

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

Mir Masood Ali <sup>†</sup>, David G. Balash <sup>⊙</sup>, Chris Kanich <sup>†</sup> and Adam J. Aviv <sup>◇</sup>

<sup>†</sup> *University of Illinois at Chicago*, <sup>⊙</sup> *University of Richmond*, <sup>◇</sup> *The George Washington University*

Last Update: June 30, 2023

## 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 do also require developers to provide a privacy policy, which offers an important comparison point. In this paper, we applied the NLP framework of Polisis to extract features of the privacy policy for 515,920 apps on the iOS App Store comparing the output to the privacy labels. We identify discrepancies between the policies and the labels, particularly as it relates to data collected that is linked to users. We find that  $287 \pm 196\text{K}$  apps' privacy policies may indicate data collection that is linked to users than what is reported in the privacy labels. More alarming, a large number of ( $97 \pm 30\%$ ) of the apps that have *Data Not Collected* privacy label have a privacy policy that indicates 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 there is still significant work to be done to help developers more accurately labeling their apps. Incorporating a Polisis-like system as a first-order check can help improve the current state and better inform developers when there are possible misapplication of privacy labels.

## 1 Introduction

Privacy policies are ubiquitous and required in many settings [46, 48, 47, 71], 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 [65, 42] as a tool to inform users clearly about the behaviors and data management practices. *Privacy nutrition labels*, or *privacy labels*, (similar

to a food nutrition label [60, 35]) offer an alternative to both simplify and standardize communication of privacy behavior similar to food nutrition labels. In December of 2020, Apple began requiring privacy labels [27] 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 standardizes the presentation of privacy behavior that was previously described in the natural language text of the privacy policy.

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 515,920 apps' privacy policies and privacy labels using a validated implementation of Polisis [56], an NLP privacy policy tool that reports features of the privacy policies' data management practices. We mapped Polisis' features to Apple's privacy labels to identify discrepancies between the reported behavior of apps based on their labels as compared to their privacy policies.

We find that there are large differences between privacy labels and privacy policies. Most prominently, according to Polisis' analysis of the privacy policies, nearly  $287 \pm 196\text{K}$  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, ( $97 \pm 30\%$ ) apps that report no data collection in their privacy label have statements in their privacy policy to the contrary. In many cases mislabeling vary from the privacy policy in terms of the kinds of data that is 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 6% of paid apps report collecting data that is linked to users, but the policies suggest

\*Primary contacts: Mir Masood Ali - mali92@uic.edu

that  $83 \pm 38\%$  apps perform such collection. We further analyzed privacy-relevant data practices that aren't covered by privacy labels. We found that while most apps (81%) had a self-assigned content rating of 4+, on the App Store, to indicate age appropriateness and enforce parental controls. Of these apps, only  $46 \pm 0.01\%$  of such apps had a policy in place to handle data collected from children. Our case study further reveals that their policy might just be to claim no responsibility of collecting and handling data collected from users under 13 years of age. We also employ a similarity metric and identify that 58% of evaluated apps potentially make use of templates, providing insight into a possible source of discrepancies. We further analyzed the network traffic from 30 apps, showing that their data collection practices diverge from those declared in privacy labels and privacy policies.

Our analysis indicates that privacy labels are likely misapplied in great numbers, even considering that Polisis is an imperfect tool for analyzing privacy policies. More guidance for developers would go a long way to improving the accuracy of privacy labels, but there are also more concerning misapplications that could and should be addressed more broadly, such as 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 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 (e.g., using Polisis) 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.

## 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 Kelly et al. [62]. The structure of a label is hierarchical (see Figure 1 for details) and describes data collection practices under four levels: **(1) Privacy Type:** Describes the way in which the collected data is handled. This includes collected for tracking users (with third parties), collected and linked to users' identity, and collected but aggregated/anonymized. An app's privacy

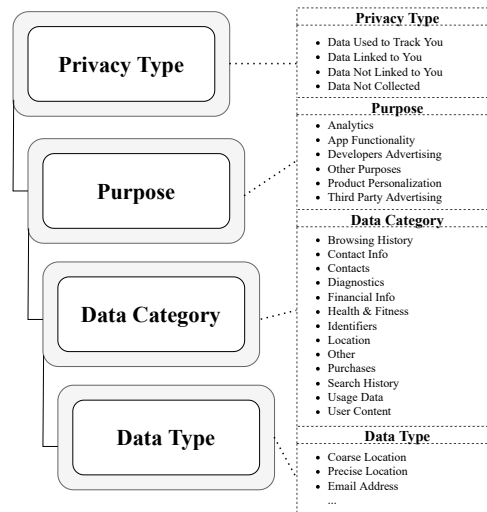


Figure 1: Anatomy of a Privacy Label.

label may contain a combination of one, two, or all three of these 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 type of data collected. **(4) Data Type:** Granular information to describe the information collected under the Data Category.

**Privacy nutrition labels.** Privacy nutrition labels have been studied from a variety of perspectives [80, 41, 30, 60, 61, 62, 43, 44, 78], but Apple's privacy label is the first wide-scale deployment [27]. Balash et al. [29] performed a 36-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. [89] conducted an in-depth interview study to determine the usability of iOS privacy labels from a user perspective. Most users found the privacy labels useful despite misunderstandings that included unfamiliar terms and a confusing structure. Garga et al. [?] discovered that privacy label disclosures of sensitive information does reduce app demand, and thus the accuracy of the labels is important to help users make informed choices.

Gardner et al. [49] developed a tool to assist developers

by prompting them while coding of possible functionality that would require a privacy label. Li et al. [64] studied the usability of Apple’s privacy nutrition label creation process for developers by conducting semi-structured interviews and observing developers as they applied privacy labels. 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 among others. We observe the same when comparing to the privacy policies, and Li et al.’s findings regarding “knowledge blindspots” and misinterpreted Apple’s definitions likely leads to many of the missapplications we identified.

**Privacy behavior on mobile apps.** Numerous studies have measured the privacy behaviors of mobile applications [21, 22, 33, 34, 37, 79, 88, 90, 91, 69]. One of the first approaches to automatically identify problems in privacy policies was PPChecker [88] which combined an NLP analysis of privacy policy text with bytecode analysis. Andow et al. [21] developed PolicyLint to identify contradictions within an individual policy. Andow et al. [22] also created PoliCheck which takes into account third-party versus first-party entity access to personal data for an entity sensitive consistency check. Bui et al. [34] extended PoliCheck to develop PurPliance that checks if data, entity, and purpose are equivalent to those extracted from the data flows. In this paper, we choose Polisis [56] as the policy analysis tool as it produces output that is similar to that of the privacy labels.

Zimmeck et al. [90] evaluated 1,035,853 Android apps using 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. [63] analyzed a small number (1,759) of iOS apps using a combination of network traffic monitoring, and they found that 80% of the apps that claimed to not collect any data in the privacy labels actually contained at least one tracker library. We find that this discrepancy probably exists at scale. Xiao et al. [87] analyzed 5,102 apps by checking the privacy labels against actual data flows, discovering 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 this finding, where mentioning of unique identifiers in the pri-

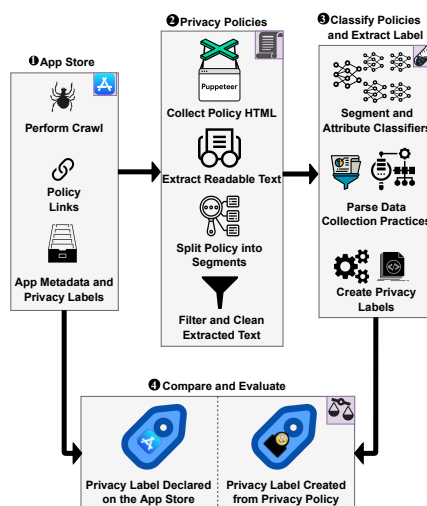


Figure 2: An overview of the measurement workflow.

vacy policy are not reflected in the data linking labels.

### 3 Measurement Workflow

In Figure 2, we present the primary measurement workflow, which we describe in detail below. Note that our analysis consider 515,920 apps with an accessible privacy policy and privacy labels, which constitutes roughly half of the 1M apps with a privacy label on the App Store at the time of our measurement in July 2022. During all scans, we followed best practices of limiting the number of requests and respecting 403 Errors by using exponential back-offs.

**1 Crawling the App Store.** We began by first 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 meta-data, such as the app name, version, size, type, user rating, genre, content rating, release date, seller name, and price. Importantly, included in the meta-data is a link the privacy policy. We also parsed the extended privacy label details, such as the purposes and data types, we performed an additional GET request to the Apple Catalog API [23]. In July 2022, there were 1,569,057 apps on the App Store. Of them, 1,041,469 apps had a privacy label and we iden-

tified 1,038,930 apps with links to 716,823 unique policies (note some apps link to the same policy).

**② Collecting Privacy Policies.** We extracted the HTML for each policy using a Puppeteer script in *headful mode* [13] attached to a virtual display [72]. We then leverage the `readability` library [68, 82], a standalone version of the Firefox browser *reader mode*. The library employs a complex set of heuristics to extract relevant text from web pages [10], leaving us with de-cluttered HTML that can be divided into segments based on the `<p>` tags. 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 or less – and merged them into the preceding paragraph, thereby treating the entire list as a single segment. After cleaning, each segment was individually processed by the classifiers and mapped back to the original policy from which it was extracted. After excluding policies that had name resolution and HTTP response errors, Firefox’s `readability` library successfully extracted relevant text from the policies of 763,400 apps ( $n = 448,429$  policies) which we classified in the next stage.

**③ Classifying Policies and Extracting Labels.** For analyzing the privacy policies, we used Polisis [56], an NLP that classifies behavior based on the privacy policy. Unfortunately, the prior published Polisis implementation is proprietary, but the implementation details and the datasets used to train the algorithm are public, including a recent open-sourced version [86]. We adapted and performed a full re-implementation of Polisis to the same standards as in prior work to perform large-scale analysis. An overview of Polisis structure is provided in Figure 3, and Polisis accuracy/precision evaluation results can be found in Table 3 in the Appendix. A more detailed description of Polisis’ functionality and implementation can be found in Appendix A.

Each segment is first passed through Polisis’ Segment Classifier to extract the high-level data practice. If the Segment Classifier indicated that the segment addressed some data collection practice, i.e., *First Party Collection/Use* or *Third Party Collection/Sharing*, we passed these segments through four Attribute Classifiers – *Does/Does Not*, *Identifiability*, *Purpose*, and *Personal In-*

*formation Type* – to extract annotations relevant to privacy labels. Prior to creating a Polisis based label entry from any segment, we first verified the output of the *Does/Does Not* classifier. Any segment claiming to *not* engage in a data collection practice was excluded from consideration, except for the *Data Not Collected* label. We detected data collection in 552,495 apps ( $n = 351,453$  policies), which we considered for further filtering. It is important to reiterate that the *Does Not* attribute was used as a filter while extracting classifier results. The performance of this attribute classifier is included in Table 3 in Appendix A.

Apple requires that developers link to their app’s privacy policies on the App Store [2, 25]. However, developers may instead link to policies that apply to a website or a range of other services. We further passed relevant segments gathered from the previous stage through additional attribute classifiers. For segments addressing *First Party Collection*, we filtered out any segments for which the *Action First Party* attribute classifier identified collection on website but did not identify collection being performed in a mobile app. For segments addressing *Third Party Collection*, we will filtered out any segments for which the *Action Third Party* attribute classifier indicated that the data is seen but is neither collected nor tracked on the first party website/app. While imperfect, this filtering mechanism helps exclude policy segments that explicitly address data collection specific to websites. We report on the metrics of these attribute classifiers in Table 3 in Appendix A.

We considered the union of Polisis labeling across the segments of a given policy as the derived privacy label. We successfully detected at least one segment addressing data collection in the policies of 515,920 apps ( $n = 338,917$  policies), which we used for analysis in the next stage.

**④ Compare and Evaluate.** The taxonomy of Polisis labeling does not always have a one-to-one mapping with Apple’s privacy labels, and so we developed a grounded strategy, based in qualitative coding, to convert outputs from Polisis into equivalent privacy labels. Three researchers independently coded the conversions and then discussed to reach agreement on the mappings between Polisis and privacy labels. The coders were asked to complete three amatching tasks:

- First, the coders determined which of the data

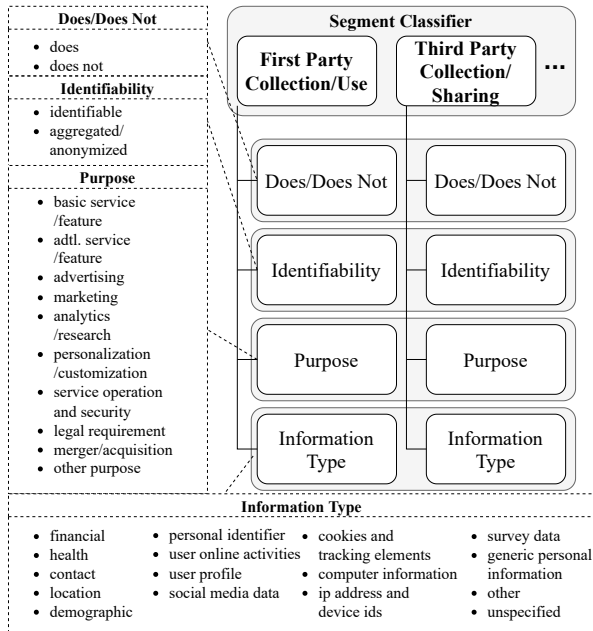


Figure 3: The hierarchical structure of the Polisis classifiers.

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,” or “Does Not”, match to an appropriate Apple privacy label type, such as *Data Linked to You* or *Data Not Collected*. For example, when Polisis identifies a segment with a data practice of “First Party Collection/Use” and the data is “Identifiable,” then that would associate with an Apple privacy label type of *Data Linked to You*.

- Next, the coders matched the output of the Polisis Purpose Attribute Classifier against Apple’s privacy label purposes. For example, Polisis may provide a classification of “Basic Services/Features” which 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 la-

Table 1: Deriving privacy label entries directly from segment annotations created using the Polisis framework. [Numbers] indicate uncertainty calculated from the F1 score for the corresponding classifier result.

Apple Privacy Label	Polisis	
Privacy Type	High-level Data Practice	Identifiability
Data Linked to You [±0.38]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	Identifiable [±0.25]
Data Not Linked to You [±0.29]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	Aggregated/anonymized [±0.15]
Privacy Type	High-level Data Practice	Does/Does Not
Data Not Collected [±0.30]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	Does Not [±0.16]
Purpose	High-level Data Practice	Purpose
App Functionality [±0.33]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	Basic Service/feature [±0.20] Additional Service/feature [±0.20] Service operation & security [±0.17]
Analytics [±0.29]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	Analytics/Research [±0.15]
Developers Advertising [±0.31]	First Party Collection/Use [±0.20]	Advertising [±0.14]
Other Purposes [±0.29]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	Merger/Acquisition [±0.07] Legal requirement [±0.13] Unspecified [±0.24]
Third Party Advertising [±0.26]	Third Party Collection/Sharing [±0.14]	Advertising [0.14]
Product Personalization [±0.32]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	Personalization/Customization [±0.18]
Data Category	High-level Data Practice	Personal Information Type
Contact Info [±0.24]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	Contact [±0.09]
Location [±0.27]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	Location [±0.12]
Financial Info [±0.24]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	Financial [±0.08]
Identifiers [±0.24]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	Cookies & Tracking Elements [±0.09] IP address & Device IDs [±0.08]
Usage Data [±0.27]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	User Online Activities [±0.12]
User Content [±0.29]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	User Profile [±0.14] Social Media Data [±0.14]
Health & Fitness [±0.39]	First Party Collection/Use [±0.20] Third Party Collection/Sharing [±0.14]	Health [±0.26]
Browsing History [±0.24]	Third Party Collection/Sharing [±0.14]	User Online Activities [±0.12]

bel data category of *Contact info*.

The combination of these three matching tasks provides a single privacy label 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*). The full list of the direct conversions identified in this manner can be found in [Table 1](#). The coding process also revealed additional, inferred privacy labels from Polisis classification that included a combination of classifications, which are relevant for *Data Used to Track You* and remaining *Data Categories*. The inferred privacy labels are shown in [Table 2](#).

Table 2: Inferring privacy label entries from segment annotations created using the Polisis framework. [Numbers] indicate uncertainty calculated from the F1 score for the corresponding classifier result.

Apple Privacy Label	Polisis	
Privacy Type	High-level Data Practice	Purpose
Data Used to Track You [ $\pm 0.26$ ]	Third Party Collection/Sharing [ $\pm 0.14$ ]	Advertising [ $\pm 0.14$ ]
Data Category	High-level Data Practice	Personal Information Type
Diagnostics [ $\pm 0.24$ ]	First Party Collection/Use [ $\pm 0.20$ ] Third Party Collection/Sharing [ $\pm 0.14$ ]	Computer Information [ $\pm 0.08$ ] IP address & Device IDs [ $\pm 0.08$ ]
Contacts [ $\pm 0.29$ ]	First Party Collection/Use [ $\pm 0.20$ ] Third Party Collection/Sharing [ $0.14$ ]	Social Media Data [ $\pm 0.14$ ] 'contact', 'friend' 'address book', 'phone book'
Purchases [ $\pm 0.25$ ]	First Party Collection/Use [ $\pm 0.20$ ] Third Party Collection/Sharing [ $\pm 0.14$ ]	Financial [ $\pm 0.08$ ] User Online Activities [ $\pm 0.12$ ]
Search History [ $\pm 0.30$ ]	First Party Collection/Use [ $\pm 0.20$ ]	User Online Activities [ $\pm 0.12$ ] 'search'
Sensitive Info [ $\pm 0.27$ ]	First Party Collection/Use [ $\pm 0.20$ ] Third Party Collection/Sharing [ $\pm 0.14$ ]	Demographic [ $\pm 0.12$ ] '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'

## 4 Limitations

Before proceeding, it is important to note the limitations of our approach in comparing the privacy labels with the privacy policies. Foremost that we do not have not have verified ground truth, and neither the labels nor the policies are directly verified in any meaningful way that can provide that ground truth. We can report on observed discrepancies between the policies and the labels, but validating which is actually more correct is beyond the scope of this paper. However, we argue that discrepancies at the scale identified (as reported in the next section) more likely than not indicates the prominence of misapplication of privacy labels.

We also use our own implementation of Polisis, rather than original code base. This was necessitated since the original code is proprietary, but we worked with the authors to develop a validated version using the same training and testing data sets. We are confident that our implementation is as accurate and precise as prior work.

Polisis is not without its own limitations. The outputs of the classifiers introduce a level of uncertainty that propagates further when combined. We highlight the resulting uncertainties in Table 1 and Table 2, add error bars to pre-

sented figures, and append all reported numbers to ensure that the results are read in-context. As a result of these inaccuracies, we can only report on the *possible* presence of statements addressing data collection practices in privacy policies, and differences when compared with privacy labels. However, many of the discrepancies are *much* larger than the associated uncertainties, and the results of the classifiers are comparable to prior work. The goal of this work is not to provide a ground truth from the policies, but to note that that automated analysis of the policies seem to greatly disagree with the current labeling in the App Store.

Finally, there are some important caveats regarding the mapping of Polisis classifications to privacy labels. First, Polisis analyzes privacy policies on a per-paragraph/per-segment basis, so Polisis will not be able to detect explanations of app behaviors that span multiple segments. Many privacy policies are designed to be self-contained discussion per-segment, but we cannot be sure that this is the case throughout our dataset. Second, many privacy policies contain links to subordinate privacy policies of third parties. We did not analyze the transitive closure of all privacy policies as part of this work. It is possible that additional clarifying details are found in subordinate policies, but more likely, additional data collection practices would be found there. One could consider our analysis as a lower-bound of behavior, particularly related to third-parties. It is also Apple’s policy for privacy labels to include all collection and tracking mechanisms, even performed done by third parties.

## 5 Results

In this section, we directly compare developers’ reported privacy labels to the output of Polisis following the hierarchical structure of the privacy labels (see Figure 1). We first compare based on the privacy types of the labels, such as if data is linked, not linked, used to tracked, or not collected. Next, we compare the purposes reported in the privacy labels for data collection to what is presented in the policies. Next, we consider the data categories of the label, such as which data is subject of collection. We also offer a metadata analysis comparing discrepancies in the labels and policies based on free vs. paid apps, as well as self-declared content ratings of apps. Finally, analyze the

privacy labels for a subset of apps whose privacy policies are similar to templates.

The numbers reported in this section indicate uncertainty reported by considering the F1 scores of the underlying classifier results used to arrive at that value. For example, consider *Data Linked to You* privacy type, which relies on the outputs of the Segment Classifier, specifically *First Party Collection/Use* and *Third Party Collection/Sharing*, which have corresponding F1 scores of 0.80 and 0.86 respectively. Combining these scores gives us an average F1 score of 0.83. Additionally, we also takes into account the *Identifiable* result obtained from the Identifiability attribute classifier in Table 3, which has an F1 score of 0.75. By multiplying the F1 scores of the Segment Classifier and *Identifiable* result, we obtain a combined F1 score of 0.62. Consequently, the uncertainty of the *Data Linked to You* privacy type identified from privacy policies is expressed as  $(1 - 0.62) = \pm 0.38$ . We append all reported numbers with similarly calculated  $\pm$  values to ensure that results are read in context.

**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 that have such a privacy type and if that type is also found in the policies according to Polisis. Figure 4 and Table 5 (in Appendix C) 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 that is both, linked to the user and aggregated/anonymized, and hence also not linked to the user, and additionally, they may collect data to track users.

We found that Polisis correctly captured data linking in the policies of 92% ( $n = 202,082 \pm 83,673$ ) of the apps whose privacy labels stated that they collected data in an identifiable manner, i.e., had a *Data Linked to You* privacy type ( $n = 220,191$ ). We further identified that 56% ( $n = 287,824 \pm 196,049$ ) of the more than 515K apps that we analyzed have segments in their privacy policy suggesting that they collect data linked to the user but failed to report

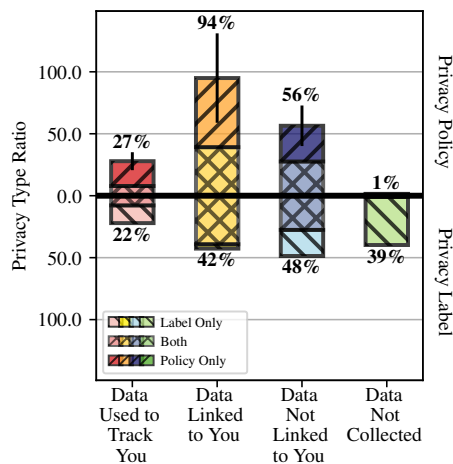


Figure 4: An overview of the privacy types associated with data collection on the App Store, from privacy labels and privacy policies. The denominator is the total apps that we analyzed, i.e., 515,920 apps. Please note that the privacy types, except for *Data Not Collected*, are *not* mutually exclusive.

such data collection in their privacy label.

We identified that 57% ( $n = 142,330 \pm 72,782$ ) 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 stated so in their policies. Of the remaining 43% ( $n = 108,642$ ) apps that had the *Data Not Linked to You* privacy type in their label but did not have an identified segment by Polisis, 70% ( $n = 75,730 \pm 41,284$ ) of those instead included segments in their privacy policy that indicated that they actually collect data linked to users.

This difference may be the result of apps *not* stating their aggregation and anonymization practices in the same paragraph of the policy that addresses data collection. Similarly, that the numbers reported for *Data Linked to You* privacy type may have higher uncertainty owing to the *Identifiable* classifier attribute. We elaborate on this potential difference in section 4. But even as an overestimation, there is likely a large gap between the privacy labels and the true privacy behavior as Polisis itself has a F1 score of 0.75 for identifying such behavior.

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., it is only added by apps that claim that they do *not* collect *any* data from their users. While 40% ( $n = 205,274$ ) of the apps that we analyzed indicated in their privacy label that they did *not* collect *any* data, only 0.03% ( $n = 5,310 \pm 61,582$ ) of these apps made similar statements in their policies. More surprisingly, 84% ( $n = 173,441 \pm 78,004$ ) of these apps stated in their policies that they collected data that is linked to the user. Note that Apple requires developers to report this label only when they do not collect *any* data. Our conversion method only creates this label after performing two passes of a policy. In the first pass, we ensure that we do not find any segment that the app collects data, and in the second pass we look for at least one segment that states that the app *does not* collect data. Note that this approach is still a per-segment analysis, and the associated privacy type is created based on the classifier results associated with a single segment. However, unlike other values, we only create this privacy type after ensuring that no other privacy label entry can be applied to the application, thereby ensuring consistency with the composition of Apple’s privacy labels. We may, as a result, find fewer policies mapping to this label, but the large number of policies that indicate other forms of data collection are a cause for concern.

Of the 22% ( $n = 114,095$ ) of apps that we analyzed which stated in their privacy label that they collected data that’s used to track users, our conversion method correctly identified tracking in 35% ( $n = 40,302 \pm 29,664$ ). We also identified an additional 20% ( $n = 103,428 \pm 134,139$ ) that mention in their privacy policies that they collect data for *Advertising*, and share this data with third parties. Recall that this privacy type is inferred from the outputs of the Polisis framework and reports on part of the tracking that apps engage in. Our identification of apps that fail to report data being collected for tracking points to the possibility of a larger number of apps under-reporting their tracking practices.

To summarize our findings, it is very likely that apps may be under-reporting the collection of data that is linked to users and data that is used to track users in their privacy labels, even considering the limitations and drawbacks of the Polisis’ NLP analysis. The majority of apps whose privacy labels indicate that they do *not* collect *any* data, state otherwise in their privacy policies.

**Purposes.** We further looked at how apps claimed to be using the data that 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 data that is both linked and not linked (anonymized). Additionally, data may be collected 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 that apps most commonly stated that the data they collect is used for *App Functionality* (94%;  $n = 484,678$ ) and *Analytics* (82%;  $n = 424,628$ ) according to either, their privacy labels and their privacy policies. Fewer apps stated that they used data for *Product Personalization* (45%;  $n = 231,152$ ), *Developer Advertising* (39%;  $n = 202,049$ ), or *Third Party Advertising* (38%;  $n = 195,071$ ). Apps are more likely to associate their data collection as being beneficial to the end user and the app itself, than they are to associate with their advertising and marketing practices.

We also find greater agreement between privacy labels and privacy policies for apps that stated they collect data for *App Functionality* and *Analytics*. Of the 103,548 apps indicated in their privacy label that they collect data linked to users for *App Functionality*, 67% ( $n = 69,795 \pm 34,170$ ) also included a corresponding statement in their privacy policy. Similarly, of the 114,974 apps who stated in their privacy label that they collect data linked to users for *Analytics*, 61% ( $n = 69,765 \pm 33,342$ ) also included a corresponding statement in their privacy policy. Recall that the reported uncertainty is also exacerbated by considering the high-level data practice (see Table 1). Regardless, even considering uncertainty, we found the highest overlap for these two purposes.

Finally, we find that of the 402,948 ( $78 \pm 29\%$ ) apps that stated in their privacy policies that they collected data for a purpose that does not fit into any of the options that Apple provides in their privacy label, only 43,929 (11%) of these apps also addressed this in their privacy label. Given that privacy policies are free form, the information that developers provide in these policies can be more elaborate. It is likely that Apple’s predefined set of purposes may restrict developers from being more forthcoming about additional uses of the data they collect.



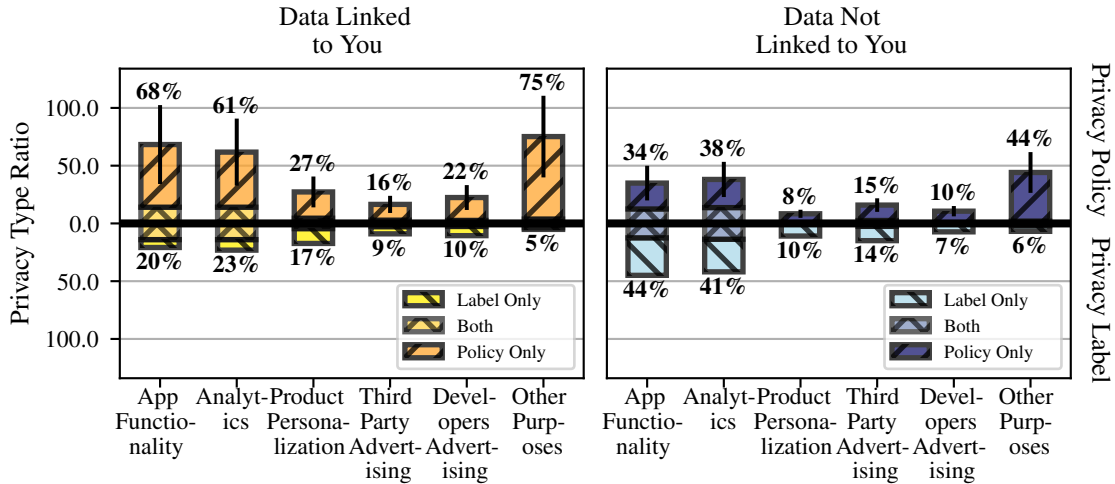


Figure 5: The ratios of the six purposes for the *Data Linked to You* and *Data Not Linked to You* privacy types. The denominator is the number of apps with the designated privacy type either in their privacy label or their privacy policy, i.e., 497,124 apps with a *Data Linked to You* label and 386,519 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.

To summarize our findings, app developers are more likely to declare data collection that’s used for *App Functionality* and *Analytics* in either, privacy labels or privacy policies. Apps are less likely to declare data collection in their privacy labels when the data is used for purposes outside of the predefined set of values provided by Apple, i.e., *Other Purposes*.

**Data Categories.** We additionally analyze the collection of data categories that apps state in their privacy labels and in their privacy policies. A visual of these results can be found in [Figure 8](#) in [Appendix C](#).

We find that apps are more more likely to declare in either their privacy policies or their privacy labels that they collect *Contact Info* ( $n = 352, 891$ ; 71%) and *Identifiers* ( $n = 320, 801$ ; 65%) that is linked to users. On the other hand, we only identified 23,895 apps (5%) stating that they collect *Health and Fitness* data that is linked to users.

Apps that collect data to track users are more likely to use *Browsing History* (43%;  $n = 90, 315$ ), *Identifiers* (61%;  $n = 127, 466$ ), and *Usage Data* (67%;  $n = 141, 422$ ). These 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*) [19, 45].

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), *Financial Info* (i.e., payment methods, credit score, income, etc.), and *Sensitive Info* (such as racial/ethnic data, sexual orientation, etc.) that is linked to users are less likely to declare this collection in their privacy labels. Surprisingly, of the  $167,659 \pm 119, 309$  apps that stated in their privacy policy that they collect *Browsing History* that is linked to the user, only 719 (0.4%) of these apps declared this practice in their privacy labels. Of the  $132,277 \pm 119, 309$  apps that stated in their privacy policy that they collect *Financial Info* that is linked to the user, only 20,144 (15%) apps also stated this collection in their privacy labels. Finally, while we found that  $170,924 \pm 134, 223$  apps indicated in their privacy policy that they collect some form of *Sensitive Info*, only 2% ( $n = 3, 733$ ) apps also declared this collection in their privacy labels. Note that the reported uncertainties here are high, especially after considering

the high-level data practice (see [Table 1](#)). Regardless, even after considering uncertainty, the discrepancies are at least an order of magnitude off, indicating that a few tens of thousands of apps are under-reporting collected data categories.

To summarize our findings, apps most commonly state in either, their privacy labels or their privacy policies, that they collect *Identifiers* and *Contact Info* that’s linked to users. Apps that state in their privacy policies that they collect *Browsing History*, *Financial Info*, or *Sensitive Info* that’s linked to users are less likely to also 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.** There are four pricing models on the App Store: free apps, free apps with in-app purchases, paid apps, and paid apps with in-app purchases. Interestingly, when only observing privacy labels (the bottom half of [Figure 9](#) in [Appendix C](#)), it would appear that paid apps have better privacy behaviors than their free counterparts. However, the altruism of paid apps compared to free apps disappear when considering the privacy policies (the top half of [Figure 9](#) in [Appendix C](#)). The privacy policy analysis better aligns with observations of Han et al. [[54](#), [55](#)] who compared free and paid apps in the Android Play Store based on 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 this privacy type has the largest discrepancies of potentially under-reporting data collection practices in their privacy labels, as compared to the privacy policies. Paid apps also appear to under-report their privacy types. The privacy policies suggest that 83% ( $n = 16,568 \pm 7,592$ ) of paid apps collect data linked to users, only 6% ( $n = 998$ ) of these apps have a privacy label of this type, and when collecting data not linked to users, 48% ( $n = 9,542 \pm 5,794$ ) have a segment/paragraph in their privacy policy discussing such actions while only 21% ( $n = 1,984$ ) have a privacy label. Note the range of uncertainty resulting from classifier results. However, considering that values inferred from policies align with prior findings, our observations indicate discrepancies between declaration in labels and policies.

**Content Rating.** Developers provide a *Content Rating* as part of the app metadata to indicate the age appropriateness of their apps. *Content Rating* indicates the objectionable nature of material within the app, and does not directly indicate whether they are targeted towards children. These ratings are reviewed by Apple [[24](#)], and used to enforce parental control features that restrict children from accessing the app [[26](#)]. At the outset, we highlight that unless an application specifically targets children, they are not obligated to include a corresponding segment in their privacy policy. The discussion that follows sheds light on a potential area of concern given the content rating that apps adopt and how parental controls work. We find that most apps that we analyzed had a 4+ content rating on the App Store (81%;  $n = 419,762$ ), while fewer apps had 9+ (3%;  $n = 16,687$ ), 12+ (9%;  $n = 46,737$ ), or 17+ (13%;  $n = 69,309$ ) content rating values. Since privacy labels do not indicate the app’s data practices specific to children, the only other option that users have for learning this information is to review the privacy policy. Given parental control settings, an app with a 4+, 9+, or 12+ rating could be used by minors, although they may not be the intended audience for the app. But when an app specifically targets children, it is subject to additional regulation that may require parental consent. Polisis can identify policy segments that address *International/Specific Audiences* and further identify if the segment addresses *Children*, and then compare this output to the content rating. Only 46% ( $n = 191,120 \pm 4,197$ ) apps with a 4+ content rating also included a privacy policy segment that addresses data practices specific to children. We were more likely to find similar policy segments for apps with different content ratings that can also be accessed by children, 9+ (64%;  $n = 10,700 \pm 167$ ) and 12+ (50%;  $n = 23,529 \pm 467$ ).

We further looked at app content ratings for different privacy label types associated with data collection. These findings are presented in [Figure 6](#). Considering apps with a 4+ content rating, we find that across privacy types, roughly half of these apps had a policy in place that specifically addressed children. While 18% ( $n = 75,346$ ), 37% ( $n = 154,972$ ), and 44% ( $n = 184,722$ ) of the apps with a 4+ content rating declare in their privacy label that they collect data that is used to track users, linked to users, and not linked to users respectively, only 53% ( $n = 40,081 \pm 19,590$ ), 48%

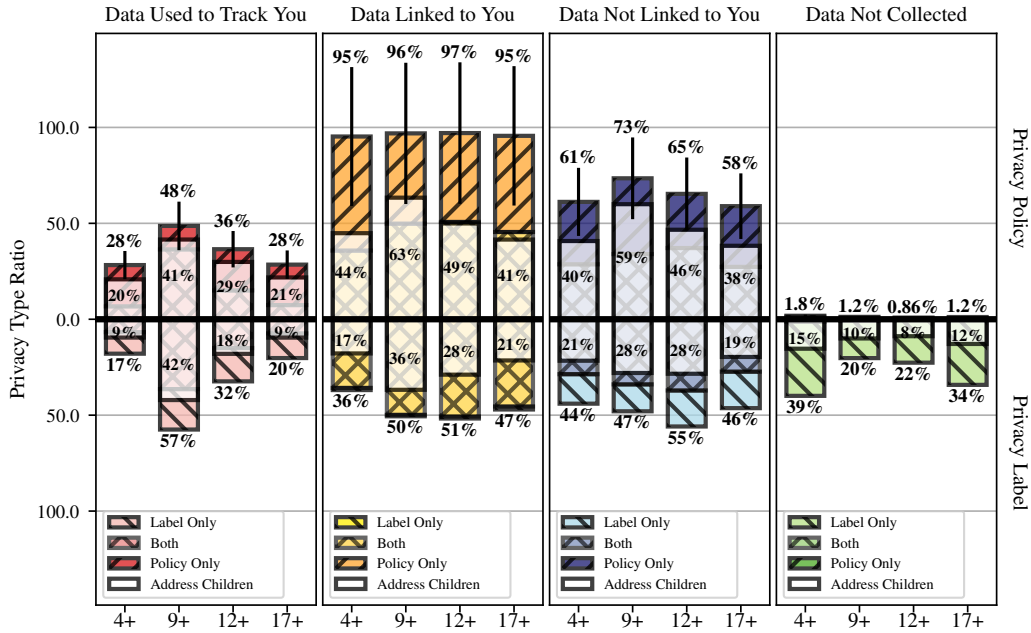


Figure 6: The ratios of the content ratings for each of the four privacy types, with an overlay (white bar) indicating the ratio of apps that also include a segment in their privacy policy, where they address privacy practices specific to children who engage with their services. The denominator is the number of apps with the designated content rating that have a privacy label. Please note that privacy types shown here are *not* mutually exclusive.

( $n = 74,829 \pm 58,889$ ), and 49% ( $n = 90,755 \pm 53,279$ ) of those apps also addressed children in their privacy policies. 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 [38] 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 [24]; 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.

We acknowledge that our findings do not implicate the evaluated apps of being in violation of COPPA [47], which, for example, allows PII collection some specific restrictions (e.g., geolocation) provided that this such not used for targeted/profiling of the minor and is obtained with informed parental or legal tutor consent. We highlight discrepancies between declared content ratings on the App Store, discrepancies in addressing this choice in privacy policies, and the need to ensure consistency in stated practices across platforms. Additionally, third party libraries offer configuration options for apps to help applications comply with COPPA regulations but prior work has shown that they are often misconfigured [76].

**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 the handling of user data. These pre-designed templates serve as a starting point that

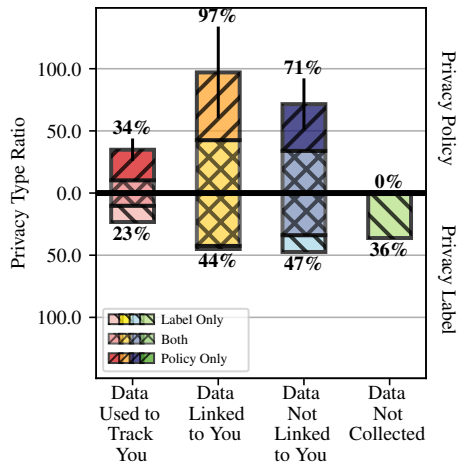


Figure 7: An overview of the privacy types associated with data collection on the App Store, from privacy labels and privacy policies, specific to apps whose policies are similar to templates. The denominator is the total number of such apps, i.e., 300,535 apps. Please note that the privacy types, except for *Data Not Collected*, are *not* mutually exclusive.

can be customized to meet specific requirements and legal obligations. By utilizing templates, businesses can save time and effort by avoiding the need to create a privacy policy 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. To achieve this, we did a broad search 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 the same, in-domain word embeddings that we used to train classifiers.

For each policy in our dataset, we conducted a compre-

hensive sentence-level comparison. We compared each sentence in a given policy against every sentence in a template. We employed the cosine similarity metric to measure the semantic resemblance between two sentences. Sentences were deemed similar if their cosine similarity exceeded a threshold of 0.8. In order to determine if a policy was derived from a template, we established a criterion: if over half the sentences in a policy were similar to over half of the sentences in the template being evaluated against, we identified the policy as being template-based.

We find that the privacy policies of 58% ( $n = 300,535$ ) of the apps potentially use templates. We further looked at the privacy labels declared by these apps on the App Store. We present our findings in Figure 7. Considering privacy types, 23%, 44%, 47%, and 36% 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. These findings are in line with those of all evaluated apps (see Figure 4).

We find that a majority of evaluated apps use template-like privacy policies, and that the use of templates possibly affects the discrepancies we see between declaration of data collection practices in privacy labels and privacy policies. Privacy policy templates are often made using generators that offer significant value by ensuring that developers thoroughly consider various data collection and sharing practices. These generators are similar to the process of creating privacy labels on the App Store. However, it is essential to recognize that templates are not one-size-fits-all solutions. It is crucial for developers to review and tailor policies derived from templates to accurately reflect the unique data collection practices of individual apps. By carefully reviewing and customizing policies, developers can ensure that their privacy practices are accurately communicated to users.

We find that a large number of policies in our dataset are similar to the policy template generated by the *Securiti Privacy Center*. The service provides multiple products that unify “data controls” including consent management, data requests, and compliance, among others. With support for third-party integrations, services like this automate the handling of numerous privacy-relevant mechanisms including privacy policies. While helpful in reducing that developers may face in the comprehensive generation and management of privacy policies, it may also

make it difficult for them to manually fill out privacy labels on the App Store in an accurate manner, serving as one potential reason for observed inconsistencies. Further research covering developer studies may help reveal the complicated nature of multiple disclosure formats, and their impact on the accuracy of said disclosures.

While our approach to template analysis illuminated noteworthy discoveries, this comparison methodology has some limitations. The set of templates used in our analysis is not exhaustive, representing only a sample of available templates. Additionally, our approach is optimized for shallow and direct reuse of template content, and may not capture all nuances of similarity. The templates we investigated are themselves similar and often overlap. As a result, some policies were similar to more than one template. We provide an overview of the templates used and the number of matches in [Table 4](#) in [Appendix B](#). Nevertheless, this approach allows us to identify a subset of policies that exhibit template-like characteristics. By doing so, we gain valuable insights into one of the factors influencing the transparency of data collection practices.

## 6 Case Studies

In the absence of verification by Apple itself, privacy labels (and privacy policies) may not provide full clarity into actual app practices. To shed light on the potential disparities between stated data collection practices and real-world app behavior, we conducted case studies of app behaviour, using network requests captured during app installation and usage to reported behaviors in the labels and policies.

We used an iPhone running iOS 15.7.5 (released April 2023) with a man-in-the-middle (MiTM) proxy [40] to gather outgoing traffic to determine which domains were being 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 proceeding to evaluate the next app in the sequence.

We included 30 apps in the analysis, split between (a) 15 apps that declare data collection for advertising pur-

poses in their privacy policies, but not on their privacy labels, and (b) 15 apps that declared a “Data Used to Track You” privacy label on the App Store, but could not be inferred from their privacy policies. We then compared the domains in the captured network requests against EasyList, EasyPrivacy, and WhoTracksMe to identify trackers [18, 51, 59]. We provide an overview of our findings in [Table 6](#) in [Appendix D](#).

The analysis presented in this study is an exploratory case study of 30 apps’ network behavior, and should not be considered representative of the practices of all apps on the App Store. The goal is to highlight examples of differences between stated data collection practices and observed app behavior, and provide insights into potential causes for discrepancies, which do not mean that developers are dishonest.

We find that the evaluated apps contact numerous tracking domains, with Facebook and Google being most prominent. Further, developers often do not include the use of analytics libraries within their purview of tracking, but guides from these libraries show that their practices are more nuanced [28, 53]. Additionally, we find that inconsistencies between privacy disclosures and network traffic persist across different app categories. When privacy policies mention third party libraries, they refer 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.

**Template Reuse.** Developers with multiple apps on the App Store reuse the privacy policies that are linked with individual apps. While for some developers this practice may result from the reuse of generic templates, this reuse can also be observed across organizations with multiple services. For example, Lexington Law and CreditRepair (#1 & #2 in [Table 6](#)) are associated with different developer accounts and 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. It is important that developers update templates to ensure accurate data collection practices, which can then reflect in 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, DishCult, Hello Neighbor, Chemistry & Periodic Table, Demo (#4, #8, #10, #11 in [Table 6](#)) have policies that cover the collection and sharing of data from third-parties. To their credit, third party libraries provide guidelines and disclosure links for developers review before filling out their privacy labels and privacy policies (examples, [28, 83, 66, 53]). However, these guides include multiple caveats that can further complicate developers’ understanding, requiring them to process against their own 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 that wherein disclosure is deemed optional [2]. While apps like Modern Milkman, Script, and Best Buy (#6, #9, #14 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 advertising measurement purposes, indicating the developers’ understanding of such activity. However, despite the App Store requiring the disclosure of all data collection practices, it is possible that the developers’ interpretation of optional caveats affect their creation of privacy labels. For example, the period tracking app, Maya (#15 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 [66, 53].

**Understanding Apple’s Definition of Tracking.** Apple provides definition of practices that it considers to fall under *Tracking*, along with examples and caveats [2]. However, recent work has found that developers find it difficult to understand this definition and correctly declare data collection that is used to track users [64]. Apps like Axolochi, WebMD, and Food Network Magazine (#7,

#12, and #13 in [Table 6](#)) acknowledge the use of tracking technologies in their privacy policies. However, the absence of similar declaration in privacy labels can possibly source from confusion around their understanding of Apple’s definition of tracking.

Next, we provide an analysis of the possible reasons for discrepancies for apps that have a *Data Used to Track You* privacy type in the label on the App Store but prove difficult to automatically capture tracking practices from their privacy policies.

**Absence of Relevant Policy Statement.** The privacy policies of Shake Shack, Kika Keyboard, Photo Prints CVS, Everpix, and FloatMe (#16, #17, #18, #19, #20 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 in the absence of ground truth.

**Unclear Policy Statements.** Even when developers declare the 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 (#21, #22, #23 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 purposes.

**Complex Privacy Policy 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, the content of the policy can only be accessed behind further link(s), as is the case with apps like McDonalds, Episode (#26, #27 in [Table 6](#)). Additionally, the policy for Brain-Boom (#24 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 (#28, #29, #30 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 515,920 apps on the iOS App Store comparing their reported privacy behavior of their privacy policies to the privacy labels by performing automated NLP classification of the privacy policies using Polisis [56]. We find that the majority of apps are likely under-reporting data collection practices in their privacy labels as compared to their privacy policies. As many as  $56\pm 38\%$  of analyzed apps collect data that is potentially linked to users. We also find that almost all ( $97\pm 30\%$ ) 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 81% of paid apps indicate that they do not collect any data, while privacy policies indicate that the true number may be closer to only  $10\pm 30\%$  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  $46\pm 0.01\%$  of these apps mention data collected from children in their privacy policies.

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 misapplication of labels and recommendations for improving the current state.

**Privacy Behavior Ground Truth.** Since Apple’s labels are not validated, as far as we can tell, we considered the privacy policies as a reasonable reference point of comparison. However, it is difficult to know the exact actual ground truth of privacy behavior, even if every app was fully dynamically and statically analyzed. In this paper, we compare privacy labels against privacy policies as a point of comparison of declaration of data practices across platforms. Privacy policies do not serve as ground truth for actual app behavior. Additionally, policies are not always written to clearly state behavior but instead to provide “cover” for organizations from legal and professional consequences. While there are a number of limitations to the approach we take in analyzing privacy policies using Polisis, we argue that the NLP methods of extracting free form text levels gets us closer to a viable understanding of data collection practices than the privacy la-

bels, as currently used. We believe that this is the case for two reasons.

First, Polisis could be prone to over-reporting privacy behavior. This stems from the fact that policies are analyzed on a per-segment basis, so discussions of data aggregation or anonymization that occurs in one paragraph, separate from the data that is collected, might appear as data linking when it is in fact not linked. However, even in these case, the behavior of the app may still remain ambiguous according to the privacy policy with regard to which data is aggregated or anonymized. In many cases data could be linked based on the use of unique identifiers stated in other places in the policies. Even if Polisis misclassifies and over-reports a reasonable fraction of these, it would still suggest that large numbers of apps are mislabeled.

At the same time, we also believe that there are significant cases of under-reporting by Polisis due to both how Apple links to privacy policies and the use of secondary privacy policies from third-party libraries. Many privacy policies link to other privacy policies that were not analyzed as part of our analysis. The App Store also often links to the developers’ and not the specific apps’ privacy policy. These policies usually address all services provided by the developer. For example Subsplash [17] and ChowNow [38] affect thousands of apps, and it is unknown how that data is used by the eventual customer and if that is reflected in the policies. In each of these cases, it is likely that additional behaviors or practices are *not* being captured by Polisis, again suggesting that Polisis is far from a ground truth, but does at least suggest that the labeling is much farther from accurate than we would like.

**Source of Confusion Around Privacy Labels.** It may also be the case that the processes for generating a privacy policy, which may include legal staff, is quite different from those selecting the labels, which could be left to the development team submitting the app to the store. This could be part of the cause for confusion with respect to the kinds of data covered by the privacy label (as compared to what is in the policy), in addition to what would be considered linked or not linked to users. For example, a recent study by Li et al. [64] showed that developers find it difficult to correctly identify data that is linked to users and data that is used to track users.

Additionally, the privacy labels data collection *Pur-*

*poses* cover broad topics, such as *App Functionality*, *Analytics*, and *Other Purposes*. There may be confusion from developers on what should be labeled as “privacy” relevant as the the labels cover more than just obvious privacy behavior. It may be the case that some data collection may not appear to be related to privacy to many developers, and while it is covered in the privacy policy, the fact that it is collected (or perhaps even covered in the labels) may be missing from the developers model of what a privacy label is and what kind of information it is described to convey. 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 unique identifiers may also be collected, as well as when *other* kinds of data is collected that this should match to the *Other Purposes* category. We argue that this is not necessarily the fault of the developer, but that better guidance and education is 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 continually shows that privacy policies are neither well understood [75] nor actively reviewed by most users [58]. In contrast, privacy labels are now forward facing, published directly on the App Store without needing to follow any links to review. Recent results by Garg et al. [50] have even suggested that the privacy labels can reduce app demand in cases of collecting sensitive information. The incentive for privacy labels may be an economic one 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 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 privacy label information content could be expanded to include more explicit details about data col-

lection behaviors, some of which may indeed be crucial to users for making meaningful and informed decisions about whether to install an app on their personal computing devices. However, balance is needed 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. Unfortunately it appears that privacy labels suffer from the transparency paradox [70]: the inherent conflict between transparency of textual meaning and the transparency of information-handling practices.

**Improved NLP Models for Privacy Labels.** Polisis and other similar approaches [56, 90, 85] offer much promise to helping to verify additional labeling of apps, like privacy labels. However, these approaches have a number of shortcomings as they were not designed for this task. Foremost, the analysis process is on a per-segment basis, which are useful in inferring practices that are completely described in individual policy segments. However, policies that describe practices in parts, across multiple segments are not properly captured. This is in part due to the design of the models, but also the training data (OPP-115 dataset [85]) which is labeled on a per-segment basis.

Additionally, given that privacy labels are being adapted more broadly, including by Google in Android [52], 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, and this could form the basis for new NLP models that could form the basis for developing more reliable tools that can assist in this emerging space for developers, researchers, and regulators.

**Recommendations for Apple.** With recent studies highlighting that privacy labels are hard to understand [64, 89], Apple could reconsider the taxonomy and descriptions of privacy labels, perhaps even removing the word “privacy” from the description. The connotation of “privacy” may indicate that the labeling should only capture some behaviors, but in fact they are designed to capture a much wider swath of data collection that may or may not always be privacy relevant.

Apple’s lack of obvious vetting or regulation of the privacy labels also may not incentivize the creation of accurate labels, particularly without any feedback to developers. The Polisis framework, while imperfect, is capable



of scanning many thousands of privacy policies relatively quickly with reasonable resources and could provide, at least, a first level review process for developers to consider wider arrays of labels for their apps. Given that Apple imposes a short embargo to review new apps before posting to the store, some form of policy based analysis could be incorporated into the review process.

## References

- [1] ALDI USA. URL: <https://apps.apple.com/us/app/aldi-usa/id429396645>.
- [2] App Privacy Details - App Store. URL: <https://developer.apple.com/app-store/app-privacy-details/>.
- [3] Axolochi. URL: <https://apps.apple.com/us/app/axolochi/id1432184360>.
- [4] Bagelman. URL: <https://apps.apple.com/us/app/bagelman/id1254203081>.
- [5] Chemistry & Periodic Table. URL: <https://apps.apple.com/us/app/chemistry-periodic-table/id493558583>.
- [6] Credit Karma on the App Store. URL: <https://apps.apple.com/us/app/credit-karma/id519817714>.
- [7] DappyTKeys. URL: <https://apps.apple.com/us/app/dappytkeys/id1536818077>.
- [8] El Charrito. URL: <https://apps.apple.com/us/app/el-charrito/id1080457082>.
- [9] FamilyLife®. URL: <https://apps.apple.com/us/app/familylife/id903170704>.
- [10] Firefox Reader View for clutter-free web pages | Firefox Help. URL: <https://support.mozilla.org/en-US/kb/firefox-reader-view-clutter-free-web-pages>.
- [11] HomeState, A Texas Kitchen. URL: <https://apps.apple.com/us/app/homestate-a-texas-kitchen/id925093380>.
- [12] Pregnancy Tracker: Baby Bump. URL: <https://apps.apple.com/us/app/pregnancy-tracker-baby-bump/id1453373942>.
- [13] Puppeteer. URL: <https://pptr.dev/>.
- [14] Walmart - Shopping & Grocery. URL: <https://apps.apple.com/us/app/walmart-shopping-grocery/id338137227>.
- [15] WebMD: Symptom Checker. URL: <https://apps.apple.com/us/app/webmd-symptom-checker/id295076329>.
- [16] Subsplash | Church mobile apps, websites, media, giving & more, 2020. URL: <https://www.subsplash.com/>.
- [17] Subsplash Privacy Policy, 2020. URL: <https://www.subsplash.com/legal/privacy>.
- [18] EasyList, 2023. URL: <https://easylist.to/>.
- [19] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. The web never forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, CCS '14*, page 674–689, New York, NY, USA, 2014. Association for Computing Machinery. doi:10.1145/2660267.2660347.
- [20] Aldi US. Aldi: U.S. Privacy Policy, 2023. URL: <https://www.aldi.us/en/online-privacy-notice/>.
- [21] Benjamin Andow, Samin Yaseer Mahmud, Wenyu Wang, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Tao Xie. PolicyLint: Investigating internal privacy policy contradictions on google play. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 585–602, Santa Clara, CA, August 2019. USENIX Association. URL: <https://www.usenix.org/conference/usenixsecurity19/presentation/andow>.
- [22] Benjamin Andow, Samin Yaseer Mahmud, Justin Whitaker, William Enck, Bradley Reaves, Kapil

- Singh, and Serge Egelman. Actions speak louder than words: Entity-Sensitive privacy policy and data flow analysis with PoliCheck. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 985–1002. USENIX Association, August 2020. URL: <https://www.usenix.org/conference/usenixsecurity20/presentation/andow>.
- [23] Apple. Apple Catalog API. URL: <https://amp-api.apps.apple.com/v1/catalog/>.
- [24] Apple. App Store Review Guidelines - Apple Developer, 2020. URL: <https://developer.apple.com/app-store/review/guidelines/>.
- [25] Apple. Apple Developer Program License Agreement | Section 3.3.10, 2020. URL: <https://developer.apple.com/support/downloads/terms/apple-developer-program/Apple-Developer-Program-License-Agreement-20220606-English.pdf>.
- [26] Apple. Age Ratings - Apple Developer, 2022. URL: <https://developer.apple.com/help/app-store-connect/reference/age-ratings/>.
- [27] Apple Developer. App privacy labels now live on the App Store - Latest News - Apple Developer. URL: <https://developer.apple.com/news/?id=3wann9gh>.
- [28] AppsFlyer. Preparing for the App Store review—nutrition labels, February 2023. URL: <https://support.appsflyer.com/hc/en-us/articles/207032086-Preparing-for-the-App-Store-review-nutrition-labels>.
- [29] David G Balash, Mir Masood Ali, Xiaoyuan Wu, Chris Kanich, and Adam J Aviv. Longitudinal analysis of privacy labels in the apple app store. *arXiv preprint arXiv:2206.02658*, 2022. URL: <https://doi.org/10.48550/arXiv.2206.02658>.
- [30] Rebecca Balebako, Florian Schaub, Idris Adjerid, Alessandro Acquisti, and Lorrie Cranor. The impact of timing on the salience of smartphone app privacy notices. In *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*, SPSM '15, page 63–74, New York, NY, USA, 2015. Association for Computing Machinery. doi:10.1145/2808117.2808119.
- [31] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146, 2017. doi:10.1162/tacl\_a\_00051.
- [32] Braincake. Privacy Policy, 2023. URL: [http://braincake.net/pregnancytracker\\_privacy.html](http://braincake.net/pregnancytracker_privacy.html).
- [33] Travis D. Breaux and Ashwini Rao. Formal analysis of privacy requirements specifications for multi-tier applications. In *2013 21st IEEE International Requirements Engineering Conference (RE)*, pages 14–23, 2013. doi:10.1109/RE.2013.6636701.
- [34] Duc Bui, Yuan Yao, Kang G. Shin, Jong-Min Choi, and Junbum Shin. Consistency analysis of data-usage purposes in mobile apps. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, CCS '21, page 2824–2843, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3460120.3484536.
- [35] Center for Food Safety and Applied Nutrition. How to Understand and Use the Nutrition Facts Label, February 2022. Publisher: FDA. URL: <https://www.fda.gov/food/new-nutrition-facts-label/how-understand-and-use-nutrition-facts-label>.
- [36] Chemistry. Privacy Notice, 2023. URL: <https://getchemistry.io/privacy-policy/ios/>.
- [37] 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. Demystifying hidden privacy settings in mobile apps. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 570–586, 2019. doi: [10.1109/SP.2019.00054](https://doi.org/10.1109/SP.2019.00054).
- [38] Chownow. Online food ordering system for restaurants, 2022. URL: <https://get.chownow.com/>.
- [39] Chownow. Privacy policy - chownow, 2022. URL: <https://get.chownow.com/privacy-policy>.
- [40] Aldo Cortesi, Maximilian Hils, Thomas Kriebaumer, and contributors. mitmproxy: A free and open source interactive HTTPS proxy, 2010–. [Version 9.0]. URL: <https://mitmproxy.org/>.
- [41] Lorrie Faith Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *J. on Telecomm. & High Tech. L.*, 10:273, 2012. URL: [http://jthtl.org/content/articles/V10I2/JTHTLv10i2\\_Cranor.PDF](http://jthtl.org/content/articles/V10I2/JTHTLv10i2_Cranor.PDF).
- [42] Lorrie Faith Cranor, Candice Hoke, Pedro Leon, and Alyssa Au. 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, March 2014. URL: <https://papers.ssrn.com/abstract=2418590>.
- [43] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. Ask the Experts: What Should Be on an IoT Privacy and Security Label? In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 447–464, San Jose, CA, USA, May 2020. IEEE. ISSN: 2375-1207. doi: [10.1109/SP40000.2020.00043](https://doi.org/10.1109/SP40000.2020.00043).
- [44] Pardis Emami-Naeini, Janarth Dheenadhayalan, Yuvraj Agarwal, and Lorrie Faith Cranor. Which Privacy and Security Attributes Most Impact Consumers’ Risk Perception and Willingness to Purchase IoT Devices? In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 519–536, San Francisco, CA, USA, May 2021. IEEE. doi: [10.1109/SP40001.2021.00112](https://doi.org/10.1109/SP40001.2021.00112).
- [45] Steven Englehardt and Arvind Narayanan. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, CCS ’16*, page 1388–1401, New York, NY, USA, 2016. Association for Computing Machinery. doi: [10.1145/2976749.2978313](https://doi.org/10.1145/2976749.2978313).
- [46] Federal Trade Commission. Privacy Online: Fair Information Practices in the Electronic Marketplace: A Federal Trade Commission Report to Congress, May 2000. URL: <https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf>.
- [47] Federal Trade Commission. Children’s Online Privacy Protection Rule (COPPA), July 2013. URL: <https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa>.
- [48] Federal Trade Commission. FTC Safeguards Rule: What Your Business Needs to Know, April 2022. URL: <https://www.ftc.gov/business-guidance/resources/ftc-safeguards-rule-what-your-business-needs-know>.
- [49] Jack Gardner, Yuanyuan Feng, Kayla Reiman, Zhi Lin, Akshath Jain, and Norman Sadeh. Helping mobile application developers create accurate privacy labels. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 212–230, 2022. doi: [10.1109/EuroSPW55150.2022.00028](https://doi.org/10.1109/EuroSPW55150.2022.00028).
- [50] Rajiv Garg and Rahul Telang. Impact of app privacy label disclosure on demand: An empirical analysis. *Workshop on the Economics of Information Security (WEIS)*, 2022. URL: <https://weis2022.econinfosec.org/wp-content/uploads/sites/10/2022/06/weis22-telang.pdf>.
- [51] Ghostery. WhoTracks.me - Bringing Transparency to Online Tracking, 2023. URL: <https://whotracks.me/>.

- [52] Google. Provide information for Google Play’s Data safety section - Play Console Help, 2020. URL: <https://support.google.com/googleplay/android-developer/answer/10787469>.
- [53] Google. Privacy Disclosures Policy, May 2023. URL: <https://support.google.com/analytics/answer/7318509>.
- [54] Catherine Han, Irwin Reyes, Amit Elazari Bar On, Joel Reardon, Álvaro Feal, Serge Egelman, Narseo Vallina-Rodriguez, et al. 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.*, San Francisco, CA, USA, 2019. IEEE. URL: <https://www.ieee-security.org/TC/SPW2019/ConPro/papers/han-conpro19.pdf>.
- [55] Catherine Han, Irwin Reyes, Álvaro Feal, Joel Reardon, Primal Wijesekera, Narseo Vallina-Rodriguez, Amit Elazari Bar On, Kenneth A. Bamberger, and Serge Egelman. The price is (not) right: Comparing privacy in free and paid apps. *Proc. Priv. Enhancing Technol.*, 2020(3):222–242, 2020. doi: 10.2478/popets-2020-0050.
- [56] Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G. Shin, and Karl Aberer. Polis: Automated analysis and presentation of privacy policies using deep learning. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 531–548, Baltimore, MD, August 2018. USENIX Association. URL: <https://www.usenix.org/conference/usenixsecurity18/presentation/harkous>.
- [57] HyperBeard Games. Privacy Policy: HyperBeard Games, 2023. URL: [https://docs.google.com/document/d/1bJpQtjxoxmf1leATVgnRbZWkV9W-iDngLx\\_pTsvu8Vk/edit](https://docs.google.com/document/d/1bJpQtjxoxmf1leATVgnRbZWkV9W-iDngLx_pTsvu8Vk/edit).
- [58] Carlos Jensen and Colin Potts. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 471–478, 2004.
- [59] Arjaldo Karaj, Sam Macbeth, Rémi Berson, and Josep M. Pujol. WhoTracks .Me: Shedding light on the opaque world of online tracking, April 2019. arXiv:1804.08959 [cs]. URL: <http://arxiv.org/abs/1804.08959>, doi:10.48550/arXiv.1804.08959.
- [60] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W Reeder. A “nutrition label” for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, pages 1–12, 2009.
- [61] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. Standardizing privacy notices: an online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’10*, pages 1573–1582, New York, NY, USA, April 2010. Association for Computing Machinery. doi:10.1145/1753326.1753561.
- [62] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3393–3402, Paris France, April 2013. ACM. URL: <https://dl.acm.org/doi/10.1145/2470654.2466466>, doi:10.1145/2470654.2466466.
- [63] Konrad Kollnig, Anastasia Shuba, Max Van Kleek, Reuben Binns, and Nigel Shadbolt. Goodbye tracking? impact of ios app tracking transparency and privacy labels. In *2022 ACM Conference on Fairness, Accountability, and Transparency, FAccT ’22*, page 508–520, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3531146.3533116.
- [64] Tianshi Li, Kayla Reiman, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I. Hong. Understanding challenges for developers to create accurate privacy nutrition labels. In *CHI Conference on Human Factors in Computing Systems, CHI ’22*, New York, NY,

- USA, 2022. Association for Computing Machinery. doi:10.1145/3491102.3502012.
- [65] Aleecia M. McDonald and Lorrie Faith Cranor. The Cost of Reading Privacy Policies. *HeinOnline*, 4(3):543–568, 2009. URL: <https://heinonline.org/HOL/P?h=hein.journals/isjlp soc4&i=563>.
- [66] Meta. Resources for Completing App Store Data Practice Questionnaires for Apps That Include the Facebook or Audience Network SDK, July 2022. publisher: Meta. URL: <https://developers.facebook.com/blog/post/2022/07/18/resources-for-completing-app-store-data-practice-questionnaires-apps-facebook-or-audience-network-sdk/>.
- [67] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 3111–3119, 2013. URL: <https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html>.
- [68] Mozilla. Readability.js, 2020. URL: <https://github.com/mozilla/readability>.
- [69] Collins W. Munyendo, Yasemin Acar, and Adam J. Aviv. Desperate times call for desperate measures: User concerns with mobile loan apps in kenya. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 2304–2319, 2022. doi:10.1109/SP46214.2022.9833779.
- [70] Helen Nissenbaum. A contextual approach to privacy online. *Daedalus*, 140(4):32–48, 2011.
- [71] Office of the Attorney General of California. California Consumer Privacy Act (CCPA), October 2018. URL: <https://oag.ca.gov/privacy/ccpa>.
- [72] David P. Wiggins. xvfb(1): virtual framebuffer X server for X - Linux man page. URL: <https://linux.die.net/man/1/xvfb>.
- [73] Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, October 2014. Association for Computational Linguistics. URL: <https://aclanthology.org/D14-1162>, doi:10.3115/v1/D14-1162.
- [74] Sebastian Raschka. Creating Confidence Intervals for Machine Learning Classifiers | Method 3: Bootstrapping the Test Set Predictions, April 2022. URL: <https://sebastianraschka.com/blog/2022/confidence-intervals-for-ml.html>.
- [75] 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. Disagreeable privacy policies: Mismatches between meaning and users’ understanding. *Berkeley Tech. LJ*, 30:39, 2015.
- [76] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpahan, Narseo Vallina-Rodriguez, and Serge Egelman. “won’t somebody think of the children?” examining coppa compliance at scale. *Proceedings on Privacy Enhancing Technologies*, 3:63–83, 2018. URL: <https://petsymposium.org/2018/files/papers/issue3/popets-2018-0021.pdf>.
- [77] Benjamin Sanchez-Lengeling, Jennifer N. Wei, Brian K. Lee, Richard C. Gerkin, Alán Aspuru-Guzik, and Alexander B. Wiltschko. Machine Learning for Scent: Learning Generalizable Perceptual Representations of Small Molecules, October 2019. arXiv:1910.10685 [physics, stat]. doi:10.48550/arXiv.1910.10685.
- [78] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. A design space for effective

- privacy notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 1–17, Ottawa, July 2015. USENIX Association. URL: <https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub>.
- [79] Rocky Slavin, Xiaoyin Wang, Mitra Bokaei Hosseini, James Hester, Ram Krishnan, Jaspreet Bhatia, Travis D. Breaux, and Jianwei Niu. Toward a framework for detecting privacy policy violations in android application code. In *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*, pages 25–36, 2016. doi:10.1145/2884781.2884855.
- [80] Staff FTC. Protecting consumer privacy in an era of rapid change—a proposed framework for businesses and policymakers. *Journal of Privacy and Confidentiality*, 3(1), 2011. URL: <https://www.ftc.gov/reports/protecting-consumer-privacy-era-rapid-change-recommendations-businesses-policymakers>.
- [81] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. Learning sentiment-specific word embedding for Twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1555–1565, Baltimore, Maryland, June 2014. Association for Computational Linguistics. URL: <https://aclanthology.org/P14-1146>, doi:10.3115/v1/P14-1146.
- [82] The Alan Turing Institute. ReadabiliPy, 2018. URL: <https://github.com/alan-turing-institute/ReadabiliPy>.
- [83] Unity. Apple privacy survey for Unity Ads, 2023. URL: <https://docs.unity.com/ads/en/manual/ApplePrivacySurvey>.
- [84] WebMD. WebMD Privacy Policy, 2023. URL: <https://www.webmd.com/about-webmd-policies/about-privacy-policy>.
- [85] 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. The creation and analysis of a website privacy policy corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1330–1340, Berlin, Germany, August 2016. Association for Computational Linguistics. URL: <https://aclanthology.org/P16-1126>, doi:10.18653/v1/P16-1126.
- [86] Maximiliane Windl, Niels Henze, Albrecht Schmidt, and Sebastian S. Feger. Automating contextual privacy policies: Design and evaluation of a production tool for digital consumer privacy awareness. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI '22*, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3491102.3517688.
- [87] Yue Xiao, Zhengyi Li, Yue Qin, Xiaolong Bai, Jiale Guan, Xiaojing Liao, and Luyi Xing. Lalaine: Measuring and characterizing non-compliance of apple privacy labels at scale. *CoRR*, abs/2206.06274, 2022. arXiv:2206.06274, doi:10.48550/arXiv.2206.06274.
- [88] Le Yu, Xiapu Luo, Xule Liu, and Tao Zhang. Can we trust the privacy policies of android apps? In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 538–549, 2016. doi:10.1109/DSN.2016.55.
- [89] Shikun Zhang, Yuanyuan Feng, Yaxing Yao, Lorie Faith Cranor, and Norman Sadeh. How usable are ios app privacy labels? *Proceedings on Privacy Enhancing Technologies*, 4:204–228, 2022.
- [90] Sebastian Zimmeck, Peter Story, Daniel Smullen, Abhilasha Ravichander, Ziqi Wang, Joel R. Reidenberg, N. Cameron Russell, and Norman M. Sadeh.

Maps: Scaling privacy compliance analysis to a million apps. *Proceedings on Privacy Enhancing Technologies*, 2019:66 – 86, 2019.

- [91] Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven M. Bellovin, and Joel Reidenberg. Automated analysis of privacy requirements for mobile apps. In *Proceedings 2017 Network and Distributed System Security Symposium*, 2017. doi:10.14722/ndss.2017.23034.

## A Reimplementing and Training Polisis

**Polisis framework.** At its core, Polisis relies on 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. [85], 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. Of these, 4 attributes were relevant to the creation of privacy labels – we expand on their use later in this section.

**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. [56], 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.

**Word Embeddings.** Text classifiers deal with text by representing their features as building blocks. A simple example of this would consider the *frequency* of occurrence of each word as a feature used to train classifiers. This approach is limited in its ability to interpret words outside of the dataset used to train classifiers, hence limiting its ability to generalize.

Word embeddings offer a different approach by extracting vector representations of words, in an unsupervised manner, from a large corpus of text. These representations can directly be used as features to classifiers, and also account for words that the classifier has not observed in its training phase.

General purpose embeddings like GloVe [73] and Word2vec [67], have been trained on large text corpuses and provide useful word representations for most use cases. However, domain-specific embeddings generates better classification results [81]. We therefore reached out to Harkous et al. [56] and requested a copy of their corpus of 130K privacy policies of apps on the Google Play Store. We then trained a word-embeddings model using *fastText* [31], which also helped create representations for subwords, and account for words outside of the policies

Table 3: Classification results for the attributes that were used in the creation of Privacy Labels.

Polis Output	Classification Report				Bootstrapping Test Set (200 samples)		
	Precision	Recall	F1	Support (Presence/Absence)	Accuracy	Lower 95% CI	Upper 95% CI
<b>Segment Classifier</b>							
First Party Collection/Use	0.80	0.79	0.80	300/457	0.80	0.77	0.83
Third Party Collection/Sharing	0.88	0.85	0.86	271/486	0.88	0.85	0.90
Average	0.84	0.82	0.83		0.84	0.81	0.87
<b>Identifiability</b>							
Identifiable	0.75	0.76	0.75	166/109	0.76	0.71	0.81
Aggregated or Anonymized	0.85	0.85	0.85	131/144	0.85	0.81	0.89
Average	0.80	0.80	0.80		0.81	0.76	0.85
<b>Does/Does Not</b>							
Does Not	0.91	0.80	0.84	146/461	0.89	0.87	0.92
Average	0.91	0.80	0.84		0.89	0.87	0.92
<b>Purpose</b>							
Additional Service/Feature	0.82	0.79	0.80	583/1040	0.82	0.80	0.84
Advertising	0.87	0.84	0.86	487/1136	0.88	0.86	0.90
Analytics/Research	0.86	0.85	0.85	595/1028	0.86	0.85	0.88
Basic Service/Feature	0.80	0.80	0.80	738/885	0.80	0.78	0.82
Legal Requirement	0.92	0.83	0.87	219/1404	0.94	0.93	0.95
Marketing	0.84	0.82	0.83	554/1069	0.85	0.83	0.86
Merger	1.00	0.88	0.93	59/1564	0.99	0.99	1.00
Personalization	0.86	0.80	0.82	388/1235	0.88	0.86	0.89
Service Operation and Security	0.86	0.81	0.83	251/1372	0.91	0.90	0.93
Unspecified	0.81	0.73	0.76	437/1186	0.83	0.81	0.84
Average	0.86	0.81	0.83		0.88	0.86	0.89
<b>Personal Information Type</b>							
Computer Information	0.94	0.91	0.92	325/1551	0.96	0.95	0.97
Contact	0.91	0.90	0.91	612/1264	0.92	0.90	0.93
Cookies and Tracking Elements	0.95	0.87	0.91	475/1401	0.94	0.93	0.95
Demographic	0.90	0.86	0.88	271/1605	0.94	0.92	0.95
Financial	0.94	0.90	0.92	227/1649	0.97	0.96	0.97
Generic Personal Information	0.82	0.81	0.81	617/1259	0.83	0.82	0.85
Health	0.95	0.66	0.74	31/1845	0.99	0.98	0.99
IP Address and Device IDs	0.97	0.89	0.92	372/1504	0.95	0.94	0.96
Location	0.91	0.85	0.88	319/1557	0.93	0.92	0.95
Personal Identifier	0.95	0.77	0.83	75/1801	0.98	0.97	0.99
Social Media Data	0.93	0.82	0.86	60/1816	0.99	0.98	0.99
Survey Data	0.93	0.84	0.88	60/1816	0.99	0.98	0.99
User Online Activities	0.88	0.87	0.88	629/1247	0.89	0.87	0.90
User Profile	0.90	0.82	0.86	212/1664	0.95	0.94	0.96
Unspecified	0.81	0.81	0.81	870/1006	0.81	0.80	0.83
Average	0.91	0.84	0.86		0.94	0.92	0.95
<b>Audience Type</b>							
Children	0.99	0.99	0.99	59/65	0.99	0.98	1.00
Average	0.99	0.99	0.99		0.99	0.98	1.00
<b>Action First Party</b>							
Collect on website	0.77	0.66	0.67	445/164	0.79	0.76	0.82
Collect in mobile app	0.82	0.75	0.78	66/543	0.92	0.91	0.95
Average	0.80	0.71	0.73		0.86	0.84	0.89
<b>Action Third-Party</b>							
Collect on first party website/app	0.84	0.80	0.82	115/377	0.87	0.84	0.90
See	0.90	0.73	0.79	58/434	0.93	0.91	0.95
Average	0.87	0.77	0.80		0.90	0.88	0.93



in the training corpus.

**Hyperparameters and Evaluation Metrics.** The hyperparameters for each classifier were determined using a randomized grid-search. We adopted a similar approach to that of Harkous et al. [56], and evaluated classifiers based on the precision, recall, and F1 scores, macro-averaged per label. We present the evaluation results in Table 3, where the numbers show the classifiers’ ability to detect both, the *presence* and *absence* of a label in a given text segment. We also report the accuracy and confidence intervals for each classifier result after performing bootstrap sampling on the test set [74, 77].

## B Templates

We provide an overview of 15 templates in Table 4.

## C Additional Tables and Figures

We have moved additional tables and figures here. We intend to move these back to the main text in the final submission.

## D Network Traffic Collection

We provide an overview of the analysis of 30 apps in Table 6.

## E 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 [16] is used by 8,015 apps of local churches on the App Store (examples, [9, 7]).

All of the hosted apps link to the same privacy policy [17] 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 Polisio 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, Polisio 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.

Table 4: Overview of privacy policy templates used to compare policies. Please note that matched policies are not mutually exclusive, i.e., policies can be found to be similar to more than one template.

Template	#Policies	#Apps	Data Used to Track You			Data Linked to You			Data Not Linked to You			Data Not Collected		
			Label Only	Policy Only (+/- 0.26)	Overlap	Label Only	Policy Only (+/- 0.38)	Overlap	Label Only	Policy Only (+/- 0.29)	Overlap	Label Only	Policy Only (+/- 0.30)	Overlap
Securiti Privacy Center	145759	221211	30877	51371	21619	4789	121108	94959	28267	85358	76230	78976	348	188
WebsitePrivacyPolicyGenerator.com	94170	131258	20363	23638	9555	4049	80955	45629	19539	49929	40513	56827	174	362
Termly	89275	183697	25870	47995	21985	4909	94748	84013	27524	64195	61567	60331	335	166
TermsFeed	86199	126844	19964	26507	9751	1947	74712	49945	14464	52646	42936	48868	57	68
Enzuzo	80837	140996	21543	31389	14184	2531	76193	61901	19317	51293	47137	47615	75	95
FreePrivacyPolicy	36224	88209	15513	18560	9921	2152	45845	40121	15949	28306	25494	29493	33	68
App Privacy Policy Generator	31786	39479	10787	1510	821	1015	27752	10895	4641	16892	13048	19115	20	59
PrivacyPolicyGenerator.info	25504	53020	10479	8202	5205	3535	32314	17753	13379	14379	12686	24267	94	230
PrivacyPolicies.com	18292	41740	5880	9391	4610	1118	23125	17404	7321	12579	13402	14796	19	28
PandaDoc	15306	30135	6729	4403	2189	2572	19933	8250	8698	7388	5834	15859	105	161
PrivacyPolicyGenerator.org	13429	24383	5900	3631	1798	2415	16028	6382	6940	6436	4600	13933	67	171
GetTerms	5274	11840	1757	1945	1395	319	6326	5089	2358	3567	3372	4250	1	6
iubenda	3718	6632	566	1465	699	22	4372	2176	1435	2168	2283	2041	0	0
WebsitePolicies.com	1993	3302	132	1165	355	2	2130	1170	132	1628	1362	1347	0	0
Shopify	371	540	123	55	22	230	274	254	468	32	18	188	0	0

Table 5: 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 Polisii.

Label \ Policy	Data Used to Track You	Data Linked to You	Data Not Linked to You	Data Not Collected
Data Used to Track You	<b>40,302</b>	66,390 (39,535)	79,506 (49,827)	35,930 (35,930)
Data Linked to You	101,974 (33,696)	<b>202,082</b>	228,260 (114,309)	173,441 (173,441)
Data Not Linked to You	62,074 (17,013)	124,861 (54,322)	<b>142,330</b>	94,502 (94,502)
Data Not Collected	1,048	1,382	1,865	<b>5,310</b>

**ChowNow.** ChowNow [38] is an app platform used by 3,182 different apps of local restaurants to receive online orders for takeout and delivery (examples, [8, 11, 4]).

All apps using the ChowNow platform link to the same privacy policy [39] 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 Polisii 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 [24]; 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 [14] 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 [6] 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 [1]. 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 [20], 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 [3]. The app's privacy policy [57] 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 [15]. Their privacy policy [84] 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.

**Chemistry & Periodic Table.** An app under the *Education* category, this application assists users in solving chemical equations [5]. The app’s privacy label on the App Store declares the collection of *Identifiers*, *Usage Data*, and *Diagnostics*. However, their privacy policy [36] reveals that the app automatically collects information concerning the user’s country and location. Unfortunately, the developers did not include corresponding entry for the *Location* data category within the privacy label that they declared on App Store.

**Pregnancy Tracker.** The pregnancy tracking app developed by Fitness Labs has concerning discrepancies between its privacy label on the App Store [12] and its privacy policy [32]. 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 fails to provide a specific privacy type for *Data Linked to You* in its privacy label. Further, the privacy label does not adequately highlight the collection of multiple data categories, including *Identifiers*, *Financial Information*, *Contact Information*, and *Sensitive Information*.

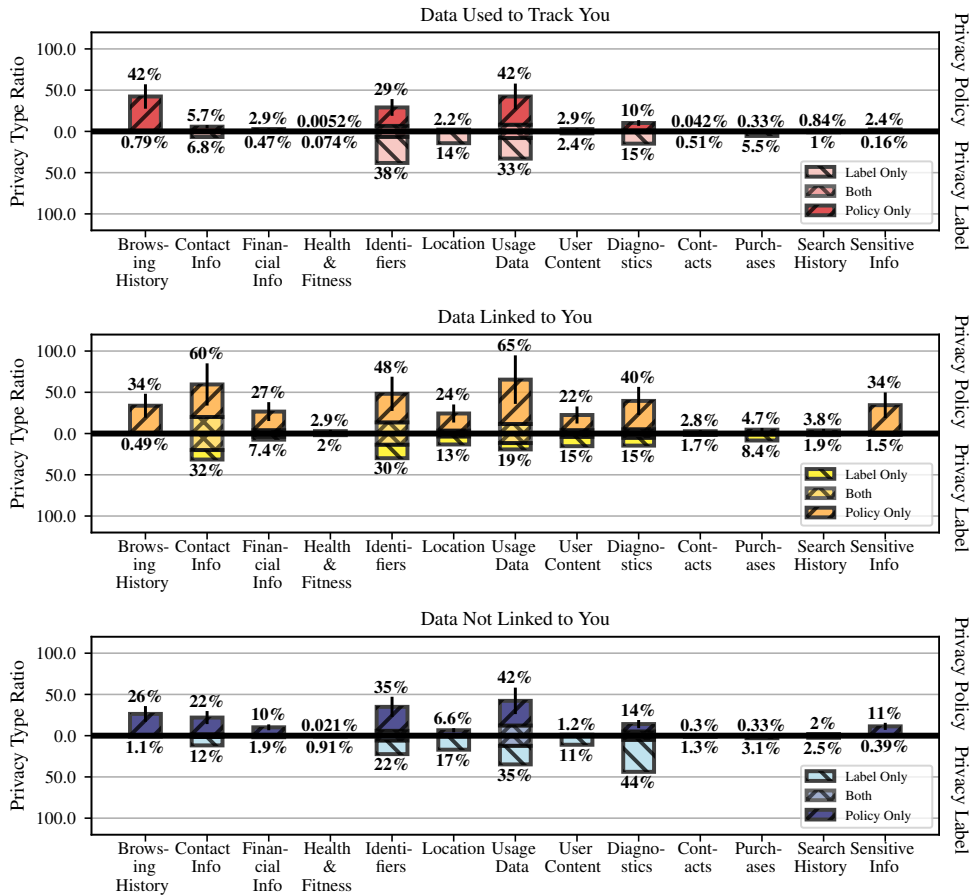


Figure 8: The ratios of data categories against privacy types. The denominator is the number of apps with the designated privacy type either in their privacy label or their privacy policy, i.e., 210,205 apps with *Data Used to Track You*, 497,124 apps with *Data Linked to You*, and 386,519 apps with *Data Not Linked to You*. The three privacy types shown here are *not* mutually exclusive.

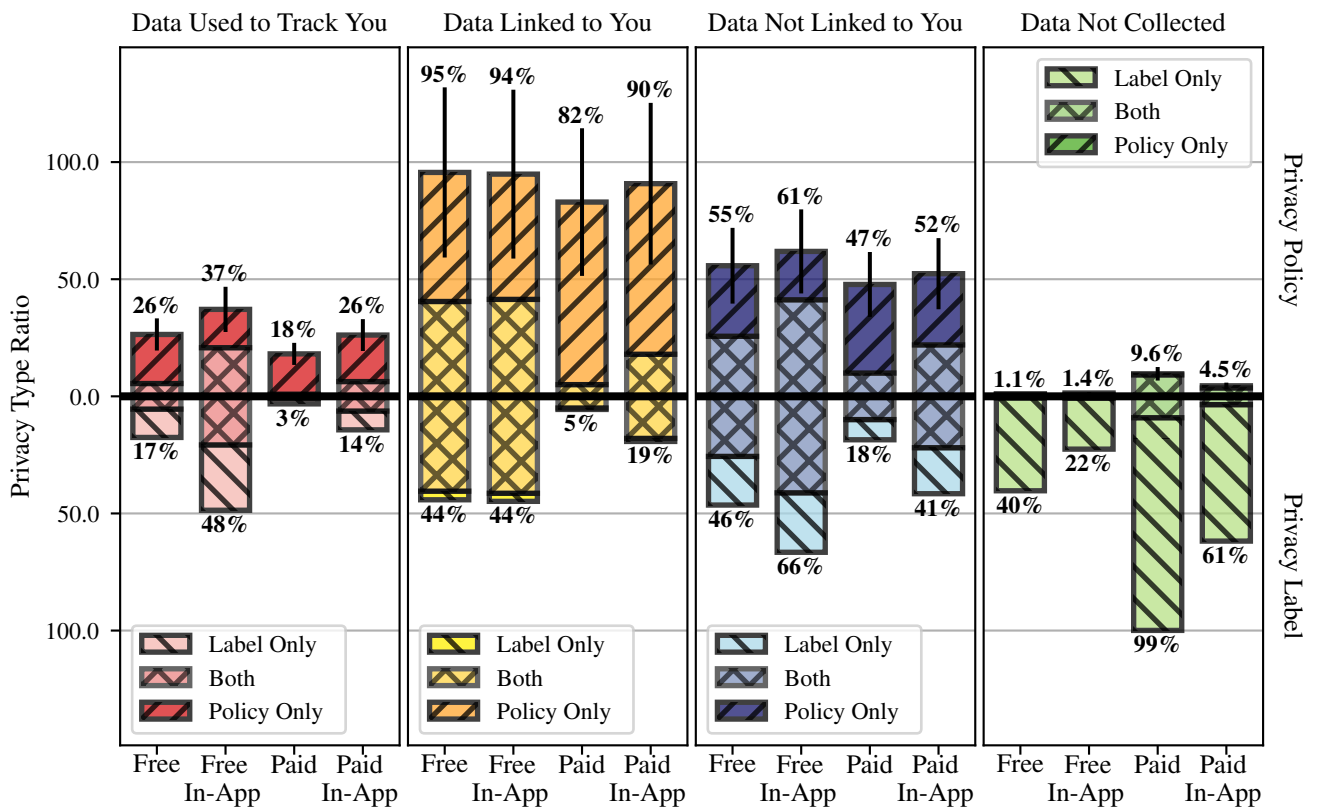


Figure 9: The ratios of app costs for each of the four privacy types. The denominator is the number of apps with the designated app cost that have a privacy label. Free apps are more likely than paid apps to collect data, including data used to track and linked to users. Please note that privacy types shown here are *not* mutually exclusive.

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

#	App Name	Category	Declared in Privacy Label	Declared in Privacy Policy	Trackers	Notes
<b>Apps that do not declare tracking in their privacy label</b>						
1	Lexington Law	Finance	N	Y	Facebook, Google, Adobe, TheTradeDesk, LiveIntent, StackAdapt, Bing, TikTok, Taboola, Snapchat, Twitter	Policy Template Reuse
2	CreditRepair	Finance	N	Y	Facebook, Google, Adobe, StackAdapt, TTD, Twitter, Yahoo, LiveIntent, Taboola	Policy Template Reuse
3	Aldi	Shopping	N	Y	Adobe, Google	Incomplete understanding of App Store requirements
4	DishCult	Food & Drink	N	Y	Facebook, AppsFlyer	Incomplete understanding of third party collection
5	FoodHub	Food & Drink	N	Y	Facebook, Google, MoEngage	Incomplete understanding of third party collection
6	Modern Milkman	Food & Drink	N	Y	Facebook, Google, AppsFlyer, LiveIntent, BidSwitch, ShareThis	Incomplete understanding of App Store requirements
7	Axolochi	Games	N	Y	Google, SuperSonic, Unity	Incomplete understanding of App Store's definition of tracking
8	Hello Neighbor	Games	N	Y	Google, SuperSonic	Incomplete understanding of third party collection
9	Scripts - Learn Chinese Writing	Education	N	Y	Facebook, Google	Incomplete understanding of App Store's requirements
10	Chemistry & Periodic Table	Education	N	Y	Google	Incomplete understanding of third party collection
11	Demo - Song-writing Studio	Music	N	Y	Facebook, Google, AppsFlyer	Incomplete understanding of third party collection
12	WebMD	Medical	N	Y	Adobe, Google	Incomplete understanding of App Store's definition of tracking
13	Food Network Magazine	Food & Drink	N	Y	Facebook, Google	Incomplete understanding of App Store's definition of tracking
14	Best Buy	Shopping	N	Y	Adobe, Google, Criteo	Incomplete understanding of App Store's requirements
15	Maya Period Tracker	Health & Fitness	Partial	Y	Facebook, Google	Incomplete understanding of App Store's requirements
<b>Apps that declare tracking in their privacy label but have an unclear privacy policy</b>						
16	Shake Shack	Food & Drink	Y	N	Facebook, Google	Not Stated in Policy
17	Kika Keyboard	Utilities	Y	N	AppLovin, Facebook, Google	Not Stated in Policy
18	Photo Prints CVS	Photo & Video	Y	N	Facebook, Google	Not Stated in Policy
19	Everpix	Entertainment	Y	N	AppLovin, Facebook, Google, Liftoff	Not Stated in Policy
20	FloatMe	Finance	Y	N	Facebook, Google, AppsFlyer	Not Stated in Policy
21	Buffalo Wild Wings	Food & Drink	Y	N	Google	Not Clearly Stated in Policy
22	The General Auto-Insurance App	Finance	Y	N	Facebook, Google	Not Clearly Stated in Policy
23	Conservative News	News	Y	N	Amazon, AppLovin, Flurry, Google	Not Clearly Stated in Policy
24	BrainBoom	Games	Y	Y	AppLovin, Facebook, Google, Supersonic Ads, InMobi, TapJoy, IronSource, Vungle, AdColony	Presented as an image, difficult to parse
25	Stickman Boxing	Games	Y	Y	Amazon, AppLovin, Facebook, Google, IronSource, Supersonic Ads, TapJoy, Vungle, Yandex	Separate Declaration of Data Collection and Purpose.
26	McDonalds	Food & Drink	Y	Y	Adobe, Facebook, Google, Kochava	Policy segments linked on landing page
27	Episode: Choose Your Story	Games	Y	Y	Adjust, Facebook, Google	Policy linked behind a link on the landing page from App Store
28	JCPenney	Shopping	Y	Y	Adobe, Facebook, Google, UrbanAirship	Incorrect Policy Link. Different part of website
29	Dosh	Shopping	Y	Y	AppsFlyer, Google	Incorrect Policy Link. Different part of website
30	CDL Prep Test	Reference	Y	<del>N</del>	Google	Incorrect Policy Link. Link broken.