Honesty is the Best Policy: On the Accuracy of Apple Privacy Labels Compared to Apps' Privacy Policies

Mir Masood Ali
University of Illinois Chicago
mali92@uic.edu

David G. Balash
University of Richmond
david.balash@richmond.edu

Monica Kodwani
The George Washington University
monicakodwani@gwu.edu

Chris Kanich
University of Illinois Chicago
ckanich@uic.edu

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

ABSTRACT

Apple introduced *privacy labels* in Dec. 2020 as a way for developers to report the privacy behaviors of their apps. While Apple does not validate labels, they also require developers to provide a privacy policy, which offers an important comparison point. In this paper, we fine-tuned BERT-based language models to extract privacy policy features for 474,669 apps on the iOS App Store, comparing the output to the privacy labels. We identify discrepancies between the policies and the labels, particularly as they relate to data collected linked to users. We find that 228K apps' privacy policies may indicate data collection linked to users than what is reported in the privacy labels. More alarming, a large number (97%) of the apps with a *Data Not Collected* privacy label have a privacy policy indicating otherwise. We provide insights into potential sources for discrepancies, including the use of templates and confusion around Apple's definitions and requirements. These results suggest that significant work is still needed to help developers more accurately label their apps. Our system can be incorporated as a first-order check to inform developers when privacy labels are possibly misapplied.

1 INTRODUCTION

Privacy policies are ubiquitous and required in many settings [35–37, 64], and for better or worse, are an important tool for communicating about the behavior of systems. Natural language policies have many shortcomings and are full of technical details and jargon that significantly impact their usability as a tool to inform users clearly about the behaviors and data management practices [28, 58]. *Privacy nutrition labels*, or *privacy labels*, offer an alternative to both simplify and standardize the communication of privacy behavior similar to food nutrition labels [20, 51]. In December 2020, Apple began requiring privacy labels [31] for all new and updated apps in the App Store. Apple's privacy labels ask developers to self-label (without verification) the data collection and sharing practices of their apps, the purposes, the types of data, and if that data is linked to user identities (see Figure 1 for more details). Essentially, privacy labels standardize the presentation of privacy behavior described in the privacy policy's natural language text.

In this paper, we answer the question: *How do privacy labels compare to the behavior described in the privacy policies?*

We conducted a large-scale analysis of the Apple App Store by reviewing 474,669 apps' privacy policies and privacy labels using a validated implementation of PrivBERT [70], a transformer-based privacy policy language model. We fine-tuned PrivBERT with the OPP-115 corpus and mapped its features to Apple's privacy labels to identify discrepancies between the reported behavior of apps based on their labels compared to their privacy policies.

We find that there are large differences between privacy labels and privacy policies. Most prominently, according to our analysis of the privacy policies, nearly 228K *more* apps may be performing some amount of data linking than the number of apps that reported similar data collection in the labels. More alarming, 97% of apps that report no data collection in their privacy label have statements in their privacy policy to the contrary. In many cases, mislabeling varies from the privacy policy regarding the kinds of data collected, particularly around app functionality and analytics or "other" functionality not prescribed by a privacy label.

We also compared free and paid apps. While paid apps use fewer privacy labels compared to free apps, the policies tell a different story: only 4% of paid apps report collecting data that is linked to users, but the policies suggest that 76% paid apps perform such collection. We further analyzed privacy-relevant data practices that are not covered by privacy labels. We found that most apps (76%) had a self-assigned content rating of 4+ on the App Store to indicate age appropriateness and enforce parental controls. Of these apps, only 50% of such apps had a policy in place to handle data collected from children. Our case study further reveals that their policy might be to claim no responsibility for collecting and handling data collected from users under 13 years of age. We also employ a similarity metric and identify that 65% of evaluated apps potentially use templates, providing insight into a possible source of discrepancies. We further analyzed the network traffic from 30 apps, showing that their data collection practices diverged from those declared in privacy labels and privacy policies.

Our analysis indicates that privacy labels are likely misapplied in great numbers, even considering that classifiers are imperfect for analyzing privacy policies. More guidance for developers would go a long way toward improving the accuracy of privacy labels. Still, there are also more concerning misapplications that could and should be addressed more broadly, such as the collection of data used to track users and apps falsely reporting that they do not collect any data. In these cases, the privacy policies are often explicit

Proceedings on Privacy Enhancing Technologies 2024(4), 142–166

https://doi.org/10.56553/popets-2024-0111

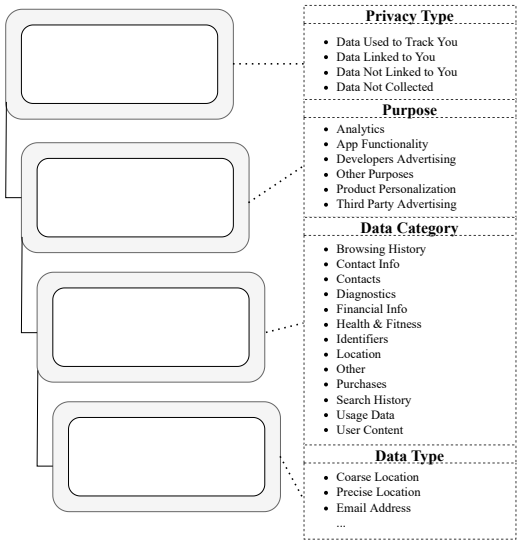

Figure 1: Anatomy of a Privacy Label.

in this behavior, and the absence of a corresponding entry in the privacy label could lead to misunderstandings of the risks associated with using these apps and potentially violate Apple's App Store policies. First-level checks of the privacy policies when apps are submitted to the App Store could go a long way in highlighting and correcting some of the more common and egregious privacy label inaccuracies. In this work, we make the following contributions.

- We build and validate a hierarchical framework that uses fine-tuned transformer models to extract multiple features from privacy policies.
- We develop and validate a mapping between features extracted from classifiers and App Store privacy labels.
- We collect and analyze the privacy labels of 474,669 apps against their policies and find large differences in their reported practices.
- We use a similarity metric to compare policies against templates and find that their use might indicate a likely source of observed discrepancies. We also present examples from a case study of traffic collected 30 apps and show evidence of discrepancies.
- We publicly release our code and dataset of app metadata and privacy policies to facilitate further research. The artifact associated with this paper can be accessed at https://github.com/masood/2024-pets-privacy-labels-policies.

2 BACKGROUND AND RELATED WORK

Anatomy of a privacy label. The Apple privacy labels are similar in style and content to the "Privacy Facts" label developed by Kelley et al. [53]. The structure of a label is hierarchical (see Figure 1) and describes data collection practices under four levels: **(1) Privacy Type:** Describes how the app handles collected data, which includes data collected for tracking users (with third parties), data collected and linked to users' identities, and data collected but aggregated/anonymized. An app's privacy label may contain a combination of one, two, or all three types. An app may also report that data is not collected, which is mutually exclusive with the other types. **(2) Purpose:** Discloses the intended reason for the data collection, e.g., for advertising, analytics, personalization. **(3) Data Category:** Reports at a high level the category under which collected data falls. **(4) Data Type:** Granular information that describes the data collected under the Data Category.

Privacy nutrition labels. Privacy nutrition labels have been studied from a variety of perspectives [16, 27, 32, 33, 51–53, 68, 71], but Apple's privacy label is the first wide-scale deployment [31]. In an exploratory study, Li et al. [56] observed the adoption of iOS privacy labels on the App Store and found that very few developers *voluntarily* created privacy labels. Balash et al. [15] performed a 66-week analysis of the privacy label adoption on the Apple App Store and identified a steady increase in the number of apps with privacy labels and likely under-reporting by developers forced to provide a label on a version update.

Zhang et al. [81] conducted an in-depth interview study to determine the usability of iOS privacy labels from a user perspective. Most users found the privacy labels helpful despite misunderstandings that included unfamiliar terms and a confusing structure. Garg et al. [40] discovered that privacy label disclosures of sensitive information reduce app demand, and thus, the accuracy of the labels is important to help users make informed choices.

Gardner et al. [39] developed a tool to assist developers by prompting them while coding functionality that would potentially require a privacy label. Li et al. [55] studied developers' creation of Apple's privacy nutrition labels and conducted semi-structured interviews. They found that errors and misunderstandings were prevalent in the privacy label generation process. These errors included under-reporting linked data, third-party data use, and missing data types. We observe the same when comparing the privacy policies and Li et al.'s findings regarding "knowledge blindspots" and misinterpreted Apple's definitions, likely leading to many of the misapplications we identified.

Privacy behavior of mobile apps. Numerous studies have measured the privacy behaviors of mobile applications [8, 9, 18, 19, 21, 61, 69, 80, 82, 83]. One of the first approaches to automatically identify problems in privacy policies was PPChecker [80], which combined an NLP analysis of privacy policy text with bytecode analysis. Andow et al. [8] developed PolicyLint to identify contradictions within an individual policy. Andow et al. [9] also created PoliCheck, which considers third-party versus first-party entity access to personal data for an entity-sensitive consistency check. Bui et al. [19] extended PoliCheck to develop PurPliance that checks if data, entity, and purpose are equivalent to those extracted from data flows. In this paper, we choose Polisis [46] as the policy analysis tool as it produces output similar to the privacy labels.

Zimmeck et al. [82] evaluated 1,035,853 Android apps using the Mobile App Privacy System (MAPS), a pipeline based on code analysis and supervised machine learning classifiers, to identify potential non-compliance with privacy standards. Kollnig et al. [54] analyzed 1,759 iOS apps using a combination of code analysis and network traffic monitoring, and they found that 80% of the apps that claimed not to collect any data in the privacy labels contained at least one tracker library. We find that this discrepancy probably exists at scale.

Xiao et al. [79] analyzed 5,102 apps (~ 1% of our dataset) by checking the privacy labels against actual data flows and focused on two levels of labels, *Purposes* and *Data Types*. They discovered

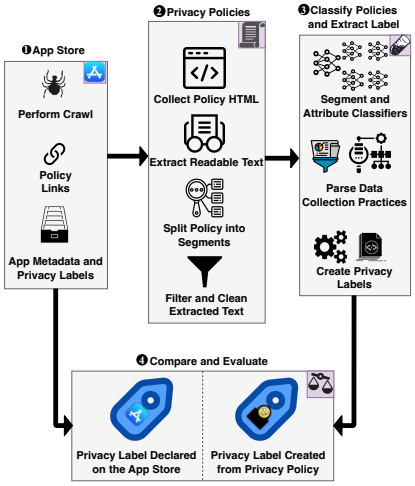

Figure 2: An overview of the measurement workflow.

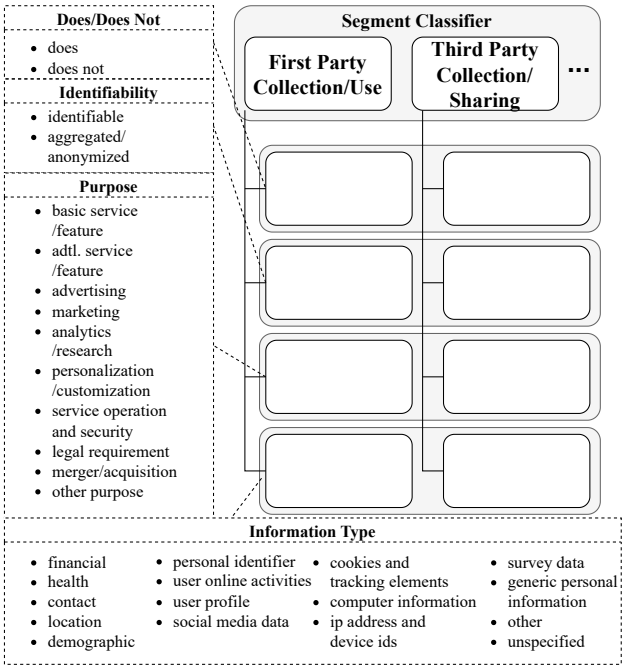

Figure 3: The hierarchical structure of the Polisis classifiers.

that 67% of those apps failed to accurately disclose their data collection practices, particularly around the use of *User ID*, *Device ID*, and *Location* data. Our results complement their findings, where mentioning the collection of unique identifiers in an identifiable manner in the privacy policy is not reflected in the privacy labels. Further, our work analyzes apps at a much larger scale and covers *Privacy Types*, *Purposes*, and *Data Categories*.

Apple's deviations from recommendations. Although derived from Kelley et al.'s [53] work, Apple's implementation deviates from its recommendations. While Kelley et al. noted, "presenting *[labels]* clearly and simply we could affect user decisions," Apple displays the nutrition label embedded down on the App Store, requiring interested users to scroll through details, where users may not see the labels before deciding to install an app. Additionally, Apple's labels do not give users choices or allow them to compare labels between apps. Further, recent user studies have found the labels to be confusing for developers [55], showing the possibility that developers misapply labels. Finally, Kelley et al. highlighted the need for the labels to be accurate and noted, "users believe this information is correct, is being verified, and will assume they misunderstand something before they would believe the displays are incorrect." Since Apple's privacy labels are not vetted and are not trustworthy, this points to a serious concern about providing disinformation to end users. These factors further highlight the necessity to verify and demonstrate the discrepancies we present.

3 MEASUREMENT WORKFLOW

In Figure 2, we present the primary measurement workflow described in detail below. During all scans, we followed best practices of limiting the number of requests and respecting `403 Errors` by using exponential back-offs.

❶ ***Crawling the App Store.*** We began by parsing the `XML` site map from Apple's App Store, which lists all apps currently published on the store, and then crawled each URL, parsing the privacy labels and associated metadata, such as the app name, version, size, type, user rating, genre, content rating, release date, seller name, and price. Notably, the metadata includes a link to the privacy policy. We also parsed the extended privacy label details, such as the purposes and data types, by performing an additional GET request to the Apple Catalog API [12]. In January 2024, there were 1.2M apps on the App Store. Of them, 995K apps had a privacy label, and we identified 993K apps with links to 669K unique policies (note that some apps link to the same policy).

❷ ***Collecting Privacy Policies.*** We extracted the HTML for each policy using a Python script. We leveraged the *readability* library [60, 73], a standalone version of the Firefox browser *reader mode*. The library employs a complex set of heuristics to extract relevant text from web pages [72], leaving us with de-cluttered HTML that we divided into segments based on the <p> tags. We then used a wrapper library on Google's language detection to discard non-English policies [63]. When policies included lists where each list entry was not self-contained, we merged these lists into the preceding text to provide relevant context. We scanned short lists, i.e., where each list item was composed of <20 words, and merged them into the preceding paragraph, thereby treating the entire list as a single segment. We then eliminated segments comprising <20 words. After cleaning, the classifiers individually processed each segment and mapped it back to the original policy. After excluding links that returned response errors, the *readability* library successfully extracted relevant text from 286,717 policies, which we classified in the next stage.

❸ ***Classifying Policies and Extracting Labels.*** We analyzed policies with a similar approach to Polisis [46], an NLP framework that classifies data collection behavior from privacy policy text. Unfortunately, the prior published Polisis implementation is

proprietary, and on reaching out, the authors informed us that their website can only take up to 30K policies. We completely re-implementated the classification framework to the same standards as prior work. We replaced their CNN-based approach with a state-of-the-art language model to improve classifier performance. We used PrivBERT [70], a transformer-based privacy policy language model, which was developed by pre-training the RoBERTa$_{BASE}$ model [57] on 1M privacy policies. We fine-tuned PrivBERT on the OPP-115 corpus [78]. We present an overview of the framework structure in Figure 3. In Table 5 in the Appendix, we show that the PrivBERT classifiers perform better than CNNs. We provide more details about training and evaluating the models in Appendix A.

We first passed each segment through the Segment Classifier to extract the high-level data practice. We passed any segments addressing *First Party Collection/Use* or *Third Party Collection/Sharing* through six Attribute Classifiers – *Does/Does Not*, *Identifiability*, *Purpose*, *Personal Information Type*, *Action First Party*, and *Action Third Party* – to extract annotations relevant to privacy labels. We used the *Action First-Party* attribute to filter any segments explicitly addressing collection on websites (and not mobile apps). We used the *Action Third-Party* attribute to eliminate instances wherein the third party only 'sees' and does not collect data. We successfully detected segments addressing data collection in the policies of 474,966 apps ($n = 280,767$ policies), which we then used to create privacy labels.

❹ ***Compare and Evaluate.*** The taxonomy of policy labeling does not always have a one-to-one mapping with Apple's privacy labels. So, we developed a grounded strategy based on qualitative coding to convert outputs from classifiers into equivalent privacy labels. Three researchers *independently* coded the conversions and then discussed to reach an agreement on the mappings between OPP-115 and privacy labels. The coders completed three matching tasks:

- First, the coders determined which of the data practices found by the Segment Classifier, such as *First Party Collection/Use* or *Third Party Collection/Sharing*, that when combined with the Identifiability Attribute Classifier, such as "Identifiable," "Aggregated/Anonymized," "Does", or "Does Not", match to an appropriate Apple privacy label type, such as *Data Linked to You* or *Data Not Collected*. For example, when the framework identifies a segment with a data practice of "First Party Collection/Use" and the data is "Identifiable," that would associate with an Apple privacy label type of *Data Linked to You*.
- Next, the coders matched the output of the Purpose Attribute Classifier against Apple's privacy label purposes. For example, a framework output of "Basic Services/Features" gets mapped to *App Functionality* for privacy label purposes.
- Finally, the coders matched the outputs of the Personal Information Type Attribute Classifier to the data categories provided in Apple's privacy label. For example, Polisis may identify that a segment discusses "Contact," which then maps to the privacy label data category of *Contact info*.

The combination of these three matching tasks provides a single privacy label entry for an app, according to the privacy policy, describing the privacy type (e.g., *Data Linked to You*), the purpose (e.g., *App Functionality*), and the data category collected (e.g., *Contact Info*). We can find the full list of the direct conversions in this mannerTable 1. The coding process also revealed additional, inferred privacy labels from Polisis classification that included a combination of classifications and keywords relevant for *Data Used to Track You* and remaining *Data Categories*. Table 2 shows the inferred privacy labels. We further verified the mapping by randomly sampling labels generated from classifier outputs. In the Appendix, we present our evaluation in Table 4.

Table 1: Deriving privacy label entries directly from segment annotations created using the Polisis framework.

Apple Privacy Label	**Polisis**	
Privacy Type	**High-level Data Practice**	**Identifiability**
Data Linked to You	First Party Collection/Use Third Party Collection/Sharing	Identifiable
Data Not Linked to You	First Party Collection/Use Third Party Collection/Sharing	Aggregated/anonymized
Privacy Type	**High-level Data Practice**	**Does/Does Not**
Data Not Collected	First Party Collection/Use Third Party Collection/Sharing	Does Not
Purpose	**High-level Data Practice**	**Purpose**
App Functionality	First Party Collection/Use Third Party Collection/Sharing	Basic Service/feature Additional Service/feature Service operation & security
Analytics	First Party Collection/Use Third Party Collection/Sharing	Analytics/Research
Developers Advertising	First Party Collection/Use	Advertising
Other Purposes	First Party Collection/Use Third Party Collection/Sharing	Merger/Acquisition Legal requirement Unspecified
Third Party Advertising	Third Party Collection/Sharing	Advertising
Product Personalization	First Party Collection/Use Third Party Collection/Sharing	Personalization/Customization
Data Category	**High-level Data Practice**	**Personal Information Type**
Contact Info	First Party Collection/Use Third Party Collection/Sharing	Contact
Location	First Party Collection/Use Third Party Collection/Sharing	Location
Financial Info	First Party Collection/Use Third Party Collection/Sharing	Financial
Identifiers	First Party Collection/Use Third Party Collection/Sharing	Cookies & Tracking Elements IP address & Device IDs
Usage Data	First Party Collection/Use Third Party Collection/Sharing	User Online Activities
User Content	First Party Collection/Use Third Party Collection/Sharing	User Profile Social Media Data
Health & Fitness	First Party Collection/Use Third Party Collection/Sharing	Health
Browsing History	Third Party Collection/Sharing	User Online Activities

4 LIMITATIONS

Before proceeding, it is essential to note the limitations of our approach in comparing the privacy labels with the privacy policies.

Ground truth. Foremost, we note that neither the labels nor the policies can provide comprehensive ground truth of app behavior, and even statistical and dynamic analysis has limitations. Here, we report only on observed discrepancies between the policies and the labels, but validating which is more in line with app behavior is beyond the scope of this paper. However, as these discrepancies

Table 2: Inferring privacy label entries from segment annotations created using the Polisis framework.

Apple Privacy Label	Polisis	
Privacy Type	**High-level Data Practice**	**Purpose**
Data Used to Track You	Third Party Collection/Sharing	Advertising
Data Category	**High-level Data Practice**	**Personal Information Type**
Diagnostics	First Party Collection/Use Third Party Collection/Sharing	Computer Information IP address & Device IDs
Contacts	First Party Collection/Use Third Party Collection/Sharing	Social Media Data 'contact' , 'friend' 'address book', 'phone book'
Purchases	First Party Collection/Use Third Party Collection/Sharing	Financial User Online Activities
Search History	First Party Collection/Use	User Online Activities 'search'
Sensitive Info	First Party Collection/Use Third Party Collection/Sharing	Demographic 'race', 'racial','ethnic', 'ethnicity', 'sexual orientation', 'sexual preference', 'pregnancy', 'pregnant', 'childbirth', 'child birth', 'child-birth', 'disability', 'religion', 'religious', 'religious belief', 'trade union', 'union member','politics', 'political' 'genetic','genetic information', 'biometric'

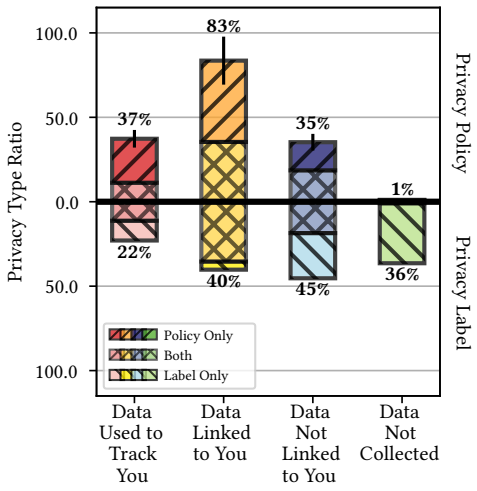

Figure 4: An overview of apps declaring data collection with corresponding *Privacy Types* within their privacy policies (top) and on the App Store via privacy labels (bottom). The denominator is the total apps that we analyzed, i.e., 474,669 apps. Please note that the privacy types, except for *Data Not Collected*, are *not* mutually exclusive.

occur at scale (as reported in the next section), there are strong indications of prominent misapplication of privacy labels according to the privacy policies provided by app developers. Additionally, we present examples via case studies (section 6 and Appendix B) to show how such discrepancies occur with popular apps.

Classifier Predictions. The outputs of language models introduce uncertainty that propagates further when combined. As a result of these inaccuracies, we can only report on the presence of statements addressing data collection practices in privacy policies and differences when compared with privacy labels. However, the reported discrepancies are *much* larger than the associated uncertainties. Additionally, our framework analyzes privacy policies on a per-paragraph/per-segment basis, so it cannot detect explanations of app behaviors that span multiple segments.

Train/Test Dataset. Without an updated corpus with equivalent robustness, we used the OPP-115 corpus to fine-tune language models [78], an extensive dataset comprising manual annotations of 23k fine-grained data practices gathered from multiple graduate-level law students. However, the dataset includes old privacy policies that the researchers collected before the introduction of present-day privacy laws. We identify the limitations introduced by this dataset and recognize the need for an updated dataset. Additionally, specific annotations in the OPP-115 corpus do not directly map to the Apple privacy label taxonomy. As such, the independent annotators used a grounded approach to develop an inferential mapping to address this limitation (see Table 2). Finally, in Table 5 performance, we manually evaluate the classifier outputs on new policies by randomly sampling segments from our dataset of app policies.

Information Extraction. Privacy policies comprise varying formats, reducing the amount of information we can gather from our framework. As previously highlighted, our per-segment approach misses information that spans multiple, non-contiguous segments. Next, policies present information in various media formats (e.g., images) that we do not include in our analysis. Finally, many privacy policies contain links to third parties' privacy policies. We did not analyze the transitive closure of all privacy policies as part of this work. Apple's policy is for privacy labels to include all collection and tracking mechanisms, including third-party practices. Our analysis is a lower bound of data collection performed within an app, particularly related to third parties.

5 RESULTS

In this section, we directly compare developers' reported privacy labels to the output of language models following the hierarchical structure of the privacy labels (see Figure 1).

Privacy Types. We first consider the top level of privacy labels, the privacy types: *Data Used to Track You*, *Data Linked to You*, *Data Not Linked to You*, and *Data Not Collected*. We are primarily concerned with determining the number of apps with such a privacy type and if we can also find that privacy type in the policies. Figure 4 and Table 3 provide a snapshot of the overlap of privacy types extracted from privacy policies and the privacy types declared in the privacy labels for the app on the App Store. As a helpful reminder while reading the numbers reported in this table, three of the privacy types, *Data Used to Track You*, *Data Linked to You*, and *Data Not Linked to You*, are *not* mutually exclusive. Apps may collect data linked to the user and aggregated/anonymized (i.e., not linked to the user), and they may also collect data to track the user.

The *Data Linked to You* privacy type indicates that the app collects data linked to users, i.e., in an identifiable manner. Of the 190,965 apps indicated such collection on the App Store, our framework identified 88% (n = 168, 121) (Fig. 4; lower half; yellow bar; hatches). More concerning, we observed an additional 228,539 apps

Table 3: The number of apps with three of the privacy types associated with their data collection practices, as stated in privacy labels, against practices found in privacy policies. Please note that three of the *Privacy Types* shown here, *Data Used to Track You, Data Linked to You* and *Data Not Linked to You,* are not mutually exclusive. (values) indicate the number of apps that did *not* also declare the corresponding privacy type found by Polisis.

Policy \ Label	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected
Data Used to Track You	**53,359**	83,160 (48,039)	91,665 (53,912)	45,074 (45,074)
Data Linked to You	97,333 (34,294)	**168,121**	188,041 (97,029)	131,310 (131,310)
Data Not Linked to You	44,479 (13,636)	77,171 (35,815)	**88,172**	43,354 (43,354)
Data Not Collected	391	431	796	**4,359**

that reported this practice in their policies but did not report it on the App Store (Fig. 4; top half; yellow bar; stripes).

We identified that 41% ($n = 88,172$) of the apps whose privacy labels stated that they collected data in an aggregated/anonymized manner, i.e., had *Data Not Linked to You* privacy type, also said so in their policies (Fig. 4; lower half; blue bar; hatches). Of the remaining 59% ($n = 127,020$) apps that had the *Data Not Linked to You* privacy type in their label but did not have a corresponding policy segment (Fig. 4; lower half; blue bar; stripes), 76% ($n = 97,029$) of those instead included segments in their privacy policy that indicated that they collect data linked to users (Table 3; row2; col3). This difference may result from apps *not* stating their aggregation practices in the same segment of the policy that addresses data collection. Despite factoring in uncertainty, there is a large gap between the practices declared in privacy labels and privacy policies.

Perhaps more problematic is apps that report they do not collect any data. Recall that the *Data Not Collected* privacy type *is* mutually exclusive, i.e., developers only added this label to apps that claim *not* to collect *any* data from users. While 36% ($n = 172,924$) of the apps that we analyzed indicated in their privacy label that they did *not* collect *any* data, only 0.03% ($n = 4,359$) of these apps made similar statements in their policies (Fig. 4; lower half; green bar). More surprisingly, 84% ($n = 173,441$) of these apps stated in their policies that they collected data linked to users (Table 3; row 4; col 2).

Finally, of the 108,937 apps that stated on the App Store that they collected data to track users, our framework also reported similar practices in the privacy policies of 49% ($n = 53,359$) (Fig. 4; bottom half; red bar; hatches). We identified an additional 123,675 apps that did not declare this practice on the App Store (Fig. 4; top half; red bar; stripes). Recall that the framework infers this privacy type, and we, therefore, partially report user tracking that apps engage in, presenting a lower bound of mislabeling. Our identification of apps that fail to report data collected for tracking indicates that many apps are under-reporting their tracking practices.

Takeaways. Developers are very likely under-reporting their collection of identifiable data on the App Store. Most apps that indicate on the App Store that they do *not* collect *any* data state otherwise in their privacy policies.

Purposes. We look at how apps claim to use the data they collect. Figure 5 presents a snapshot of the purposes associated with data collection, as identified from privacy labels and privacy policies. As a reminder, apps may collect both linked and not linked (anonymized) data. Additionally, apps may collect data for multiple purposes. For example, an app may collect your *Location* in an anonymized manner to personalize user experience (*Product Personalization*) and in an identifiable manner to help advertisers and agencies tailor the advertisements they display (*Third Party Advertising*).

We find greater agreement between privacy labels and privacy policies for apps that collect data for *App Functionality* and *Analytics*. Of the 161,587 apps indicated in their privacy label that apps collect data linked to users for *App Functionality*, 81% ($n = 130,108$) also included a corresponding statement in their privacy policy. Similarly, of the 105,729 apps that stated in their privacy label that they collect data linked to users for *Analytics*, 68% ($n = 71,883$) also included a corresponding statement in their privacy policy (Fig. 5; bottom half; left plot; yellow bars 1 & 2; hatches).

We find notable discrepancies in developers' reporting of *Third-party Advertising* in their privacy policies and on the App Store. Considering data collection that is linked to users (Fig. 5; left plot; bar 4), 139,765 apps exclusively declare this purpose in their privacy policies (top half) and do not report this practice on the App Store. Our findings are concerning since this is a lower bound. Privacy policies link to third-party policies instead of including details here. The results indicate that developers focus on their app's data collection practices when filling out privacy labels without considering third parties. We further highlight the problem of incomplete labeling with examples in §6 and Appendix B.

Finally, we find that while 366,840 (77%) apps stated in their privacy policies that they collected data in an identifiable manner for a purpose that does not fit into any of the options that Apple provides in their privacy label, only 17,487 (5%) of these apps also addressed this on the App Store (Fig. 5; left plot; yellow bar 6). It appears that developers are less forthcoming about declaring data collection in their privacy labels for purposes beyond Apple's taxonomy, making limited use of the catch-all: *Other Purposes*.

Takeaways. Developers are *more* likely to declare data collection for *App Functionality* and *Analytics* in either, privacy labels or privacy policies. Developers are also *less* likely to declare data collection in their privacy labels for purposes beyond Apple's taxonomy, i.e., *Other Purposes*.

Data Categories. We additionally analyze the data categories collected by apps as stated in their privacy labels and policies. Figure 6 provides visual results of our findings, and we present additional details in Table 7 in Appendix D.

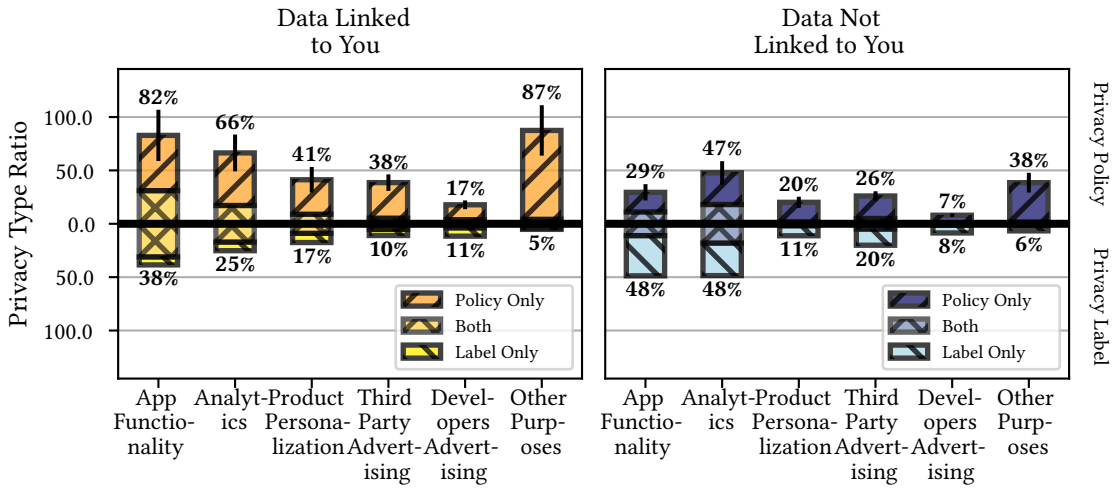

Figure 5: The ratios of the six purposes for the *Data Linked to You* and *Data Not Linked to You* privacy types. The denominator is the number of apps with the designated privacy type either in their privacy label or their privacy policy, i.e., 419,504 apps with a *Data Linked to You* label and 294,391 with a *Data Not Linked to You* label. It is helpful to note here that privacy types shown here are *not* mutually exclusive. Two other *Privacy Types* are not shown here; the *Data Used to Track You* privacy type refers to collection for the purpose of tracking, while the *Data Not Collected* refers to the absence of *any* data collection.

We find that apps are more likely to declare in either their privacy policies or their privacy labels that they collect *Contact Info* (n = 273,351; 65%) and *Identifiers* (n = 320,607; 76%) linked to users (Fig. 6; middle plot; yellow bars 2 & 5). Apps that collect data to track users are more likely to use *Browsing History* (46%; n = 106,816), *Identifiers* (71%; n = 164,732), and *Usage Data* (65%; n = 150,651) (Fig. 6; upper plot; red bars 1, 5, & 7). Our findings are in line with previous work that showed tracking activities target users with cookies and tracking pixels (*Identifiers*) and monitor their browsing practices across sites and services (*Browsing History* and *Usage Data*) [5, 34].

However, we find that apps that state in their privacy policy that they collect *Browsing History* (i.e., how users browse the Internet outside of the app) and *Sensitive Info* (such as racial/ethnic data, sexual orientation, etc.) linked to users are less likely to declare this collection in their privacy labels (Fig. 6; middle plot; top half; yellow bars 1 & 13). Surprisingly, of the 212,121 apps that stated in their privacy policy that they collect *Browsing History* linked to users, only 658 (0.3%) of these apps declared this practice in their privacy labels. While 96,837 apps indicated in their privacy policy that they collect some form of *Sensitive Info*, only 2% (n = 2,144) apps also declared this collection in their privacy labels. Of notable concern, we find 22,171 apps and 11,710 apps mislabeling their collection of *Identifiers* and *Contact Info* respectively as being linked to users when their policies indicate that they use collected data to track users (see Table 7).

Takeaways. Developers most commonly state that they collect *Identifiers* and *Contact Info* that are linked to users. Developers that state in their privacy policies that they collect *Browsing History* or *Sensitive Info* linked to users are less likely to declare this collection in their privacy labels. Apps that track users are more likely to use *Browsing History*, *Identifiers*, and *Usage Data*, which is in line with prior findings about tracking practices.

Free vs. Paid Apps. The App Store has four pricing models: free apps, free apps with in-app purchases, paid apps, and paid apps with in-app purchases. Interestingly, when only observing privacy labels (Fig. 7; all plots; bottom half), it would appear that paid apps have better privacy behaviors than their free counterparts. However, the altruism of paid apps compared to free apps disappears when considering the privacy policies (the top half of Figure 7). The privacy policy analysis better aligns with the observations of Han et al. [44, 45], who compared free and paid apps in the Android Play Store based on the inclusion of third-party advertising software, finding no differences between free and paid apps.

As a result of apparent under-reporting by paid apps, we find that they have the largest discrepancies of potentially under-reporting data collection practices in their privacy labels compared to the privacy policies. While the privacy policies suggest that 75% (n = 21,330) of paid apps collect data linked to users, only 4% (n = 1,145) paid apps have a privacy label of this type (Fig. 7; second plot; yellow bar 3). More concerning, while the privacy policies of 21% (n = 6,118) paid apps report collecting data to track users, only 2% (n = 643) paid apps report this practice on the App Store (Fig. 7; first plot; red bar 3).

Content Rating. Developers provide a *Content Rating* as part of the app metadata to indicate the age appropriateness of their apps. These ratings are reviewed by Apple [10] and used to enforce parental control features that restrict children from accessing the app [11]. We find that most apps that have a 4+ content rating on the App Store (81%; n = 419,762), while fewer apps have 9+ (3%; n = 16,687), 12+ (9%; n = 46,737), or 17+ (13%; n = 69,309) content ratings. Since privacy labels do not indicate the app's data practices specific to children, users must review the privacy policy to learn this information. Given parental control settings, an app with a 4+, 9+, or 12+ rating could be used by minors, although they may not be

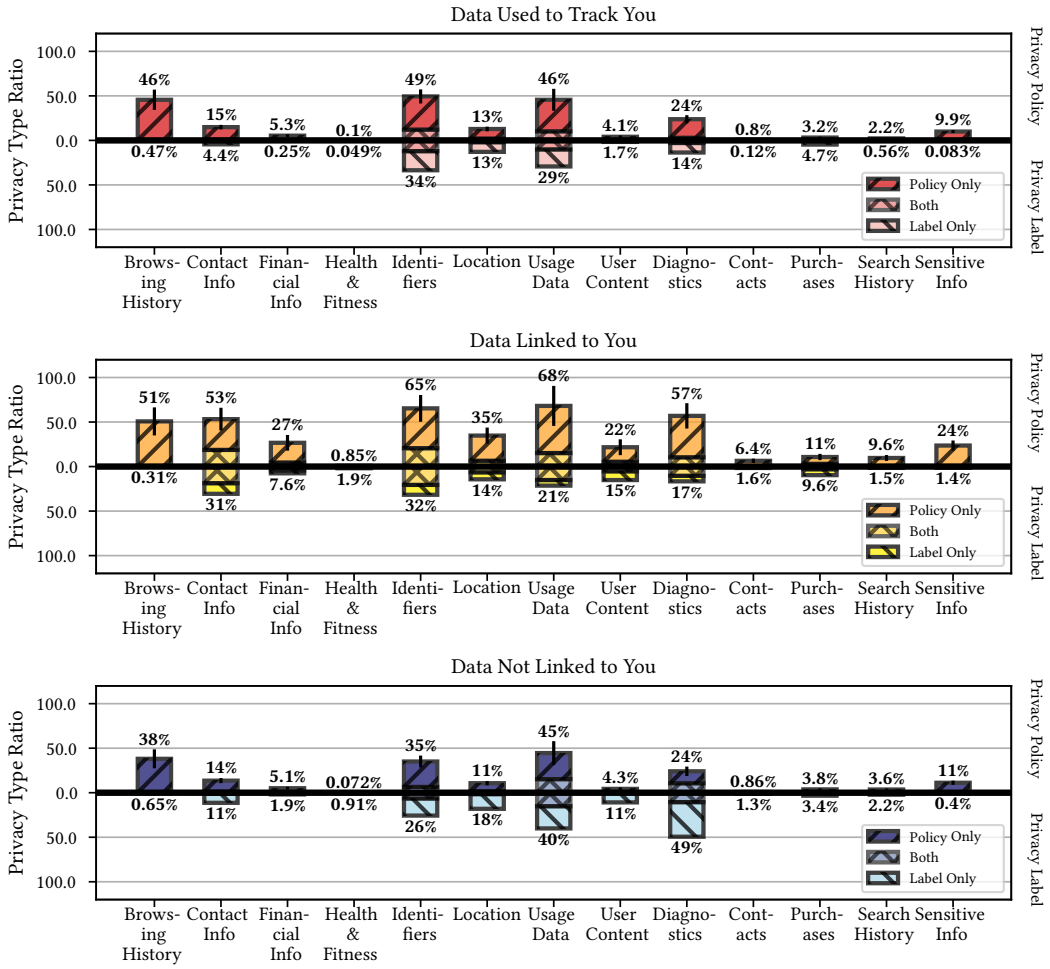

Figure 6: The ratios of data categories against privacy types. The denominator is the number of apps with the designated privacy type either in their privacy label or their privacy policy, i.e., 232,648 apps with *Data Used to Track You*, 419,504 apps with *Data Linked to You*, and 294,391 apps with *Data Not Linked to You*. The three privacy types shown here are *not* mutually exclusive.

the intended or target audience for the app. However, when an app specifically targets children, it is subject to additional regulations that may require parental consent. We fine-tuned language models to identify policy segments that address *International/Specific Audiences* and to identify further if the segment addresses *Children*, then compare this output to the content rating. Only 50% ($n = 179,168$) apps with a 4+ content rating also included a privacy policy segment that addresses data practices specific to children (Fig. 8; all plots; left-most bar). We were more likely to find similar policy segments for apps with different content ratings that can also be accessed by children, 9+ (65%; $n = 10,118$) and 12+ (59%; $n = 22,293$).

We further looked at content ratings for different privacy types associated with data collection. Considering apps with a 4+ content rating, roughly half had a policy explicitly addressing children across privacy types. While 20% ($n = 74,320$), 37% ($n = 134,076$), and 44% ($n = 159,512$) of the apps with a 4+ content rating declare in their privacy label that they collect data used to track users, linked to users, and not linked to users respectively, only 58% ($n = 43,536$), 51% ($n = 68,715$), and 54% ($n = 86,743$) of those apps also addressed children in their privacy policies (Fig. 8; plots 1, 2, & 3; bottom half; left-most bars; white overlay indicates addressing children).

While adding a 4+ content rating may help developers reach a wider audience, we only identified half of these apps consider data practices specific to children in the privacy policy. Additionally, even when apps address data collection from children in their privacy policies, these segments may absolve the developer of any responsibility. For example, ChowNow [24] is an app platform used by 3,182 different apps of local restaurants to receive online orders for takeout and delivery. ChowNow adds a content rating of 4+ to its apps on the App Store, making it accessible for children. Recall that developers choose a content rating according to Apple's guidelines [10]; Apple does not assign this value. However, ChowNow's privacy policy absolves themselves of the responsibility of dealing with data collected from children.

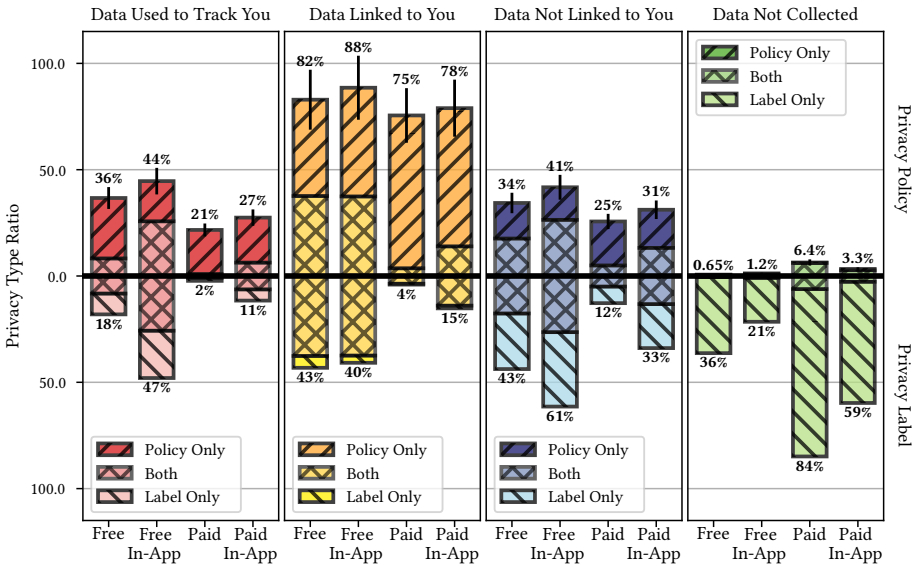

Figure 7: The ratios of app costs for each of the four privacy types. The denominator is the number of apps with the designated app cost that have a privacy label. Free apps are more likely than paid apps to collect data, including data used to track and linked to users. Please note that privacy types shown here are *not* mutually exclusive.

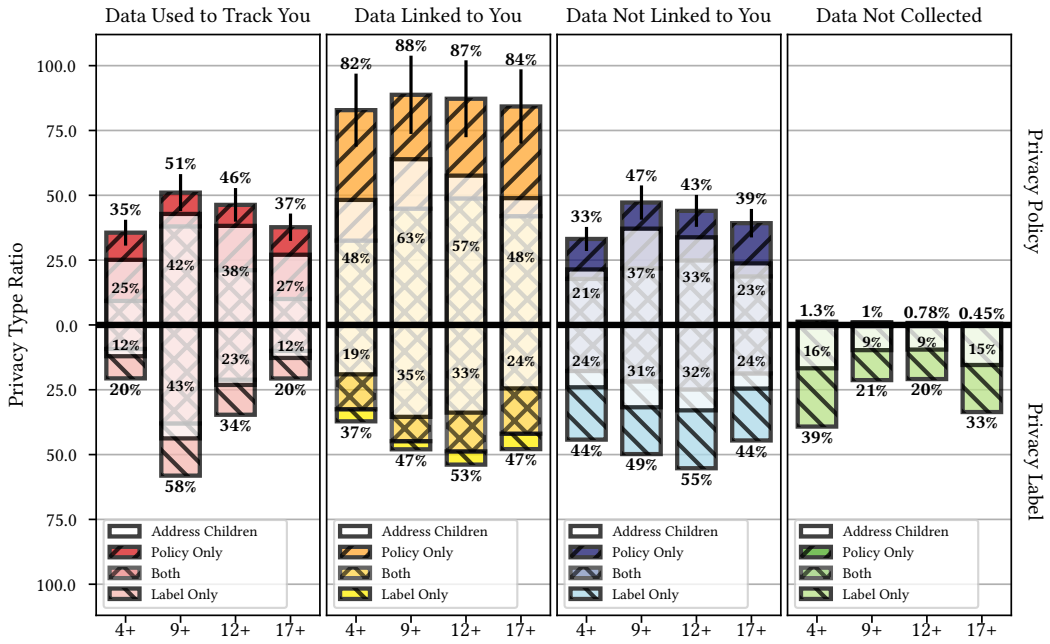

Figure 8: The ratios of the content ratings for each of the four privacy types, with an overlay (white bar) indicating the ratio of apps that also include a segment in their privacy policy, where they address privacy practices specific to children who engage with their services. The denominator is the number of apps with the designated content rating that have a privacy label. Please note that privacy types shown here are *not* mutually exclusive.

We acknowledge that our findings do not implicate the evaluated apps of violating COPPA [36], which, for example, allows PII collection with specific restrictions (e.g., geolocation) provided that developers do not use data for targeting/profiling of minors and that they obtain informed parental or legal tutor consent. We highlight the lack of declaration of data practices in privacy policies, especially when considered optional, and the need to ensure transparency across platforms. Additionally, third-party libraries offer options to help applications comply with COPPA regulations, but prior work has shown that they are often misconfigured [67].

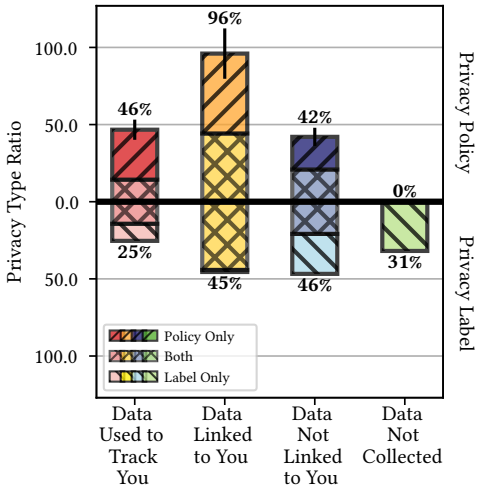

Figure 9: An overview of the privacy types associated with data collection on the App Store, from privacy labels and privacy policies, specific to apps whose policies are similar to templates. The denominator is the total number of such apps, i.e., 300,535 apps. Please note that the privacy types, except for *Data Not Collected*, are *not* mutually exclusive.

App Genre. We present an overview of our findings by app genre in Figure 11 in Appendix D. We find that *Games* apps are most likely to collect data used to track users (60%) and linked to users (59%) (Fig. 11; plots 1 & 2; bar 8). Notably, while 83% apps associated with the *Stickers* genre stated on the App Store that they do not collect *any* data, our analysis found that 66% apps collected data linked to users (Fig. 11; plots 2 & 4; bar 23). Apps under the *Stickers* genre are mostly lightweight apps made by smaller developers. They tend to have a 4+ content rating to reach a larger audience. They can include a few ad spaces and analytics libraries. Our intuition is that individual developers may not be aware of the data collection from third-party analytics and advertising libraries.

App Popularity. Since the App Store does not reveal the number of downloads for an app, we instead rely on the number of user ratings as a proxy for app popularity. To better represent their disclosures, we bin rating counts within the same order of magnitude in a single category and present our findings in Figure 10 in Appendix D. We find that with increased popularity, apps are more likely to declare data collection linked to users and used to track users. Our findings suggest that popular apps are more likely to be more thorough in their declaration of data collection practices because they receive more scrutiny.

Privacy Policy Templates. Templates offer a valuable solution for creating privacy policies, as they provide a ready-made framework for organizations to establish clear guidelines regarding handling user data. These pre-designed templates serve as a starting point that developers can customize to meet specific requirements and legal obligations. By utilizing templates, businesses can save time and effort by avoiding the need to create privacy policies from scratch. Additionally, templates help ensure compliance with privacy regulations by incorporating standard clauses and disclosures, ensuring that the privacy policy aligns with applicable laws such as GDPR or CCPA. However, it is essential for organizations to carefully review and tailor the template's content to accurately reflect their unique practices, guaranteeing transparency in communicating their privacy practices to users.

We evaluated the policies in our dataset to identify the use of templates. We searched for privacy policy templates and generators and gathered a list of services. We then visited each service and signed up, if required. We collected a set of 15 privacy policy templates, which we cleaned and divided into individual sentences. We represented the text in both the templates and the policies using in-domain word embeddings derived from privacy policies shared by Harkous et al. [46]. For each policy in our dataset, we conducted a comprehensive sentence-level comparison. We compared each sentence in a policy against every sentence in a template. We employed the cosine similarity metric to measure the semantic resemblance between two sentences. We deemed sentences similar if their cosine similarity exceeded a threshold of 0.8. We established a criterion to determine if a policy derived from a template: if over half of the sentences in a policy were similar to over half of the sentences in the template, we identified the policy as template-like.

We find that the privacy policies of 65% ($n = 306,404$) apps potentially use templates. We looked at the privacy labels these apps have declared on the App Store. Considering privacy types, 23%, 45%, 46%, and 31% of these apps declare *Data Used to Track You*, *Data Linked to You*, *Data Not Linked to You*, and *Data Not Collected* privacy types in their labels on the App Store (Fig. 9; bottom half; all bars). These findings align with all evaluated apps (see Figure 4). A majority of evaluated apps use template-like privacy policies. The use of templates possibly affects the discrepancies between the declaration of data collection practices in privacy labels and privacy policies. Templates often use generators, which offer significant value by ensuring developers thoroughly consider various data collection and sharing practices. These generators are similar to creating privacy labels on the App Store. However, it is essential to recognize that templates are not one-size-fits-all solutions. Developers must review and tailor policies derived from templates to accurately reflect individual apps' unique data collection practices. By carefully reviewing and customizing policies, developers can ensure the accuracy of their disclosures.

6 CASE STUDIES

Without Apple verifying privacy labels (and policies), their contents may not wholly clarify actual app practices. We present case studies of app behavior to shed light on the potential disparities between stated data collection practices and real-world app behavior. We use network requests captured from app usage to behavior developers report in labels and policies.

We used an iPhone running iOS 17.3.1 (released Feb 2024) with a man-in-the-middle (MiTM) proxy [26] to gather outgoing traffic to determine domains that apps accessed. We evaluated each app in the following manner: (1) We installed the app directly from the App Store. (2) We established a connection between the iPhone and the proxy. (3) Upon opening the app, the proxy captured and stored any outgoing requests made by the app. (4) After closing the app

and terminating the proxy connection, we deleted the app before evaluating the next app in the sequence.

We included 39 apps in the analysis, split between (a) 24 apps that declare data collection for advertising purposes in their privacy policies but not on their privacy labels and (b) 15 apps that declare a "Data Used to Track You" privacy type in their label on the App Store, but we could not infer such a practice from their privacy policies. We then compared the domains in the captured network requests against EasyList, EasyPrivacy, and WhoTracks.Me to identify trackers [4, 41, 50]. We provide an overview of our findings in Table 6 in Appendix C. The analysis presented in this study is an exploratory case study of 30 apps' network behavior. It should not be considered representative of the practices of all apps on the App Store.

The evaluated apps contact numerous tracking domains, with Facebook and Google being the most prominent. Further, developers often do not include analytics libraries within their purview of tracking, but guides from these libraries show that their practices are more nuanced [14, 43]. Additionally, inconsistencies between privacy disclosures and network traffic persist across different app categories. When privacy policies mention third-party libraries, they refer to third-party policies, resulting in incomplete inferences from an automated approach like the one presented in this work. We elaborate on potential explanations for our observations below.

Policy Reuse. Developers with multiple apps on the App Store reuse the privacy policies linked with individual apps. While this practice may result from using generic templates for some developers, organizations can also reuse these templates with multiple services. For example, different developer accounts publish Lexington Law and CreditRepair (#1 & #2 in Table 6), and the apps link to different privacy policies on the App Store. However, their privacy labels and privacy policies are identical. They are subsidiaries of the same organization, PGX Holdings Inc., and reuse declaration statements even if these statements apply to those subsidiaries. Developers must update templates to ensure accurate data collection practices, which can then reflect the accuracy of privacy labels.

Understanding Third Party Collection. When applications state in their privacy policies that they do not share data with third parties except to provide certain services (not including targeted advertising), it is possible that developers do not clearly understand or parse the nuances of data collection and sharing performed by integrated third parties. For example, Paypal, Crumbl, and Discord (#3, #9, #12 in Table 6) have policies covering data collection and sharing from third parties. To their credit, third-party libraries provide guidelines and disclosure links for developers to review before filling out their privacy labels and privacy policies (examples, [14, 43, 59, 74]). However, these guides include multiple caveats that can further complicate developers' understanding, requiring them to process against their use cases and translate into Apple's data collection definitions and requirements.

Understanding App Store requirements. Apple requires that developers declare all data collected in the app, including the practices of third-party partners, except for certain scenarios wherein disclosure is deemed optional [30]. While apps like Venmo, Southwest Airlines, Open Table, and Indeed (#1, #4, #6, #11 in Table 6) fill their privacy labels with multiple data categories under the *Data Linked to You* and *Data Not Linked to You* privacy types, they fail to do the same while declaring *Data Used to Track You*. Their privacy policies include statements highlighting third-party data collection and sharing for advertising and measurement purposes, indicating the developers' understanding of such activity. However, despite the App Store requiring the disclosure of all data collection practices, the developers' interpretation of optional caveats may affect their creation of privacy labels. For example, the period tracking app, Maya (#24 in Table 6), declared the sharing of *Usage Data* for tracking users, but the third-party libraries that it uses additionally collect and use identifiers and device information to track users [43, 59].

Understanding Apple's Definition of Tracking. Apple details practices that it considers to fall under *Tracking*, along with examples and caveats [30]. However, recent work has found that developers find it difficult to understand this definition and correctly declare data collection used to track users [55]. Apps like Axolochi, WebMD, and Food Network Magazine (#19, #21, and #22 in Table 6) acknowledge the use of tracking technologies in their privacy policies. However, the absence of similar declarations in privacy labels can stem from confusion around their understanding of Apple's definition of tracking. A recent study by Li et al. [55] showed that developers find it difficult to correctly identify data linked to users and data used to track users.

Next, we present possible reasons for discrepancies for apps with a *Data Used to Track You* privacy type in the App Store label but prove it challenging to automatically capture tracking practices from their privacy policies.

Non-exhaustive Policies. The privacy policies of Shake Shack, Kika Keyboard, Photo Prints CVS, Everpix, and FloatMe (#25, #26, #27, #28, #29 in Table 6) mention third party collection and sharing in terms of legal compliance and mergers/acquisitions. These privacy policies do not comprehensively cover all practices and data collection scenarios, making it difficult to identify such practices without ground truth.

Unclear Policy Statements. Even when developers declare third-party data collection and sharing in their privacy policies, such declaration is not explicit or clear to enable automatic detection and inference. The policies of Buffalo Wild Wings, The General Auto Insurance App, Conservative News (#30, #31, #32 in Table 6) include statements of sharing of information with "non-affiliated third parties", "vendors", "third party code and libraries", but do not make explicit the specific data categories collected and the use of this data for tracking, advertising, or advertising measurement.

Complex Formats. Being free-form documents, privacy policies do not need to be presented in standard, machine-parsable formats. While developers provide correct links to their policies on the App Store, we can only access the content of the policy behind a further link(s), as is the case with apps like McDonalds, Episode (#35, #36 in Table 6). Additionally, the policy for BrainBoom (#33 in Table 6) presents information in mixed formats, i.e., text and images, further complicating our ability to identify all practices. Finally, apps like JCPenney, Dosh, and CDL Prep Test (#37, #38,

#39 in Table 6) provide incorrect or broken links on the App Store, resulting in the extraction of incorrect from automated crawls.

7 DISCUSSION AND CONCLUSIONS

We analyzed 474,669 apps on the App Store, comparing the practices reported in privacy policies to those reported in privacy labels by performing automated NLP classification of the privacy policies. We find that most apps are likely under-reporting data collection practices in their privacy labels compared to their privacy policies. We find that almost all (97%) apps that indicate in their privacy labels that they do *not* collect *any* data engage in some form of data collection according to their privacy policy. Additionally, the privacy labels of 84% of paid apps indicate that they do not collect any data. In contrast, privacy policies suggest that the actual number may be closer to only 6.4% paid apps. Privacy policy analysis also reveals additional information about data practices not captured in privacy labels, including that most apps (81%) selected a 4+ content rating, but only 50% of these apps mention data collected from children in their privacy policies.

Ethics. The analysis and findings we present are based on publicly available data. We only mention popular apps (determined from rating counts) associated with large companies or developed by services with numerous associated apps. We reached out to Apple and shared our paper before publication. We encourage communication from developers and researchers to make use of our code and data to verify privacy labels.

In the remainder of this section, we discuss some of the implications of this analysis, such as the ground truth of privacy behavior when considering privacy labels or privacy policies. We also consider what factors likely lead to the misapplication of labels and recommendations for improving the current state.

Privacy Behavior Ground Truth. Since Apple's labels are not validated, we considered the privacy policies a reasonable reference point of comparison. However, it isn't easy to know the actual ground truth of privacy behavior, even if we fully dynamically and statically analyze every app. In this paper, we compare privacy labels against privacy policies as a point of comparison of the declaration of data practices across platforms. Privacy policies do not serve as ground truth for actual app behavior. While there are limitations to the approach we take in analyzing privacy policies using classifiers, the NLP methods of extracting free-form text levels get us closer to a viable understanding of data collection practices than the privacy labels, as currently used. We believe that this is the case for two reasons. First, classifier outputs introduce uncertainties that stem from the fact that policies are analyzed on a per-segment basis, so discussions of data aggregation or anonymization that occurs in one segment, separate from the data that is collected, might appear as data linking when it is, in fact, not linked. However, even with these statements, the app's behavior remains ambiguous according to the privacy policy regarding which specific data categories are aggregated or anonymized. Apps could often link data based on unique identifiers stated in other policy segments. Our observations suggest that developers mislabel many apps even after considering uncertainties from classier outputs. Second, there are also significant cases of under-reporting from classifiers due to how Apple links to privacy policies and the use of secondary privacy policies from third-party libraries. Many privacy policies link to other policies that we did not analyze. The App Store links also point to the developers' and not the specific apps' privacy policies. These policies usually address all services provided by the developer. For example, Subsplash [2] and ChowNow [24] affect thousands of apps, and it is unknown how the eventual customer uses that data and if policies reflect such scenarios.

Takeaway. We need improved notions of ground truth, which can dynamically identify data collection within apps at scale. However, even with their shortcomings, privacy policies provide a first-level check to identify discrepancies in privacy labels.

Source of Confusion Around Privacy Labels. It may also be that the processes for generating a privacy policy, including legal staff, are quite different from those selecting the labels, leaving the onus on the development team to make an accurate submission to the App Store. This split in responsibilities could confuse the kinds of data covered by the privacy label (as compared to what is in the policy) and what Apple would consider linked or not linked to users. For example, a recent study by Li et al. [55] showed that developers find it difficult to correctly identify data linked to users and data used to track users. Our results suggest that there is a large amount of mismatch in both data linked and not linked regarding the *Purposes*, where *App Functionality* and *Analytics* are particularly confusing, especially when apps may collect unique identifiers, as well as collecting *other* kinds of data that this should match to the *Other Purposes* category.

Takeaway. We argue that inaccurate labels are not necessarily the developers' fault but that better guidance and education are required to help them match app practices to labels.

Divergent Incentive Models. Privacy policies have become a standard and accepted part of notice and consent laws, and failure to provide an accurate and comprehensive privacy policy could lead to serious legal consequences. Companies are well incentivized to provide broad privacy policies that provide legal cover for their data collection practices in a way that protects them from any jeopardy, including hiring lawyers and other policy experts to craft and review them. Given their length and legal jargon, research shows that privacy policies are neither well understood [66] nor actively reviewed by most users [49]. In contrast, privacy labels are now forward-facing and published directly on the App Store without needing to follow any links to review. Recent results by Garg et al. [40] have even suggested that privacy labels can reduce app demand in cases of collecting sensitive information. The incentive for privacy labels may be an economic rather than a legal one, and these diverging incentive models may help explain some of the large differences we observed between privacy policies and privacy labels. This setup may change, and it is reasonable to consider that privacy labels should face the same regulatory scrutiny as privacy policies due to their role. One could also argue that Apple can expand privacy labels to include more explicit details about data collection behaviors, some of which may indeed be crucial to users for making meaningful and informed decisions about whether to install an app on their computing devices. However, we need balance as adding too much information contradicts the goal of

privacy labels to provide a succinct and readable description of the app behavior without needing to read the privacy policy.

Takeaway. Unfortunately, privacy labels appear to suffer from the transparency paradox [62]: the inherent conflict between the transparency of textual meaning and the transparency of data-handling practices.

Improved NLP Models for Privacy Labels. Classification approaches [46, 78, 82] offer much promise in helping to verify additional labeling of apps, like privacy labels. However, these approaches have several shortcomings as researchers did not design them for this task. Foremost, the analysis process is on a per-segment basis, which is helpful in inferring practices that policies completely describe in individual segments. However, policies often describe practices in parts that automated frameworks do not correctly capture across multiple segments. This shortcoming is partly due to the models' design and training data (OPP-115 dataset [78]), which researchers labeled on a per-segment basis. Additionally, given that services, including Google in Android [42], are adopting privacy labels more broadly, it may be time to update the models and training data to reflect privacy labels as the outcome. For example, the OPP-115 dataset could be re-annotated with privacy labels, forming the basis for new NLP models and more reliable tools to assist developers, researchers, and regulators better.

Takeaway. The community needs new datasets that align with the taxonomies used by Apple and Google. We also need stronger NLP approaches that can consider cross-segment contexts in privacy policies and thus comprehensively extract the nuances of data collection practices highlighted within the free-form text.

Regulation and Legal Compliance. Apple requires developers to create a single privacy label for all regions and all users of an app. The App Store does not allow developers to explicitly comply with region-specific (GDPR, CCPA) and age-specific (COPPA) laws. Instead, it encourages developers to create a single, universal label that is either too extensive or too sparse — neither version accurately represents a user's experience. Further, in the absence of vetting from Apple, the responsibility for accuracy solely lies with app developers. The existing structure of the ecosystem helps the App Store appear to care about user privacy but absolves Apple of responsibility for inaccuracy and disinformation.

Recommendations for Apple. With recent studies highlighting that privacy labels are hard to understand [55, 81], Apple could reconsider the taxonomy and descriptions of privacy labels. Additionally, Apple's lack of obvious vetting or regulation of the privacy labels may not incentivize the creation of accurate labels, particularly without any feedback to developers. Our imperfect framework can provide a first-level check for developers to consider more comprehensive arrays of labels for their apps. With Apple imposing a short embargo to review new apps before posting to the store, the platform could also incorporate some form of policy-based analysis into the review process.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful comments. This material is based upon work supported by the United States National Science Foundation under Grant Nos. 2247952 and 2247953.

REFERENCES

[1] 2020. Subsplash | Church mobile apps, websites, media, giving & more. https://www.subsplash.com/
[2] 2020. Subsplash Privacy Policy. https://www.subsplash.com/legal/privacy
[3] 2022. DappyTKeys. https://apps.apple.com/us/app/dappytkeys/id1536818077
[4] 2023. EasyList. https://easylist.to/
[5] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. 2014. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (Scottsdale, Arizona, USA) *(CCS '14)*. Association for Computing Machinery, New York, NY, USA, 674–689. https://doi.org/10.1145/2660267.2660347
[6] ALDI International Services GmbH & Co. oHG. 2023. ALDI USA. https://apps.apple.com/us/app/aldi-usa/id429396645
[7] Aldi US. 2023. Aldi: U.S. Privacy Policy. https://www.aldi.us/en/online-privacy-notice/
[8] Benjamin Andow, Samin Yaseer Mahmud, Wenyu Wang, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Tao Xie. 2019. PolicyLint: Investigating Internal Privacy Policy Contradictions on Google Play. In *28th USENIX Security Symposium (USENIX Security 19)*. USENIX Association, Santa Clara, CA, 585–602. https://www.usenix.org/conference/usenixsecurity19/presentation/andow
[9] Benjamin Andow, Samin Yaseer Mahmud, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Serge Egelman. 2020. Actions Speak Louder than Words: Entity-Sensitive Privacy Policy and Data Flow Analysis with PoliCheck. In *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, 985–1002. https://www.usenix.org/conference/usenixsecurity20/presentation/andow
[10] Apple. 2020. App Store Review Guidelines - Apple Developer. Retrieved 2022-05-15 from https://developer.apple.com/app-store/review/guidelines/
[11] Apple. 2022. Age Ratings - Apple Developer. Retrieved 2023-05-23 from https://developer.apple.com/help/app-store-connect/reference/age-ratings/
[12] Apple. 2022. Apple Catalog API. Retrieved 2022-10-11 from https://amp-api.apps.apple.com/v1/catalog/
[13] FamilyLife Apps. 2022. FamilyLife®. https://apps.apple.com/us/app/familylife/id903170704
[14] AppsFlyer. 2023. Preparing for the App Store review—nutrition labels. https://support.appsflyer.com/hc/en-us/articles/207032086-Preparing-for-the-App-Store-review-nutrition-labels
[15] David G Balash, Mir Masood Ali, Xiaoyuan Wu, Chris Kanich, and Adam J Aviv. 2022. Longitudinal Analysis of Privacy Labels in the Apple App Store. *arXiv preprint arXiv:2206.02658* (2022). https://doi.org/10.48550/arXiv.2206.02658
[16] Rebecca Balebako, Florian Schaub, Idris Adjerid, Alessandro Acquisti, and Lorrie Cranor. 2015. 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* (Denver, Colorado, USA) *(SPSM '15)*. Association for Computing Machinery, New York, NY, USA, 63–74. https://doi.org/10.1145/2808117.2808119
[17] Braincake. 2023. Privacy Policy. http://braincake.net/pregnancytracker_privacy.html
[18] Travis D. Breaux and Ashwini Rao. 2013. Formal analysis of privacy requirements specifications for multi-tier applications. In *2013 21st IEEE International Requirements Engineering Conference (RE)*. 14–23. https://doi.org/10.1109/RE.2013.6636701
[19] Duc Bui, Yuan Yao, Kang G. Shin, Jong-Min Choi, and Junbum Shin. 2021. Consistency Analysis of Data-Usage Purposes in Mobile Apps. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security* (Virtual Event, Republic of Korea) *(CCS '21)*. Association for Computing Machinery, New York, NY, USA, 2824–2843. https://doi.org/10.1145/3460120.3484536
[20] Center for Food Safety and Applied Nutrition. 2022. *How to Understand and Use the Nutrition Facts Label*. U.S. Food and Drug Administration. https://www.fda.gov/food/new-nutrition-facts-label/how-understand-and-use-nutrition-facts-label Publisher: FDA.
[21] Yi Chen, Mingming Zha, Nan Zhang, Dandan Xu, Qianqian Zhao, Xuan Feng, Kan Yuan, Fnu Suya, Yuan Tian, Kai Chen, XiaoFeng Wang, and Wei Zou. 2019. Demystifying Hidden Privacy Settings in Mobile Apps. In *2019 IEEE Symposium on Security and Privacy (SP)*. 570–586. https://doi.org/10.1109/SP.2019.00054
[22] ChowNow. 2022. Bagelman. https://apps.apple.com/us/app/bagelman/id1254203081
[23] ChowNow. 2022. El Charrito. https://apps.apple.com/us/app/el-charrito/id1080457082
[24] Chownow. 2022. Online Food Ordering System for Restaurants. https://get.chownow.com/
[25] Chownow. 2022. Privacy Policy - ChowNow. https://get.chownow.com/privacy-policy
[26] Aldo Cortesi, Maximilian Hils, Thomas Kriechbaumer, and contributors. 2010–. mitmproxy: A free and open source interactive HTTPS proxy. https://mitmproxy.org/ [Version 9.0].

[27] Lorrie Faith Cranor. 2012. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *J. on Telecomm. & High Tech. L.* 10 (2012), 273. http://jthtl.org/content/articles/V10I2/JTHTLv10i2_Cranor.PDF

[28] Lorrie Faith Cranor, Candice Hoke, Pedro Leon, and Alyssa Au. 2014. *Are They Worth Reading? An In-Depth Analysis of Online Advertising Companies' Privacy Policies.* SSRN Scholarly Paper ID 2418590. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=2418590

[29] Credit Karma, Inc. 2022. Credit Karma on the App Store. https://apps.apple.com/us/app/credit-karma/id519817714

[30] Apple Developer. 2022. App Privacy Details - App Store. https://developer.apple.com/app-store/app-privacy-details/

[31] Apple Developer. 2022. App privacy labels now live on the App Store - Latest News - Apple Developer. https://developer.apple.com/news/?id=3wann9gh

[32] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. 2020. Ask the Experts: What Should Be on an IoT Privacy and Security Label?. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Jose, CA, USA, 447–464. https://doi.org/10.1109/SP40000.2020.00043 ISSN: 2375-1207.

[33] Pardis Emami-Naeini, Janarth Dheenadhayalan, Yuvraj Agarwal, and Lorrie Faith Cranor. 2021. 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)*. IEEE, San Francisco, CA, USA, 519–536. https://doi.org/10.1109/SP40001.2021.00112

[34] Steven Englehardt and Arvind Narayanan. 2016. Online Tracking: A 1-Million-Site Measurement and Analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (Vienna, Austria) *(CCS '16)*. Association for Computing Machinery, New York, NY, USA, 1388–1401. https://doi.org/10.1145/2976749.2978313

[35] Federal Trade Commission. 2000. Privacy Online: Fair Information Practices in the Electronic Marketplace: A Federal Trade Commission Report to Congress. https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf

[36] Federal Trade Commission. 2013. Children's Online Privacy Protection Rule (COPPA). https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa

[37] Federal Trade Commission. 2022. FTC Safeguards Rule: What Your Business Needs to Know. https://www.ftc.gov/business-guidance/resources/ftc-safeguards-rule-what-your-business-needs-know

[38] Fitness Labs. 2023. Pregnancy Tracker: Baby Bump. https://apps.apple.com/us/app/pregnancy-tracker-baby-bump/id1453373942

[39] Jack Gardner, Yuanyuan Feng, Kayla Reiman, Zhi Lin, Akshath Jain, and Norman Sadeh. 2022. Helping Mobile Application Developers Create Accurate Privacy Labels. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 212–230. https://doi.org/10.1109/EuroSPW55150.2022.00028

[40] Rajiv Garg and Rahul Telangb. 2022. Impact of App Privacy Label Disclosure on Demand: An Empirical Analysis. *Workshop on the Economics of Information Security (WEIS)* (2022). https://weis2022.econinfosec.org/wp-content/uploads/sites/10/2022/06/weis22-telang.pdf

[41] Ghostery. 2023. WhoTracks.me - Bringing Transparency to Online Tracking. https://whotracks.me/

[42] Google. 2020. Provide information for Google Play's Data safety section - Play Console Help. Retrieved 2022-05-15 from https://support.google.com/googleplay/android-developer/answer/10787469

[43] Google. 2023. Privacy Disclosures Policy. https://support.google.com/analytics/answer/7318509

[44] Catherine Han, Irwin Reyes, Amit Elazari Bar On, Joel Reardon, Álvaro Feal, Serge Egelman, Narseo Vallina-Rodriguez, et al. 2019. Do you get what you pay for? comparing the privacy behaviors of free vs. paid apps. In *Workshop on Technology and Consumer Protection (ConPro 2019), in conjunction with the 39th IEEE Symposium on Security and Privacy, 23 May 2019, San Francisco, CA, USA*. IEEE, San Francisco, CA, USA, 7 pages. https://www.ieee-security.org/TC/SPW2019/ConPro/papers/han-conpro19.pdf

[45] Catherine Han, Irwin Reyes, Álvaro Feal, Joel Reardon, Primal Wijesekera, Narseo Vallina-Rodriguez, Amit Elazari Bar On, Kenneth A. Bamberger, and Serge Egelman. 2020. The Price is (Not) Right: Comparing Privacy in Free and Paid Apps. *Proc. Priv. Enhancing Technol.* 2020, 3 (2020), 222–242. https://doi.org/10.2478/popets-2020-0050

[46] Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G. Shin, and Karl Aberer. 2018. Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning. In *27th USENIX Security Symposium (USENIX Security 18)*. USENIX Association, Baltimore, MD, 531–548. https://www.usenix.org/conference/usenixsecurity18/presentation/harkous

[47] HyperBeard Games. 2023. Privacy Policy: HyperBeard Games. https://docs.google.com/document/d/1bJpQtjxoxmf1leATVgnRbZWkV9W-iDngLx_pTsvu8Vk/edit

[48] HyperBeard Inc. 2023. Axolochi. https://apps.apple.com/us/app/axolochi/id1432184360

[49] Carlos Jensen and Colin Potts. 2004. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 471–478.

[50] Arjaldo Karaj, Sam Macbeth, Rémi Berson, and Josep M. Pujol. 2019. WhoTracks.Me: Shedding light on the opaque world of online tracking. https://doi.org/10.48550/arXiv.1804.08959 arXiv:1804.08959 [cs].

[51] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W Reeder. 2009. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*. 1–12.

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

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

[54] Konrad Kollnig, Anastasia Shuba, Max Van Kleek, Reuben Binns, and Nigel Shadbolt. 2022. Goodbye Tracking? Impact of IOS App Tracking Transparency and Privacy Labels. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 508–520. https://doi.org/10.1145/3531146.3533116

[55] Tianshi Li, Kayla Reiman, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I. Hong. 2022. Understanding Challenges for Developers to Create Accurate Privacy Nutrition Labels. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 588, 24 pages. https://doi.org/10.1145/3491102.3502012

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

[57] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. https://doi.org/10.48550/arXiv.1907.11692 arXiv:1907.11692 [cs].

[58] Aleecia M. McDonald and Lorrie Faith Cranor. 2009. The Cost of Reading Privacy Policies. *HeinOnline* 4, 3 (2009), 543–568. https://heinonline.org/HOL/P?h=hein.journals/isjlpsoc4&i=563

[59] Meta. 2022. Resources for Completing App Store Data Practice Questionnaires for Apps That Include the Facebook or Audience Network SDK. https://developers.facebook.com/blog/post/2022/07/18/resources-for-completing-app-store-data-practice-questionnaires-apps-facebook-or-audience-network-sdk/ publisher: Meta.

[60] Mozilla. 2020. Readability.js. https://github.com/mozilla/readability

[61] Collins W. Munyendo, Yasemin Acar, and Adam J. Aviv. 2022. Desperate Times Call for Desperate Measures: User Concerns with Mobile Loan Apps in Kenya. In *2022 IEEE Symposium on Security and Privacy (SP)*. 2304–2319. https://doi.org/10.1109/SP46214.2022.9833779

[62] Helen Nissenbaum. 2011. A contextual approach to privacy online. *Daedalus* 140, 4 (2011), 32–48.

[63] Port of Google's language-detection library to Python. 2021. langdetect. https://github.com/Mimino666/langdetect

[64] Office of the Attorney General of California. 2018. California Consumer Privacy Act (CCPA).

[65] Paytronix Systems Inc. 2022. HomeState, A Texas Kitchen. https://apps.apple.com/us/app/homestate-a-texas-kitchen/id925093380

[66] Joel R Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T Graves, Fei Liu, Aleecia McDonald, Thomas B Norton, and Rohan Ramanath. 2015. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Tech. LJ* 30 (2015), 39.

[67] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, and Serge Egelman. 2018. "Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale. *Proceedings on Privacy Enhancing Technologies* 3 (2018), 63–83. https://petsymposium.org/2018/files/papers/issue3/popets-2018-0021.pdf

[68] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. 2015. A Design Space for Effective Privacy Notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Ottawa, 1–17. https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub

[69] Rocky Slavin, Xiaoyin Wang, Mitra Bokaei Hosseini, James Hester, Ram Krishnan, Jaspreet Bhatia, Travis D. Breaux, and Jianwei Niu. 2016. Toward a Framework for Detecting Privacy Policy Violations in Android Application Code. In *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*. 25–36. https://doi.org/10.1145/2884781.2884855

[70] Mukund Srinath, Shomir Wilson, and C Lee Giles. 2021. Privacy at Scale: Introducing the PrivaSeer Corpus of Web Privacy Policies. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 6829–6839. https://doi.org/10.18653/v1/2021.acl-long.532

[71] Staff FTC. 2011. Protecting consumer privacy in an era of rapid change–A proposed framework for businesses and policymakers. *Journal of Privacy and Confidentiality* 3, 1 (2011), 112 pages. https://www.ftc.gov/reports/protecting-consumer-privacy-era-rapid-change-recommendations-businesses-policymakers

[72] Mozilla Support. 2022. Firefox Reader View for clutter-free web pages | Firefox Help. https://support.mozilla.org/en-US/kb/firefox-reader-view-clutter-free-web-pages

[73] The Alan Turing Institute. 2018. ReadabiliPy. https://github.com/alan-turing-institute/ReadabiliPy

[74] Unity. 2023. Apple privacy survey for Unity Ads. https://docs.unity.com/ads/en/manual/ApplePrivacySurvey

[75] Walmart. 2022. Walmart - Shopping & Grocery. https://apps.apple.com/us/app/walmart-shopping-grocery/id338137227

[76] WebMD. 2023. WebMD Privacy Policy. https://www.webmd.com/about-webmd-policies/about-privacy-policy

[77] WebMD. 2023. WebMD: Symptom Checker. https://apps.apple.com/us/app/webmd-symptom-checker/id295076329

[78] Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N. Cameron Russell, Thomas B. Norton, Eduard Hovy, Joel Reidenberg, and Norman Sadeh. 2016. 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)*. Association for Computational Linguistics, Berlin, Germany, 1330–1340. https://doi.org/10.18653/v1/P16-1126

[79] Yue Xiao, Zhengyi Li, Yue Qin, Xiaolong Bai, Jiale Guan, Xiaojing Liao, and Luyi Xing. 2022. Lalaine: Measuring and Characterizing Non-Compliance of Apple Privacy Labels at Scale. *CoRR* abs/2206.06274 (2022). https://doi.org/10.48550/arXiv.2206.06274 arXiv:2206.06274

[80] Le Yu, Xiapu Luo, Xule Liu, and Tao Zhang. 2016. Can We Trust the Privacy Policies of Android Apps?. In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. 538–549. https://doi.org/10.1109/DSN.2016.55

[81] Shikun Zhang, Yuanyuan Feng, Yaxing Yao, Lorrie Faith Cranor, and Norman Sadeh. 2022. How Usable Are iOS App Privacy Labels? *Proceedings on Privacy Enhancing Technologies* 4 (2022), 204–228.

[82] Sebastian Zimmeck, Peter Story, Daniel Smullen, Abhilasha Ravichander, Ziqi Wang, Joel R. Reidenberg, N. Cameron Russell, and Norman M. Sadeh. 2019. MAPS: Scaling Privacy Compliance Analysis to a Million Apps. *Proceedings on Privacy Enhancing Technologies* 2019 (2019), 66 – 86.

[83] Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven M. Bellovin, and Joel Reidenberg. 2017. Automated Analysis of Privacy Requirements for Mobile Apps. In *Proceedings 2017 Network and Distributed System Security Symposium*. https://doi.org/10.14722/ndss.2017.23034

A REIMPLEMENTING AND TRAINING THE CLASSIFICATION FRAMEWORK

Hierarchical Structure. Our implementation of the framework closely follows that used for Polisis [46], which in-turn relies on the OPP-115 corpus [78]. It comprises a *hierarchical, multi-level* set of classifiers. The framework takes a paragraph-length segment of text as input, and passes it to a Segment Classifier to first determine one or more high-level data practices addressed in the segment. These data practices may look like, *First Party Collection/Use*, *Data Security*, *International/Specific Audiences*, etc. The framework further passes the segment through multiple Attribute Classifiers, each of which determine one or more attribute values relevant to the data practice determined by the Segment Classifier. For example, if the segment addresses *First Party Collection/Use*, the *Does/Does Not* Attribute Classifier determines if the policy claims to engage in data collection, the *Identifiability* Attribute Classifier determines if the data collection can be linked to the user, the *Purpose* Attribute Classifier determines the stated reason for data collection, and the *Personal Information Type* Attribute Classifier determines the data categories addressed in the segment. The framework classifies one segment of the policy at a time, and the data practices addressed in the entire policy are determined by collating results from all segments. An overview of this structure is provided in Figure 3.

Training Dataset. The Online Privacy Policies (OPP-115) dataset, created by Wilson et al. [78], is an annotated dataset of 115 privacy policies. Each policy is divided into paragraph-length *segments*, and manually annotated by law school students. Each segment was annotated at two levels – first, the annotator chose one or more high-level data practices that the segment addresses (e.g., First Party Collection/Use, Third Party Collection/Sharing); then, depending on the initial selections, they annotated segments with multiple attribute-value pairs (e.g., information_type: financial, purpose:advertising, etc.). Overall, the task covered 10 data practices and 20 associated attributes, with 138 distinct values across attributes. We developed one classifier to determine high-level data practices addressed in a segment, followed by a classifier each for the different attributes associated with the identified data practice.

Train-Test Split. For each attribute, we collected all segments that had a relevant annotation for the attribute in the OPP-115 dataset. We then performed a separate 80-20 train-test split for each collection of segments belonging to an attribute. In this aspect, we differed from Harkous et al. [46], who instead set 65 of the 115 policies aside for training, and used relevant segments from these 65 policies to train all attribute classifiers – a choice that would have resulted in varied amounts of training data being used for each attribute.

Evaluation Metrics. The authors of PrivBERT [70] presented an example of fine-tuning a segment classifier using the OPP-115 corpus, in which they manually tuned the hyperparameters used to train the model. We followed a similar approach to develop each classifier. Table 5 presents the evaluation reports for the classifier's precision, recall, and F1 scores on an unseen test set. Following the practice highlighted by Harkous et al. [46], we evaluate each classifier's ability to detect both the *presence* and *absence* of an attribute in a given text segment. Additionally, since the OPP-115 corpus is old, we additionally manually evaluated classifier outputs on randomly sampled segments of Apple App Store policies, which we also report in Table 5. For each attribute, we randomly sampled 25 segments for which the classifier reported the presence of an attribute and also sampled 25 segments for which it reported the absence of an attribute. In this manner, we cover 50 segments each for 35 attributes across privacy policies. Table 5 also compares the classification reports for implementing the Polisis CNN-based approach against the performance of the fine-tuned BERT-based models. Finally, to verify our mapping of classifier outputs to privacy label attributes, two researchers randomly sampled and manually evaluated the outputs for 25 instances of each label output (see Table 4).

Table 4: Manual Verification of Mapping

Attribute	# Accurate (25 samples)
Privacy Types:	
Data Used to Track You	22
Data Linked to You	23
Data Not Linked to You	21
Data Not Collected	25
Purposes:	
App Functionality	23
Analytics	25
Product Personalization	25
Third-party Advertising	24
Developers Advertising	25
Other Purposes	25
Data Categories:	
Browsing History	24
Contact Info	25
Financial Info	25
Health & Fitness	25
Identifiers	25
Location	25
Usage Data	25
User Content	24
Diagnostics	25
Contacts	25
Purchases	25
Search History	24
Sensitive Info	25

B CASE STUDIES OF PRIVACY POLICIES

To further provide an understanding of the differences between policies and labels, we present a few interesting examples of popular apps and their privacy policies.

Subsplash. A platform that develops and integrates multiple church services, including donations, memberships, and services, Subsplash [1] is used by 8,015 apps of local churches on the App Store (examples, [3, 13]).

All of the hosted apps link to the same privacy policy [2] and share the same privacy label, i.e., a *Data Not Linked to You* label, which states that the app collects *Usage Data* for *Analytics*, and *Diagnostics* data for *App Functionality*. Recall that the *Data Not Linked to You* privacy type indicates that the data that is collected is aggregated or anonymized. Subsplash's policy states that they collect *Contact Info*, *Financial Info*, *Purchases*, none of which are included in their privacy label. A snippet from their policy is provided below.

> *When you interact with Subsplash, we may collect personal information relevant to the situation, such as your name, mailing address, phone number, email address, and contact preferences; your credit card information and information about the Subsplash products you own, such as their serial numbers and date of purchase; and information relating to a support or service issue.*

The apps additionally collects *Location*, and *Contacts* as stated in different segments but not included in the apps' privacy label.

At the same time, there are some examples of the structure of the privacy policy that may lead Polisis classifiers to under- or over-represent some behaviors. One example is the treatment of anonymization of data. A single segment highlighting anonymization but does not specify which data types are anonymized.

> *Subsplash may use aggregated and anonymized forms of personal information for a variety of purposes, including, but not limited to, analyzing usage trends, fraud detection, and development of new Services.*

As a result, Polisis is unable to match the data collection practice to anonymous linking and would classify most of the data collected by the app as *linked* rather than *not linked*. At the same time, since the policy is unclear on this point, it is difficult to fully know the data practices and if the labels are correct on this matter.

Another example involves the format of Subsplash's privacy policy which includes some data collection practices in varied visual formats, i.e., a table that includes different categories of data, examples of data types, and a column that states whether or not the stated data is collected. However, this table is implemented using `<div>` tags around each cell. The `readability` library interprets each of the cells as a separate paragraph, and makes it difficult to interpret the data presented here, potentially under-reporting some behavior as the segments are less complete.

ChowNow. ChowNow [24] is an app platform used by 3,182 different apps of local restaurants to receive online orders for takeout and delivery (examples, [22, 23, 65]).

All apps using the ChowNow platform link to the same privacy policy [25] and apply the same privacy label. The label indicates that all data collection is not linked, indicating that the collected data is aggregated or anonymized. However, ChowNow's privacy policy states that they use contact information to manage user accounts and inform users about products through *"electronic marketing communications"*. They also state that they use billing information, including card numbers, expiration date, security code, and billing address to process orders. Neither of these services can be provided in an anonymized manner, but the privacy labels lack a *Data Linked to You* category.

ChowNow's privacy policy also states that they share information with advertisers, but their label does not include a *Data Used to Track You* label. Additionally, the information that they share is mentioned as *Other Information*, making it difficult for the Polisis framework to identify the data categories shared with third party services. The relevant snippet is provided below.

> *We share Other Information about your activity in connection with your use of the Services with third-party advertisers and remarketers for the purpose of tailoring, analyzing, managing, reporting, and optimizing advertising you see on the Platforms, the Websites, the Apps, and elsewhere.*

ChowNow adds a content rating of 4+ to its apps on the App Store, making it accessible for children. Recall that developers choose a content rating according to Apple's guidelines [10]; this

value is not assigned by Apple. However, ChowNow's privacy policy absolve themselves of the responsibility of dealing with data collected from children, instead placing the burden of preventing such data collection on parents, guardians, and the children themselves. The relevant snippet is provided below.

> *We do not knowingly collect personal information from children under the age of 13 through the Services. If you are under 13, please do not give us any personal information. We encourage parents and legal guardians to monitor their children's Internet usage and to help enforce our Privacy Policy by instructing their children to never provide us personal information without their permission. If you have reason to believe that a child under the age of 13 has provided personal information to us, please contact us, and we will endeavor to delete that information from our databases.*

Walmart. A popular shopping and grocery delivery app with 6.6M user ratings, Walmart [75] provides a large number of privacy labels on the App Store, which includes an extensive list of data categories across three privacy types, *Data Used to Track You*, *Data Linked to You*, and *Data Not Collected*.

Apple's description of sensitive information covers a list of example data types that are considered sensitive, providing a general overview of possible values. Walmart's privacy label does *not* state that it collects *Sensitive Info*, which users may expect from a shopping and grocery delivery app. However, Walmart states in their privacy policy that they collect (i) demographic data, (ii) background & criminal information, and (iii) audio, visual and other sensory information, all of which Apple may consider sensitive information.

Credit Karma. A popular finance app with 5.4M user ratings on the App Store, Credit Karma [29] does not use a *Data Used to Track You* label on the App Store despite stating in their policy that they share personal information with *"other companies, lawyers, credit bureaus, agents, government agencies, and card associations in connection with issues related to fraud, credit, defaults, or debt collection"*.

We also observed that multiple privacy policies, including others previously mentioned in this section, ask users to refer to the policies of third party providers that they use within their services. An example snippet from Credit Karma's policy is provided below.

> *We may use third party API services, such as YouTube and Twilio, for certain product features. If you choose to use those features, you acknowledge and agree that you are also bound by the third party's privacy policy, such as Google's Privacy Policy for YouTube API services. You may manage your YouTube API data by visiting Google's security settings page at https://security.google.com/settings/security/permissions. For more information about Twilio's privacy practices, please visit https://www.twilio.com/legal/privacy.*

This practice not only increases the burden of gathering additional information for users, but it also makes it difficult for Polisis to infer potentially missing information included in these additional external policies. As a result, the analysis of Credit Karma and similar apps may be a lower bound of the true privacy related behavior.

Aldi. A popular grocery store in the United States, Aldi, has an app available on the App Store, which is ranked #59 in the Shopping category [6]. The app offers a wide range of features, enabling users to conveniently order groceries, schedule deliveries or pickups, and make secure payments for their purchases. According to their privacy policy [7], Aldi collects (1) payment information (such as credit or debit card or EBT number, security code, expiration date and billing address); (2) shopping list and purchase history information. It is worth noting, however, that their privacy label on the App Store does not include corresponding entries highlighting their collection of *Financial Info* and *Purchase History*.

Axolochi. A popular application under the *Games* category, Axolochi is ranked #78 in the *Trivia* sub-category [48]. The app's privacy policy [47] states the *automatic* collection of various identifiers, such as a unique user ID, IP address, device IDs, hardware or operating system-based identifiers, and identifiers assigned to user accounts. Surprisingly, the app's privacy label on the App Store does not include the *Identifiers* data category.

Furthermore, Axolochi offers in-app purchases for users. According to their privacy policy, when users make in-app purchases, the app collects ZIP or postal codes along with "the amount of the transaction and records of purchases" made by the user. However, it is worth noting that the privacy label on the App Store does not feature corresponding entries for *Physical Address* or *Purchase History*. This discrepancy may limit the visibility and transparency of the app's data practices, potentially leaving users with incomplete information regarding the collection and usage of their personal data within the app.

WebMD.. A widely known health-related service, WebMD hosts a flagship symptom checker app on the App Store [77]. Their privacy policy [76] explicitly mentions the collection of information from third-party vendors for targeted advertising purposes.

> *Our ad network vendors use technologies to collect information about your activities on the WebMD Sites and* ***in our flagship WebMD App*** *to provide you cookie-based* ***targeted advertising*** *on our WebMD Sites and on third party websites based upon your* ***browsing activity and your interests****.*

Surprisingly, the app does not include a specific privacy type entry for *Data Used to Track You* in their privacy label. This absence in the privacy label highlights an instance of inconsistency in declaration of data collection practices across disclosures.

Pregnancy Tracker. The pregnancy tracking app developed by Fitness Labs has concerning discrepancies between its privacy label on the App Store [38] and its privacy policy [17]. The app's privacy label only includes a *Data Not Linked to You* privacy type, mentioning the collection of *Usage Data* and *Diagnostics* data categories. However, the privacy policy reveals a much broader scope of data collection. The policy states: they may collect personal information such as name, address, email address, phone numbers, payment information (credit or debit card), and other demographic information that can identify individuals or enable contact.

> *We may collect information about you such as: personal information including, for example, your name; home or business address; e-mail address; telephone, wireless or fax number; short message service or text message address or other wireless device address; instant messaging address; credit or debit card or other payment information; demographic information or other information that may* ***identify you as an individual*** *or allow online or offline contact with you as an individual.*

Unfortunately, the app's privacy label fails to include the *Data Linked to You* privacy type or indicate the collection of multiple data categories, including *Identifiers*, *Financial Information*, *Contact Information*, and *Sensitive Information*.

C NETWORK TRAFFIC COLLECTION

We provide an overview of the analysis of 39 apps in Table 6.

D ADDITIONAL TABLES AND FIGURES

We include additional tables and figures here. Figure 10 provides an overview of our findings based on apps' popularity. Figure 11 presents our findings based on app genres. Table 7 details overlaps and discrepancies in disclosures across data categories in privacy labels and policies.

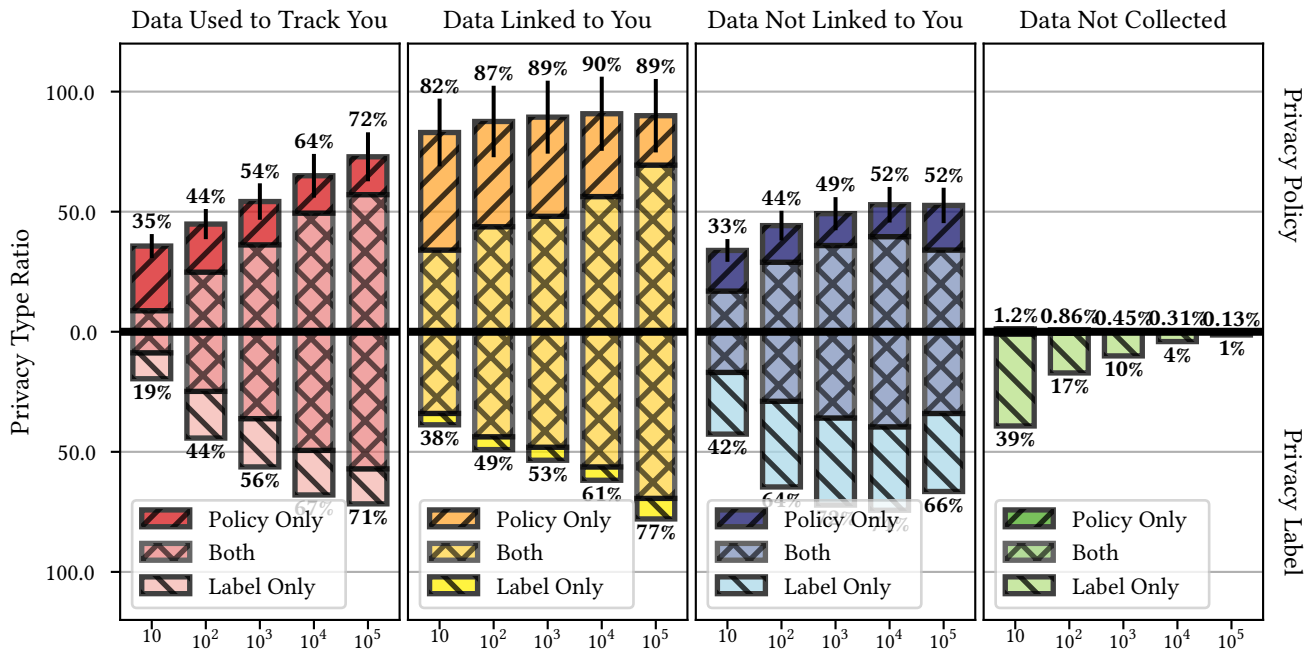

Figure 10: The ratios of the rating counts for each of the four *Privacy Types*. The denominator is the number of apps with the designated rating counts that have a privacy label. Apps with a larger number of user ratings are more likely to collect data, including data used to track users. Ratings counts are not localized metadata and apps with low ratings counts in the US region may have higher counts elsewhere.

Table 5: Classification results for the attributes that were used in the creation of Privacy Labels.

Classifier Output	PrivBERT Classification Report				Manual Check on New Segments		Polisis Classification Report		
	Precision	Recall	F1	Support (Presence|Absence)	Presence	Absence	Precision	Recall	F1
Segment Classifier									
First Party Collection/Use	0.89	0.88	0.89	298|460	25/25	24/25	0.80	0.79	0.80
Third Party Collection/Sharing	0.92	0.92	0.92	258|500	24/25	23/25	0.88	0.85	0.86
International and Specific Audiences	0.97	0.97	0.97	59|699	25/25	23/25	0.97	0.95	0.96
Average	0.93	0.92	0.93				0.88	0.86	0.87
Identifiability									
Identifiable	0.94	0.91	0.92	115|54	25/25	20/25	0.75	0.76	0.75
Aggregated or Anonymized	0.96	0.97	0.96	59|110	20/25	25/25	0.85	0.85	0.80
Average	0.95	0.94	0.94				0.80	0.80	0.80
Does/Does Not									
Does Not	0.87	0.86	0.86	47|393	25/25	25/25	0.91	0.80	0.84
Does	0.78	0.83	0.81	428|12	25/25	25/25	0.74	0.66	0.70
Average	0.83	0.85	0.84				0.82	0.73	0.77
Purpose									
Additional Service/Feature	0.83	0.86	0.84	103|399	25/25	24/25	0.82	0.79	0.80
Advertising	0.95	0.93	0.94	125|377	24/25	25/25	0.87	0.84	0.86
Analytics/Research	0.89	0.90	0.89	88|414	24/25	25/25	0.86	0.85	0.85
Basic Service/Feature	0.84	0.84	0.84	135|367	25/25	22/25	0.80	0.80	0.80
Legal Requirement	0.90	0.87	0.89	35|467	25/25	25/25	0.92	0.83	0.87
Marketing	0.86	0.85	0.86	123|379	25/25	25/25	0.84	0.82	0.83
Merger	0.94	1.00	0.97	13|489	25/25	25/25	1.00	0.88	0.93
Personalization	0.88	0.85	0.86	70|432	25/25	23/25	0.86	0.80	0.82
Service Operation and Security	0.90	0.90	0.90	37|465	25/25	23/25	0.86	0.81	0.83
Unspecified	0.79	0.79	0.78	227|275	23/25	25/25	0.81	0.73	0.76
Average	0.88	0.88	0.87				0.86	0.81	0.83
Personal Information Type									
Computer Information	0.88	0.94	0.91	30|483	25/25	24/25	0.94	0.91	0.92
Contact	0.94	0.91	0.92	129|384	24/25	25/25	0.91	0.90	0.91
Cookies and Tracking Elements	0.99	0.96	0.98	93|420	23/25	20/25	0.95	0.87	0.91
Demographic	0.94	0.89	0.92	47|466	24/25	25/25	0.90	0.86	0.88
Financial	0.85	0.77	0.81	39|474	25/25	25/25	0.94	0.90	0.92
Generic Personal Information	0.84	0.83	0.83	196|317	25/25	25/25	0.82	0.81	0.81
Health	0.91	0.95	0.93	10|503	23/25	25/25	0.95	0.66	0.74
IP Address and Device IDs	0.88	0.91	0.89	36|477	25/25	24/25	0.97	0.89	0.92
Location	0.90	0.75	0.81	14|460	25/25	24/25	0.91	0.85	0.88
Personal Identifier	0.67	0.67	0.67	14|499	25/25	24/25	0.95	0.77	0.83
Social Media Data	0.72	0.80	0.75	5|511	24/25	25/25	0.93	0.82	0.86
User Online Activities	0.86	0.78	0.81	99|439	24/25	25/25	0.88	0.87	0.88
User Profile	0.63	0.75	0.67	15|498	23/25	24/25	0.90	0.82	0.86
Average	0.82	0.82	0.81				0.86	0.81	0.83
Audience Type									
Children	0.99	0.99	0.99	35|33	25/25	25/25 0.99	0.99	0.99	
Average	0.99	0.99	0.99				0.99	0.99	0.99
Action First Party									
Collect on website	0.90	0.83	0.86	180|6	23/25	25/25	0.77	0.66	0.67
Collect in mobile app	0.97	0.79	0.85	23|163	25/25	25/25	0.82	0.75	0.78
Average	0.93	0.81	0.85				0.80	0.71	0.73
Action Third-Party									
Collect on first party website/app	0.90	0.98	0.94	43|8	25/25	24/25	0.84	0.80	0.82
See	0.89	0.87	0.87	14|47	25/25	25/25	0.90	0.73	0.79
Average	0.89	0.92	0.90				0.87	0.77	0.80

Table 6: An overview of network traffic collection for apps presented as case studies.

#	App Name	App Genre	Declared in Privacy Label	Declared in Privacy Policy	Trackers	Notes
Apps that do not declare tracking in their privacy label						
1	Venmo	Finance	N	Y	Kochava, Optimizely	Incomplete understanding of App Store Requirements
2	Bible	Reference	N	Y	Facebook, Google, Branch	Incomplete understanding of App Store's Definition of tracking
3	Paypal	Finance	N	Y	Adjust, Qualtrics	Incomplete understanding of third party tracking
4	Southwest Airlines	Travel	N	Y	Adobe, Qualtrics, Branch, Salesforce, Akamai	Incomplete understanding of App Store Requirements
5	My Verizon	Utilities	N	Y	Adobe, Google	Incomplete understanding of App Store's Definition of tracking
6	Open Table	Food & Drink	N	Y	Adjust, Mixpanel, Facebook	Incomplete understanding of App Store Requirements
7	Geico Mobile	Finance	N	Y	Adobe, Airship, Branch, Google	Incomplete understanding of App Store's Definition of tracking
8	Citi Mobile	Finance	N	Y	Adobe, Google, Mixpanel	Incomplete understanding of App Store's Definition of tracking
9	Crumbl	Food & Drink	N	Y	Branch, Google, Facebook	Incomplete understanding of third party tracking
10	Class Dojo	Education	N	Y	Datadog, Google, Zendesk	Incomplete understanding of App Store's Definition of tracking
11	Indeed Job Search	Business	N	Y	AppsFlyer, iSpot, Google	Incomplete understanding of App Store requirements
12	Discord	Social Networking	N	Y	Adjust, Google	Incomplete understanding of App Store's Definition of tracking
13	Sam's Club	Shopping	N	Y	Adobe, Branch, Google, PerimeterX, Moat Ads	Policy Template Reuse
14	Lime Ride	Travel	N	Y	Amazon, Branch, Facebook, Google, Unity, Super Sonic Ads	Incomplete understanding of App Store requirements
15	GroupMe	Social Networking	N	Y	OneSignal, MixPanel	Policy Template Reuse
16	Lexington Law	Finance	N	Y	Facebook, Google, Adobe, TheTradeDesk, LiveIntent, StackAdapt, Bing, TikTok, Taboola, Snapchat, Twitter	Policy Template Reuse
17	CreditRepair	Finance	N	Y	Facebook, Google, Adobe, StackAdapt, TTD, Twitter, Yahoo, LiveIntent, Taboola	Policy Template Reuse
18	Aldi	Shopping	N	Y	Adobe, Google	Incomplete understanding of App Store requirements
19	Axolochi	Games	N	Y	Google, SuperSonic, Unity	Incomplete understanding of App Store's definition of tracking
20	Hello Neighbor	Games	N	Y	Google, SuperSonic	Incomplete understanding of third party collection
21	WebMD	Medical	N	Y	Adobe, Google	Incomplete understanding of App Store's definition of tracking
22	Food Network Magazine	Food & Drink	N	Y	Facebook, Google	Incomplete understanding of App Store's definition of tracking
23	Best Buy	Shopping	N	Y	Adobe, Google, Criteo	Incomplete understanding of App Store's requirements
24	Maya Period Tracker	Health & Fitness	Partial	Y	Facebook, Google	Incomplete understanding of App Store's requirements
Apps that declare tracking in their privacy label but have an unclear privacy policy						
25	Shake Shack	Food & Drink	Y	N	Facebook, Google	Not Stated in Policy
26	Kika Keyboard	Utilities	Y	N	AppLovin, Facebook, Google	Not Stated in Policy
27	Photo Prints CVS	Photo & Video	Y	N	Facebook, Google	Not Stated in Policy
28	Everpix	Entertainment	Y	N	AppLovin, Facebook, Google, Liftoff	Not Stated in Policy
29	FloatMe	Finance	Y	N	Facebook, Google, AppsFlyer	Not Stated in Policy
30	Buffalo Wild Wings	Food & Drink	Y	N	Google	Not Clearly Stated in Policy
31	The General Auto-Insurance App	Finance	Y	N	Facebook, Google	Not Clearly Stated in Policy
32	Conservative News	News	Y	N	Amazon, AppLovin, Flurry, Google	Not Clearly Stated in Policy
33	BrainBoom	Games	Y	Y	AppLovin, Facebook, Google, Supersonic Ads, InMobi, TapJoy, IronSource, Vungle, AdColony	Presented as an image, difficult to parse
34	Stickman Boxing	Games	Y	Y	Amazon, AppLovin, Facebook, Google, IronSource, Supersonic Ads, TapJoy, Vungle, Yandex	Separate Declaration of Data Collection and Purpose.
35	McDonalds	Food & Drink	Y	Y	Adobe, Facebook, Google, Kochava	Policy segments linked on landing page
36	Episode: Choose Your Story	Games	Y	Y	Adjust, Facebook, Google	Policy linked behind a link on the landing page from App Store
37	JCPenney	Shopping	Y	Y	Adobe, Facebook, Google, UrbanAirship	Incorrect Policy Link. Different part of website
38	Dosh	Shopping	Y	Y	AppsFlyer, Google	Incorrect Policy Link. Different part of website
39	CDL Prep Test	Reference	Y	N	Google	Incorrect Policy Link. Link broken.

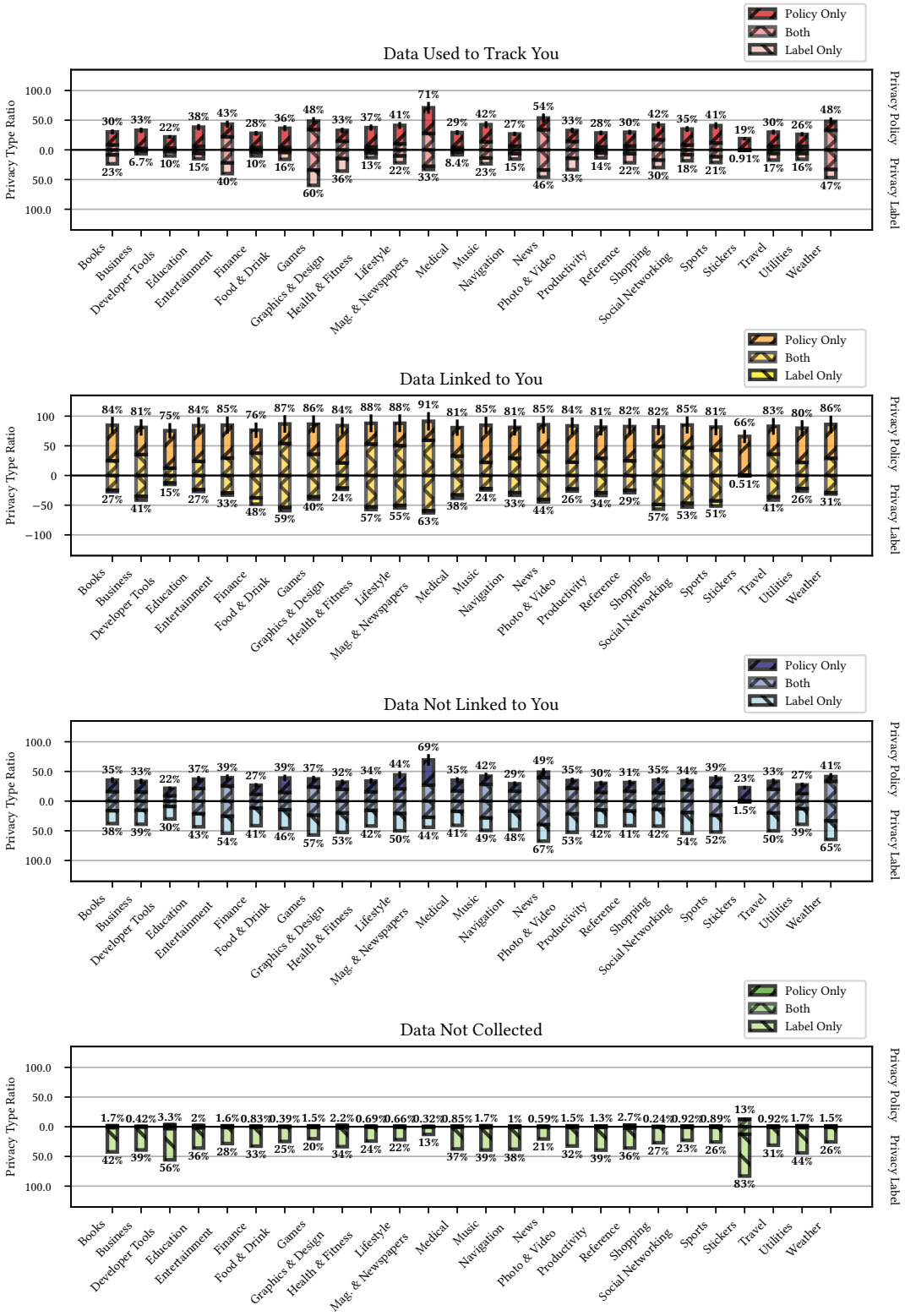

Figure 11: The ratios of app store genres for each of the four *Privacy Types*. The denominator is the number of apps with the designated app store genre that have a privacy label.

Table 7: The number of apps with three of the privacy types associated with their collection of *Data Categories*, as stated in privacy labels, against practices found in privacy policies. Please note that three of the *Privacy Types* shown here, *Data Used to Track You*, *Data Linked to You* and *Data Not Linked to You*, are not mutually exclusive. The Not Mentioned column indicates instances wherein the label or policy reports data collection, but not does not mention collecting the specific *Data Category*. (values) indicate the number of apps that did *not* also declare the corresponding privacy type found by the classifiers.

Browsing History					
Label / Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected	Not Mentioned
Data Used to Track You	**395**	421 (212)	473 (287)	23,394 (23,394)	81,830 (81,830)
Data Linked to You	624 (331)	**658**	834 (834)	61,735 (61,735)	149,552 (149,552)
Data Not Linked to You	272 (119)	324 (324)	**467**	27,573 (27,573)	83,745 (83,745)
Data Not Collected	1	0	2	**4,359**	0
Not Mentioned	301	364	642	63,597	**98,731**
Contact Info					
Label / Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected	Not Mentioned
Data Used to Track You	**1,450**	13,095 (11,710)	2,808 (2,643)	9,147 (9,147)	10,467 (10,467)
Data Linked to You	6,682 (339)	**78,411**	18,954 (13,501)	64,845 (64,845)	66,580 (66,580)
Data Not Linked to You	853 (757)	14,358 (13,638)	**3,513**	8,829 (8,829)	13,881 (13,881)
Data Not Collected	8	175	81	**4,359**	0
Not Mentioned	2,409	33,552	9,555	66,690	**64,109**
Contacts					
Label / Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected	Not Mentioned
Data Used to Track You	**0**	51 (51)	15 (15)	0	1,484 (1,484)
Data Linked to You	11	**630**	458 (458)	4740 (4740)	21,182 (21,182)
Data Not Linked to You	2 (1)	59 (59)	**32**	490 (490)	1,960 (1,960)
Data Not Collected	0	18	9	**4,359**	0
Not Mentioned	206	4,995	2,698	128,332	**237,427**
Diagnostics					
Label / Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected	Not Mentioned
Data Used to Track You	**6,725**	10,928 (6,346)	26,069 (22,465)	10,863 (10,863)	9,759 (9,759)
Data Linked to You	19,745 (5,818)	**43,487**	81,505 (73,692)	69,912 (69,912)	51,865 (51,865)
Data Not Linked to You	5,257 (2,221)	9,877 (7,692)	**31,190**	17,018 (17,018)	14,905 (14,905)
Data Not Collected	169	210	392	**4,359**	0
Not Mentioned	7,301	17,903	42,567	58,800	**32,052**
Financial Info					
Label / Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected	Not Mentioned
Data Used to Track You	**27**	3,477 (3,451)	205 (204)	2,387 (2,387)	6,276 (6,276)
Data Linked to You	194 (13)	**18,879**	2,467 (2,036)	25,044 (25,044)	65,940 (65,940)
Data Not Linked to You	23 (22)	5,055 (5,050)	**183**	2,570 (8,829)	7,067 (7,067)
Data Not Collected	1	27	6	**4,359**	0
Not Mentioned	335	9,313	2,321	107,565	**167,701**

Table 7: The number of apps with three of the privacy types associated with their collection of *Data Categories*, as stated in privacy labels, against practices found in privacy policies. Please note that three of the *Privacy Types* shown here, *Data Used to Track You*, *Data Linked to You* and *Data Not Linked to You*, are not mutually exclusive. The Not Mentioned column indicates instances wherein the label or policy reports data collection, but not does not mention collecting the specific *Data Category*. (values) indicate the number of apps that did *not* also declare the corresponding privacy type found by the classifiers.

Health					
Label / Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected	Not Mentioned
Data Used to Track You	**0**	52 (52)	18 (18)	30 30	133 (133)
Data Linked to You	11 (1)	**898**	299 (286	867 (867)	1,506 (1,506)
Data Not Linked to You	1 (1)	61 (61)	**25**	55 (55)	72 (72)
Data Not Collected	0	12	6	**4,359**	0
Not Mentioned	92	6,326	1,991	132,420	**256,974**
Identifiers					
Label / Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected	Not Mentioned
Data Used to Track You	**28,152**	41,171 (22,171)	22,056 (10,954)	25,828 25,828	28,866 (28,866)
Data Linked to You	55,763 (18,694)	**86,392**	47,009 (40,280	81,688 (81,688)	65,600 (65,600)
Data Not Linked to You	21,097 (12,635)	30,657 (28,207)	**19,027**	25,844 (25,844)	30,043 (30,043)
Data Not Collected	276	302	249	**4,359**	0
Not Mentioned	13,361	29,169	16,815	47,889	**22,411**
Location					
Label / Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected	Not Mentioned
Data Used to Track You	**4,741**	6,218 (2,961)	4,274 (2,764)	5,711 5,711	13,664 (13,644)
Data Linked to You	10,861 (3,370)	**27,248**	20,812 (20,420	30,410 (30,410)	67,749 (67,749)
Data Not Linked to You	3,860 (2,696)	5,572 (5,492)	**4,650**	6,962 (6,962)	14,344 (14,344)
Data Not Collected	178	204	137	**4,359**	0
Not Mentioned	15,047	24,228	24,818	100,638	**97,879**
Purchases					
Label / Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected	Not Mentioned
Data Used to Track You	**136**	2,664 (2,552)	275 (251)	1,194 1,194	3,255 (3,255)
Data Linked to You	1,183 (255)	**12,790**	1,339 (1,339	8,726 (8,726)	22,120 (22,120)
Data Not Linked to You	194 (142)	4,878 (4,878)	**172**	1,853 (1,853)	4,151 (4,151)
Data Not Collected	13	35	34	**4,359**	0
Not Mentioned	8,629	24,081	7,601	123,794	**197,783**
Search History					
Label / Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected	Not Mentioned
Data Used to Track You	**47**	233 (211)	123 (98)	1,118 1,118	3,708 (3,708)
Data Linked to You	154 (41)	**920**	1,303 (1,303	7,889 (7,889	30,287 (30,287)
Data Not Linked to You	61 (35)	110 (110)	**873**	1,926 (1,926)	7,679 (7,679)
Data Not Collected	1	6	8	**4,359**	0
Not Mentioned	992	4,253	4,600	124,521	**222,281**

Table 7: The number of apps with three of the privacy types associated with their collection of *Data Categories*, as stated in privacy labels, against practices found in privacy policies. Please note that three of the *Privacy Types* shown here, *Data Used to Track You*, *Data Linked to You* and *Data Not Linked to You*, are not mutually exclusive. The Not Mentioned column indicates instances wherein the label or policy reports data collection, but not does not mention collecting the specific *Data Category*. (values) indicate the number of apps that did *not* also declare the corresponding privacy type found by the classifiers.

Sensitive Info					
Label / Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected	Not Mentioned
Data Used to Track You	**32**	335 (303)	75 (75)	4,752 4,752	17,912 (17,912)
Data Linked to You	78 (5)	**2,144**	480 (480	21,567 (21,567	74,790 (74,790)
Data Not Linked to You	25 (25)	358 (358)	**109**	6,318 (6,318)	26,134 (26,134)
Data Not Collected	0	8	5	**4,359**	0
Not Mentioned	86	3,123	530	108,606	**177,261**
Usage Data					
Label / Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected	Not Mentioned
Data Used to Track You	**23,563**	28,307 (14,402)	38,829 (26,818)	23,394 23,394	18,611 (18,611)
Data Linked to You	48,575 (20,829)	**63,514**	80,399 (74,937	87,767 (87,767	59,425 (59,425)
Data Not Linked to You	21,458 (9,950)	31,074 (28,120)	**44,701**	32,792 (32,792)	25,493 (25,493)
Data Not Collected	307	254	410	**4,359**	0
Not Mentioned	9,930	14,700	19,874	39,103	**24,349**
User Content					
Label / Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected	Not Mentioned
Data Used to Track You	**226**	1,632 (1,456)	1,089 (1,031)	2,000 2,000	4,915 (4,915)
Data Linked to You	1,266 (364)	**22,550**	8,669 (6,556	20,713 (20,713	41,193 (41,193)
Data Not Linked to You	186 (116)	1,746 (1,634)	**980**	2,824 (2,824)	7,265 (7,265)
Data Not Collected	7	75	72	**4,359**	0
Not Mentioned	2,177	34,233	18,898	111,707	**144,204**