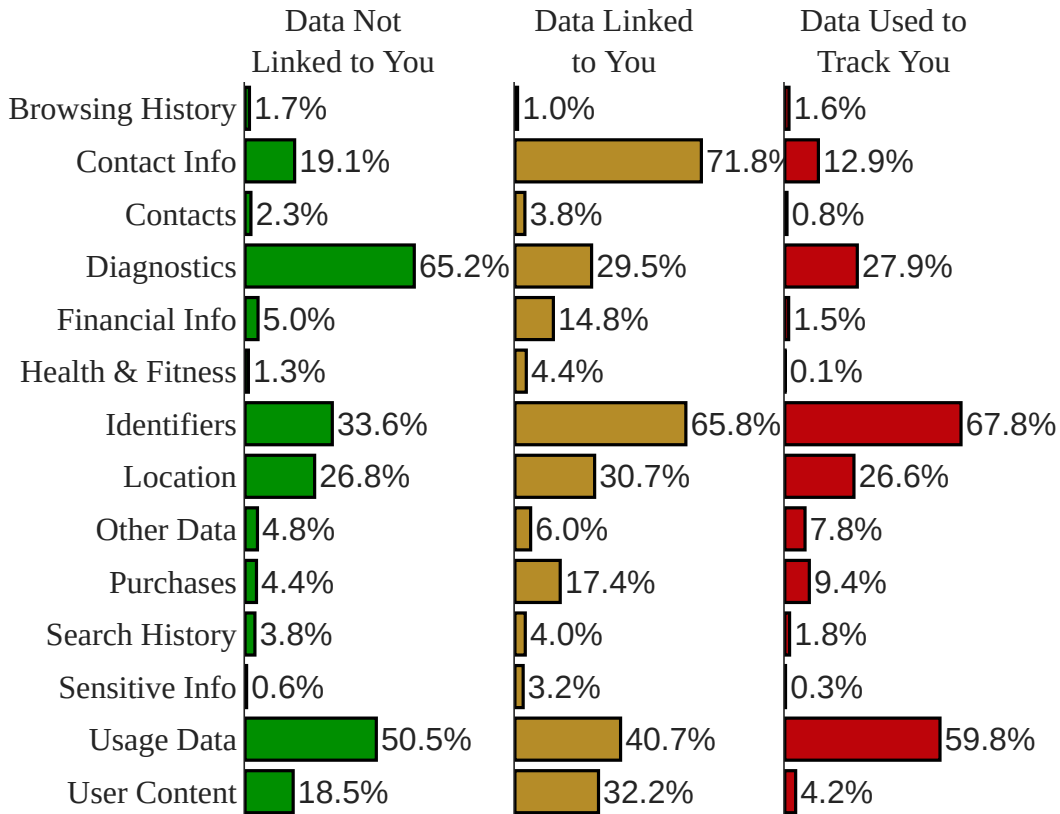
Figure 20: Plot showing the data categories used for high-level privacy practices in APL.
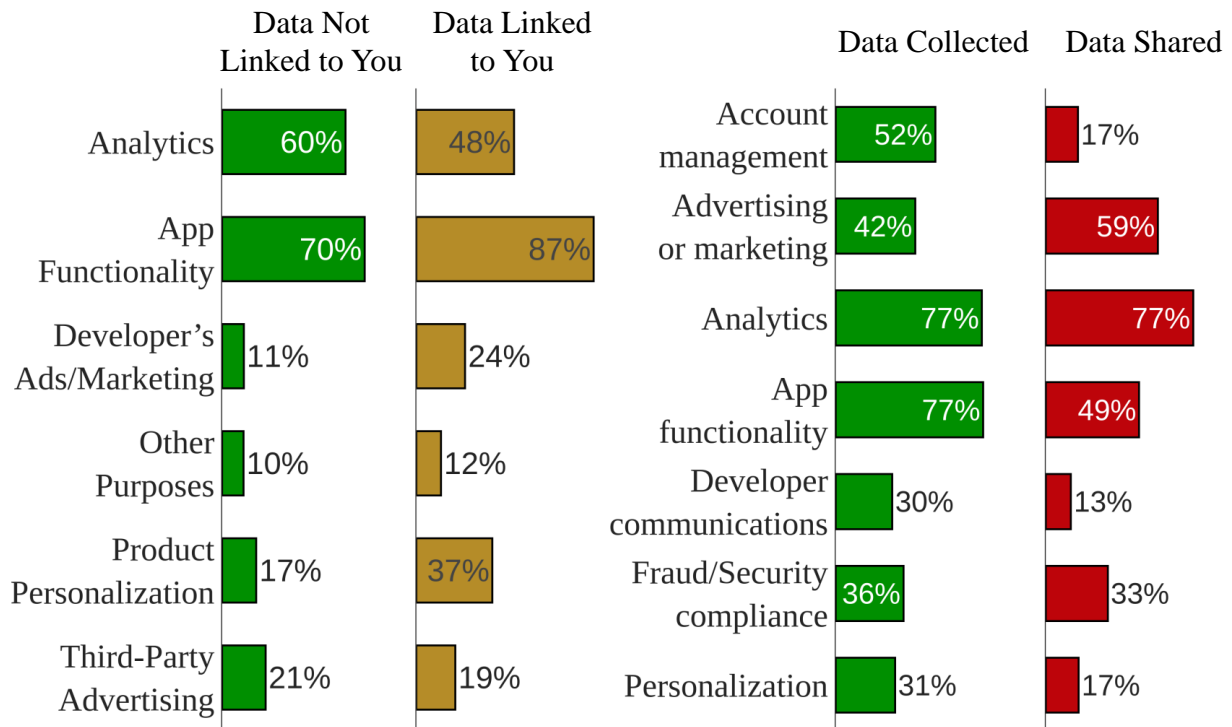

Figure 21: Plot showing the distribution of purpose for high-level privacy practices in APL