Beyond COPPA: Analysing Data Safety Across All Age Groups in Mobile Apps

Adamo Mariani
University of Twente
Enschede, The Netherlands
a.mariani@student.utwente.nl

ABSTRACT

Mobile applications have the potential to access a vast range of user personal information, and as their use in our daily lives continues to grow and diversify, so does their ability to collect user data. Studies have demonstrated the low compliance rate of apps in major stores with store developer policies and the vulnerabilities in the permission systems of Android and iOS used by such apps to access personal data. Most concerns and regulations surround the data privacy of minors, and research has been done exploring the troubling data safety practices in apps targeting children aged 12 and under. To our knowledge, current literature has yet to include Teens and Young adults when exploring the influence of an app's age demographics on patterns of permission requests. The differentiation between these age demographics is essential since, despite not having reached the age of majority, teenagers are expected to be able to grant and deny access to their data. Moreover, teenagers are treated as adults under most regulatory frameworks, giving developers more data collection freedom. Therefore, this research aims to investigate the relationship between data collection practices and privacy policy consistencies of mobile applications by targeted age groups. To investigate this correlation, this study attempts to fulfil three main contributions: (1) develop a classifier to determine an app's targeted age group based on information from its app listing, (2) compare apps from Google Play to those from the App Store to explore the differences in permission requests done by apps targeting various age groups across platforms, (3) compare an app's disclosed data safety information in their app listing with their privacy policy. The research is expected to help improve regulatory frameworks and propose a categorisation system to identify targeted age groups for further research.

KEYWORDS

Age Demographics, Data Safety, Mobile Apps, Permissions, Privacy Policy, Teens

TScIT '41, July 05, 2024, Enschede, NL

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

In the last few years, the mobile smartphone market has experienced explosive growth, which is only expected to continue. Mobile applications (apps) have become an integral part of our daily lives, with as many as 257 billion downloaded apps worldwide in 2023 [13]. The Android and iOS operating systems use permissions to control these apps' access to a user's sensitive data. Multiple studies have been conducted that assess such permission systems and shed light on their vulnerabilities [1, 3, 5, 20, 29, 33, 35]. Additionally, a variety of tools have been developed to analyse privacy policy inconsistencies, violations, and the presence of over-claimed permissions; where applications request more permissions than their functionality justifies. These will be explored further in section 2.

Results found by literature scrutinising mobile app listings on Google Play and Apple App Stores reveal a troubling disconnect between an application's stated requirements and the data it collects [7, 38]. Regulations such as the Children's Online Privacy Protection Act (COPPA) exist specifically to protect the handling of data belonging to children and safeguard their privacy [8]. However, it appears that many apps still fail to fully comply with these requirements, leading to potential breaches exploiting the vulnerabilities of young users [31]. Meyer et al. [28] concludes that 95% of apps targeting children below the age of 5 contain advertisements, and data collected by these apps is even more valuable to advertisers.

It is, therefore, important to study the relationship that an app's age demographic has on its requested permissions and determine if specific patterns emerge when examining the type and amount of permissions requested by apps targeting different age groups. As a consequence, research has been done investigating the behaviour and privacy policies of apps targeting children below the age of 12 [10, 11, 19, 22, 31, 37, 40]. However, we believe there is a need to develop a more distinct and broader categorisation of targeted age groups, which includes Teens and recognises apps for Children residing outside the "Kids" and "Family" categories of app stores. Differences between these age groups are of particular interest not only due to their high involvement with mobile applications but also because of their varying ability to read privacy policies and understand the possible privacy risks implicated, which teenagers are expected to be able to do. The European General Data Protection Regulation (GDPR) itself allows member states to lower the child age threshold to a minimum of 13[1].

Therefore, this research aims to analyse the influence an app's age demographic has on the data safety practices adopted, uncovering patterns that could inform future regulatory frameworks and raise awareness for end-users. This approach contributes to the broader topic of mobile application privacy and addresses a critical

[1] GDPR Article 8(1)

gap in understanding the privacy risks associated with younger mobile app users of varying age categories.

To achieve this goal, the following research question will be used as the basis of this study:

> How do data collection practices comply with privacy policies among applications targeting different age groups?

This can be answered through the following sub-questions:

(1) How accurately can the targeted age of mobile applications be predicted using mobile store listing information?
(2) How do data collection practices differ between mobile applications targeting specific age groups on the Google Play and Apple App stores?
(3) To what extent do data safety practices disclosed in privacy policies match those found on mobile store listings?

The official Google Play guidelines differentiate between the following targeted age groups: 5–, 6-8, 9-12, 13-15, 16-17, 18+; with additional policies applied to apps targeting children under the age of 12 [16]. The App Store considers similar age bands of 4+, 9+, 12+, and 17+ [2]. To answer this paper's main research question, while keeping the focus on younger age demographics, we have grouped the age groups as Children (12–), Teens (13-17), and Adults (18+).

2 RELATED WORK

As mentioned in section 1, plenty of research has been conducted in the realm of mobile application privacy, investigating the effectiveness of mobile permission systems used by apps and developing tools to aid in the process. This section will explore some of the related literature identified from IEEE, Scopus, Google Scholar and Research Rabbit. Considering the dynamic nature of mobile app store policies, regulations, and permissions, the focus has been placed on studies published from 2018 onwards.

Luo et al. [23] has categorised privacy policy violations by application category in some of the largest Android app stores. Results show that the "Games" category on the Xiaomi app store has the highest number of privacy violations, whereas the "Education" category ranks fourth on the Google Play store. The main target audience of apps within those categories is almost exclusively composed of children and teens, with reports citing that more than 90% are available to children below the age of 12 [30].

Several studies have been done that solely take into account apps published for children found under the "Family" and "Kids" categories of the Google Play and Apple stores [10, 11, 19, 21, 22, 31, 37, 40]. These studies show there are problems surrounding privacy policy violations in apps targeting children, but given the stricter regulations, they seem more distinct. To circumvent the problem that apps targeting children found outside the "Kids" category were not included, Liu et al. [21] has developed a machine learning model predicting whether an app is designed for children. However, the learning model takes into consideration children under the age of 12 exclusively, since it is centred around the COPPA compliance of apps.

Generic methods for detecting over-claimed permissions by mobile apps have also been explored [18, 34, 36, 39], with some studies shifting their focus on specific app categories [7, 14, 41]. Majethiya and Shah [24] conducted a useful review including many of the aforementioned papers, concluding that semantic analysis is the most efficient. In addition, Brumen et al. [7] finds interesting patterns emerging with the type of data being collected and the permission-specific information analysed per app store by category. However, the study makes no distinction between the different targeted age demographics.

Results show patterns of permission manipulation by developers used to gather as much data as possible from their users, exposing them to unnecessary privacy risks often undisclosed on the app's store listing. For instance, findings by Verderame et al. [38] reveal that more than 95% of Android apps access sensitive information, while only 1% of them comply with the Google Play Privacy guidelines.

We believe an accessible classifier that distinguishes between different target age groups through a semantic analysis of an app's store listing has not yet been made available. As a consequence, most of the related literature either fails to take into account targeted age demographics or uses apps within specific categories to build their dataset.

3 METHODOLOGY

The dataset of applications tested will be retrieved from the Google Play and Apple App stores, being the largest Android and Apple stores globally [12]. Google Play application data will be gathered using the google-play-scraper[2]. We have developed a different scraper to retrieve App Store listing information, using the selenium web-driver in combination with the app-store-scraper[3], allowing the retrieval of extra information such as an app's privacy policy and data safety information.

3.1 Answering RQ1

Due to stricter protections warranted by regulations like the GDPR and COPPA on data surrounding children, the Google Play and Apple App stores only disclose the target audience of apps found within the "Kids" category. These are split into the 5–, 6-8, and 9-12 age bands on the Play store, and 5–, 6-8, and 9-11 on the Apple store. All of the applications found within these categories are labelled as targeting "Children".

Differentiating between Teen and Adult categories is more challenging since the provided app content rating, as a PEGI[4] or ESRB[5] score, is not indicative of the true target audience but simply suggests the minimum age the app is suitable for. In other words, content rated as *PEGI 18* or *Mature* can be classified as targeting Adults. In contrast, content outside the "Kids" category with any other rating, such as *PEGI 12* or *Teen*, necessitates further processing.

To answer the first sub-question, we must develop a classifier capable of making this distinction and accurately predict the targeted age of an app. To this end, the aforementioned scrapers have been used to gather an initial dataset of 500 mobile applications, 47% from the App Store and 53% from Google Play.

Each application is then labelled with the ground truth so that it can be used to train the classifier. The mobile data and analytics platform *Data.ai* provides the empirical data necessary, displaying

[2]GitHub Repository: https://github.com/facundoolano/google-play-scraper
[3]GitHub Repository: https://github.com/facundoolano/app-store-scraper
[4]Pan-European Game Information score part of IARC
[5]Entertainment Software Rating Board score part of IARC

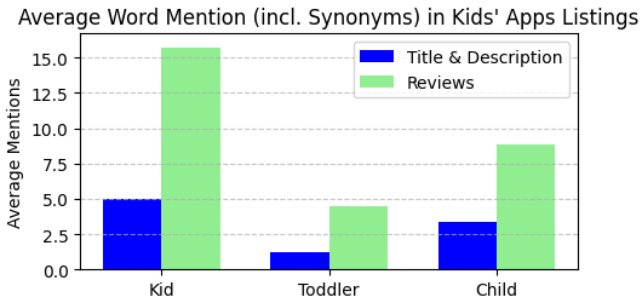

Figure 1: Average word occurrence (incl. synonyms) within kids' apps listings

the gender and age demographics of user accounts downloading a given application [15]. Data from these graphs, such as the one presented in Figure 2, together with the application listing itself, will be used to label the dataset with a simple algorithm.

Firstly, an app is checked to see whether or not it targets children, which is done in one of two possible ways. The first method is to check the app category and whether the store has marked it as targeting specific children's age bands. If this is the case, the app can be labelled as targeting children exclusively, and the rest of the algorithm can be skipped. Alternatively, a list of words synonym with *child*, *kid*, and *toddler* coming from a thesaurus is used. The occurrence of these words and their synonyms within app titles, descriptions, and the front page of reviews[6] was measured on a small set of applications within the "Family" category of the app stores. Using the results shown in Figure 1, it was decided that if these words occur three times within the app's title and description or six times within the app's first review page, the app is marked for Children. Words such as "toddler" and synonyms have been given a higher weight to account for the imbalance.

The second step involves the manual checking of an application's account download demographics coming from *Data.ai*. The platform can help distinguish between the Teenager and Adult age demographics, and a screenshot of the data of interest for a sample app has been depicted in Figure 2. The Adult age category is calculated as an average of the adult age bands considered by the platform. The age categories remaining in the top 10% are picked as the final label. If no data is found for a given app, it is discarded.

Once the data labelling is complete, the ESRB and PEGI ratings of Android applications are mapped with the Apple store content ratings, through an adapted version of the conversion table provided on the Apple Developer website[7]. The proposed mapping can be found in Table 1, with the only difference being two additional mappings: *PEGI 18* → *Mature 17+*, and *PEGI 7* → *Everyone 10+*. These can be justified given that both age bands already fall into the Children or Adults categories, and including them provides a more comprehensive dataset.

For each labelled application, four features have been saved that will be used to tune the model; these features are: *title*, *description*, *categories*, and *content rating*. The classification will be computed by

[6]The first 50 reviews ordered by helpfulness
[7]Apple content rating conversion table: https://developer.apple.com/help/app-store-connect/reference/age-ratings/

Table 1: Content rating conversion table

ESRB	PEGI	Apple
Everyone	PEGI 3	4+
Everyone 10+	PEGI 7	9+
Teen	PEGI 12	12+
Mature 17+	PEGI 16 || PEGI 18	17+

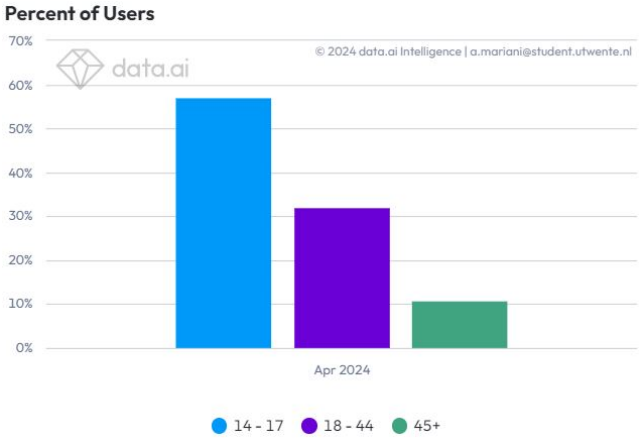

Figure 2: Data.ai user demographics for a sample mobile app

the pre-trained large language model (LLM) GPT-3.5[8] by OpenAI. The model is going to receive additional training and will be fine-tuned to determine app target age groups as accurately as possible. Multiple fine-tuned models will be evaluated, each trained with different subsets of features, varying train-test splits, and adjusted hyperparameters.

3.2 Answering RQ2

The Apple[9] and Google Play[10] stores consider different data types and different data type collection purposes. The two conversion tables used to map both variables can be found in Archive [26]. The data type collection purposes appear to match closely, having only minor discrepancies. Similarly, a suitable match is found for the majority of data types.

The few that remain unmatched, as well as any diagnostics data, are excluded from the analysis. The most notable ones are *Files and docs* and *Calendar events*, which do not map perfectly to any iOS data type. A mapping could be made to iOS existing *Other Data Types* data type, however, because of the generic nature of the category, it will instead be excluded altogether.

A dataset of 1100 Android applications will be scraped from the Google Play store. The corresponding set of applications will be retrieved from the App Store to analyse whether permission requests and data safety practices are impacted by the platform hosting the app. Therefore, a script to match application titles needs to be developed.

[8]Specifically, GPT-3.5-turbo-0125
[9]Apple data types and purposes: developer.apple.com/app-store/app-privacy-details
[10]Play Store data types and purposes: support.google.com/googleplay/android-dev

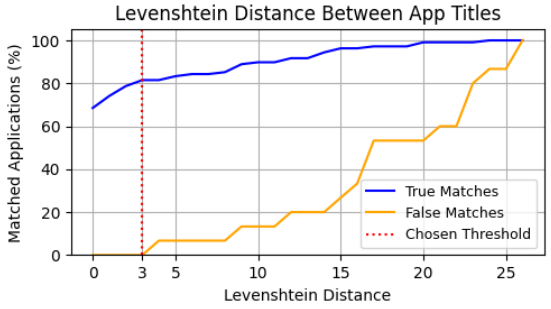

Figure 3: Matched apps with increasing distance thresholds

To understand the most effective way to pair applications, a small set of 100 Google Play Store apps was manually checked for matching listings on the App Store. While most corresponding listings had matching titles, a significant portion differed slightly. Using the Levenshtein Distance formula, we will calculate the similarity between the matched app titles and measure the best distance threshold to match application names. The Levenshtein Distance between two words describes the minimum number of single-character edits necessary to change one into the other, and the formal representation is shown below.

$$\text{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \text{lev}_{a,b}(i-1,j) + 1 \\ \text{lev}_{a,b}(i,j-1) + 1 \\ \text{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

Where $\text{lev}_{a,b}(i,j)$ indicates the Levenshtein Distance between the initial i characters of string a and initial j characters of string b, and $1_{(a_i \neq b_j)}$ denotes 0 when $a = b$ or 1 otherwise.

Apps for which a match was not found are placed in a separate set and paired with their closest match on the App Store as "False Matches". The closest corresponds to the first search result appearing after entering the unpaired application name.

The results are shown in Figure 3; by using a distance threshold of 3, we can detect over 80% of true matches and, more importantly, have a reliable margin to exclude false matches.

3.3 Answering RQ3

The dataset of applications will be analysed further to see if there are significant differences found in the reported use and collection of user personal data. Additionally, an extra 300 applications will be scraped from the App Store since the current iOS apps have only been collected as positive matches to the Play Store ones.

Figure 4 is a visualisation of the privacy policy parsing methodology. Once the privacy policy URL is retrieved, the first step consists of designing a privacy policy extraction script. Trafilatura is one of the fastest and best-performing Python text extraction packages suitable for this task [6]. Developed by Barbaresi [4], Trafilatura retrieves key textual content from the webpage found at a given URL address and filters the rest.

Secondly, the data types an application collects, shares, or tracks must be identified within the privacy policy text. This can be done using the spaCy library for Natural Language Processing (NLP), which can identify and map relevant text segments within the policies to the collection of specific data types. A mapping to different words, phrases, and patterns is created for each data type. For every detection, the library will retrieve the data type and the surrounding paragraph within the text.

The GPT 3.5 LLM will then be used to parse the paragraphs and assess whether the detected data types are being collected or not. This extra step is implemented to improve accuracy, with the idea that a contextual analysis of the paragraph will exclude cases where privacy policies mention data types not being collected.

Finally, we can compare the data types disclosed in the privacy policies to the data types disclosed on the collected application listings. The matched and unmatched data types are stored for each app so that patterns of specific undisclosed permissions can be analysed throughout different age groups and app categories.

4 RESULTS

This section will discuss the results obtained following the methodology outlined in Chapter 3.

4.1 The Classifier

Multiple classifiers have been trained and tested with various configurations to predict an app's targeted age as accurately as possible. A different combination of features, train-test splits, and hyperparameters has been used with each configuration, and the performances are recorded in Table 2. Initial results indicated that a train-test split of 80-20 generally performed worse than a split of 70-30. This is likely due to the relatively small dataset size; therefore, those models have not been included in the results table. The hyperparameters that have been altered are batch size, learning rate (LR) multiplier, and number of epochs. Batch size indicates the number of samples iterated over before the model parameters are updated. The LR Multiplier is used to scale the model's weights and, therefore, dictate how fast the model is updated with each batch. Epochs describe the number of full cycles throughout the training dataset.

To assess the gravity of false negatives and false positives, categories were mapped as follows: *Adults* $\rightarrow$ 0, *Teens* $\rightarrow$ 1, and *Children* $\rightarrow$ 2. If the classifier predicts a label of 2 on a true label of 1, a distance of $|2-1| = 1$ is recorded. Figure 5 shows the confusion matrix for the best-performing configuration, achieving a maximum accuracy of 93.6%. This is a notable improvement over the base model, indicated on the results table as model #0 and peaking at an accuracy of 66.2%. Generally, models trained using the feature set comprised of app titles, descriptions, and categories appear to outperform other models trained using different features using the same or similar configurations. The inclusion of app content rating as a feature had a negative impact on the performance of the classifier, supporting the claim that target age and content rating are not as closely related.

Surprisingly, no distances of 2 were found between the predicted and actual categories. The most problematic classifications appear to be against apps that target both teenagers and adults, where the classifier tends to choose one over the other. The same behaviour is shown with apps targeting both children and teenagers, which can

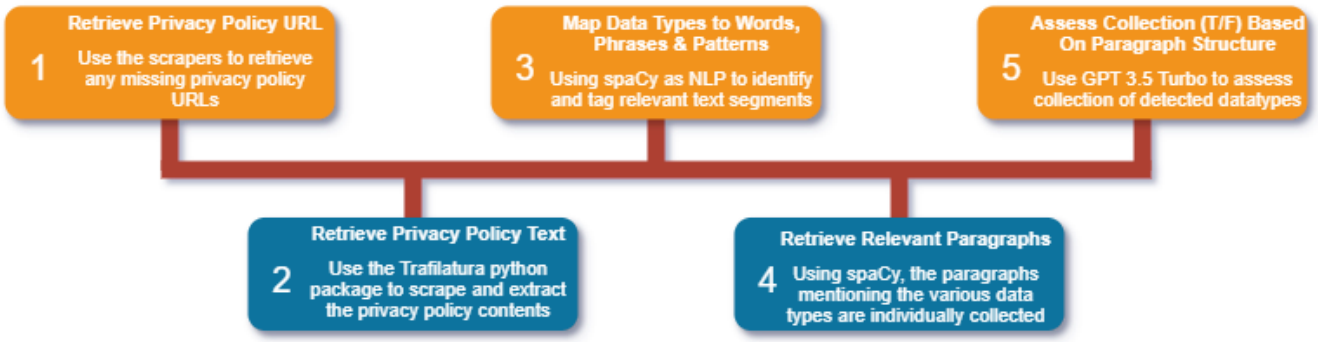

Figure 4: Privacy policy parsing methodology

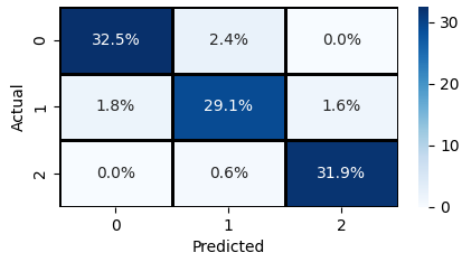

Figure 5: Confusion matrix of the best performing model

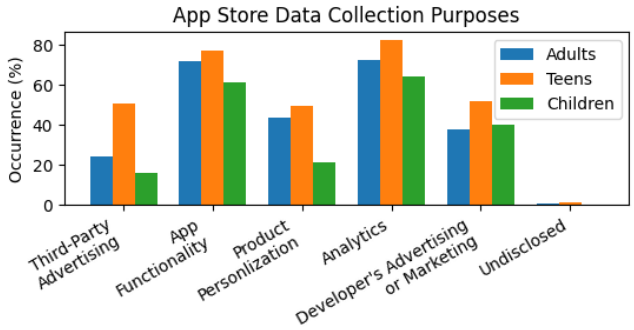

Figure 6: Apple data collection purposes by target age group

be explained by their relatively small occurrence within training and testing sets.

4.2 A cross-platform comparison

Using a Levenshtein Distance threshold of 3 to match application names, a total of 811 apps from the App Store have been paired with the original dataset of 1100. The best-performing classifier was then used to label the 1622 applications from the Apple App and Google Play stores. Table 3 shows the percentage of applications collecting each data type, categorised by platform and app target age group. A chi-squared test of independence determines if there is a significant association between two or more categorical variables. The test has been conducted to assess the relationship between platforms (under "By Platform") and age groups (under "By Age Group"), highlighting significance values below 0.05. A significance level below this threshold indicates less than a 5% risk of concluding that a difference exists between categories when there is no actual difference. With the exception of rarely requested data types and product interaction data, age groups have an average significance value of 0.0003.

Product interaction, device ID, user ID, and advertising data have the highest overall collection rates. The age group experiencing the highest collection of such data is teenagers, often by high margins. This is explicitly visible from the visual representation of this data, which can be found in Archive [27].

4.2.1 By Age Group. As explored in section 2, the related literature suggests that an overwhelming majority of apps over-claim permissions to gather as much data as possible from their users. Given the high value of children's data and the fact that most regulators treat teenagers as adults, resulting in no data collection restrictions, we hypothesised that developers would target teenagers for data collection more than any other age group. The graphs in Figures 6 and 7 show the distribution of data collection purposes, where teenagers appear to lead in every category, supporting this hypothesis. The most significant difference in data collection purposes is data collected explicitly for third-party advertising. This is the case for roughly 53% of apps targeting teens, compared to the 29% of apps targeting adults.

The distribution of data collection purposes is not subject to significant change across platforms, meaning that generally, the same application will collect data for the same purposes on the Google Play and App stores alike. However, data collected from apps targeting teenagers is the most varied, indicating that the same data is used for a larger range of purposes.

The data types most commonly collected from apps targeting teenagers are Device ID, Purchase History, and Advertising Data. To determine whether the difference between the adult and teenager categories is significant, an additional chi-square test of independence has been performed, isolating the two categories. The results, shown in Table 4, reveal that the differences all have a significance

Table 2: Fine-tuned models parameters and results

#	Feature Set	Batch Size	LR Multiplier	Epochs	Precision	Recall	F1 Score	Accuracy
0	[title, description, rating, categories]	-	-	-	0.435	0.998	0.606	0.480
	[title, description, categories]				0.537	0.983	0.695	**0.662**
	[title, description]				0.529	0.975	0.686	0.643
1	[title, description, rating, categories]	1	2	3	0.888	0.883	0.886	**0.909**
2	[title, description, contentRating]	1	2	3	0.890	0.915	0.902	**0.922**
3	[title, description]	2	1.5	5	0.858	0.926	0.891	0.911
		4	1	10	0.866	0.915	0.890	0.911
		2	2	5	0.876	0.920	0.898	0.918
		4	1.5	3	0.870	0.949	0.908	0.924
		1	2	3	0.882	0.932	0.906	0.924
		4	1	3	0.879	0.949	0.913	**0.929**
		4	0.5	3	0.883	0.943	0.912	**0.929**
4	[title, description, categories]	4	0.5	3	0.911	0.850	0.879	0.907
		1	2	4	0.921	0.911	0.916	0.933
		1	2	3	0.927	0.911	**0.919**	**<u>0.936</u>**

Table 3: Data type collection by app platform and target age

	App Store			Play Store			By Age Group		By Platform	
Data Type	**Adults**	**Teens**	**Children**	**Adults**	**Teens**	**Children**	**Chi-S**	**Sig.**	**Chi-S**	**Sig.**
Browsing History	1.83%	2.12%	1.69%	0.61%	1.33%	0.00%	1.2354	0.5392	3.5217	0.0606
Email Address	39.02%	28.38%	24.58%	54.57%	43.77%	25.42%	23.8681	**<0.0001**	18.9655	**<0.0001**
Name	28.66%	15.12%	5.08%	42.99%	32.10%	14.41%	48.9176	**<0.0001**	34.1376	**<0.0001**
Other User Contact Info	3.05%	2.92%	0.00%	23.48%	16.45%	7.63%	15.5562	**0.0004**	95.4379	**<0.0001**
Phone Number	12.50%	3.18%	1.69%	22.87%	7.69%	4.24%	66.3323	**<0.0001**	17.7805	**<0.0001**
Physical Address	4.57%	1.33%	0.00%	10.37%	1.59%	0.00%	44.8101	**<0.0001**	6.6667	**0.0098**
Contacts	3.05%	7.16%	0.85%	4.88%	7.43%	0.85%	17.3341	**0.0002**	0.5904	0.4423
Credit Info	0.30%	0.27%	0.00%	0.61%	0.00%	0.00%	2.1913	0.3343	0.0000	1.0000
Other Financial Info	3.66%	0.53%	0.00%	10.37%	0.80%	0.00%	54.1751	**<0.0001**	10.3725	**0.0013**
Payment Info	3.05%	1.06%	0.00%	7.32%	3.45%	0.85%	16.0470	**0.0003**	11.0769	**0.0009**
Fitness	0.30%	0.80%	0.00%	1.83%	1.06%	0.00%	2.4226	0.2978	2.5714	0.1088
Health	0.00%	0.80%	0.00%	1.52%	0.27%	0.00%	1.8508	0.3964	1.0000	0.3173
Device ID	49.09%	67.64%	25.42%	64.02%	74.80%	51.69%	34.6680	**<0.0001**	11.4605	**0.0007**
User ID	43.60%	56.50%	29.66%	46.65%	48.81%	26.27%	23.9866	**<0.0001**	0.6970	0.4038
Coarse Location	27.13%	35.54%	20.34%	32.93%	42.71%	34.75%	11.8020	**0.0027**	7.1257	**0.0076**
Precise Location	15.55%	5.04%	1.69%	19.51%	5.31%	0.85%	75.9083	**<0.0001**	1.0764	0.2995
Purchase History	23.48%	40.85%	22.03%	28.96%	45.62%	17.80%	45.1660	**<0.0001**	1.7633	0.1842
Search History	6.71%	7.69%	2.54%	12.80%	14.06%	3.39%	12.3456	**0.0021**	13.2353	**0.0003**
Sensitive Info	12.50%	1.59%	0.00%	29.88%	16.98%	7.63%	56.0412	**<0.0001**	70.5321	**<0.0001**
Undisclosed	0.61%	1.06%	1.69%	0.00%	0.00%	0.00%	1.1079	0.5747	8.0000	**0.0047**
Advertising Data	28.66%	44.30%	15.25%	54.27%	56.50%	45.76%	17.7424	**0.0001**	38.0608	**<0.0001**
Other Usage Data	17.38%	26.53%	5.93%	20.43%	37.67%	24.58%	34.4833	**<0.0001**	13.6219	**0.0002**
Product Interaction	61.28%	70.03%	56.78%	54.27%	56.50%	45.76%	4.8140	0.0901	7.7472	**0.0054**
Audio Data	3.05%	4.24%	2.54%	8.54%	5.84%	3.39%	2.7500	0.2528	7.5301	**0.0061**
Emails or Text Messages	5.79%	4.77%	0.00%	25.30%	21.22%	4.24%	25.4648	**<0.0001**	83.7122	**<0.0001**
Gameplay Content	3.35%	18.04%	8.47%	20.43%	37.67%	24.58%	46.5331	**<0.0001**	67.8930	**<0.0001**
Other User Content	14.33%	11.67%	0.85%	22.87%	22.81%	7.63%	24.0179	**<0.0001**	23.2214	**<0.0001**
Photos or Videos	25.00%	15.65%	3.39%	34.45%	21.22%	4.24%	59.7797	**<0.0001**	8.1895	**0.0042**

Figure 7: Play Store collection purposes by target age group

Table 4: Adults vs Teens data type significance

Data Type	Adults	Teens	Chi-S	Sig.
Device ID	56.56%	71.22%	11.7158	**0.0006**
Purchase History	26.22%	43.24%	28.7600	**<0.0001**
Advertising Data	41.47%	50.4%	6.0557	**0.0139**

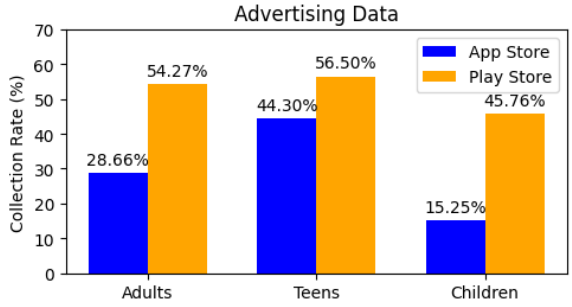

Figure 8: Advertising Data collection rates

of less than 0.05, hinting that applications treat the two age categories differently.

4.2.2 By Platform. With the exception of undisclosed data types or collection purposes, which have not been recorded on Google Play apps, the mobile platform does not impact data collection purposes. Instead, it significantly influences the type and amount of data collected.

Advertising Data is one of the most profitable data types, being of direct interest to third parties, and is subject to one of the biggest cross-platform imbalances. Figure 8 shows its collection per platform by age group. From this figure, it appears that applications on the App Store have a harder time retrieving and selling children's information to third parties. On the other hand, the same applications downloaded from the Play Store attempt to collect as much advertising data as possible between all three age groups.

Gameplay Content is the only data type collected more frequently within children's apps than those targeting adults on both stores. Interestingly, apps downloaded on the Play Store also collect Coarse Location data more often on children's applications. Figure 9 shows how often these apps collect location data, with teenage-targeted apps having the highest count.

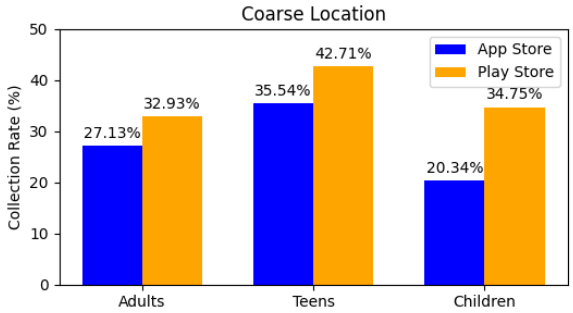

Figure 9: Coarse Location collection rates

On average, apps on the Google Play store collect 2.1 times more data than the same apps listed on the App Store. Given the higher amounts of data collected, it seems that developers have more freedom when publishing their applications on Google Play. Overall, it is evident from Table 3 and the graphs in Archive [27] that apps targeting children collect less information than the rest. This is likely the effect of current regulations adding a small layer of protection for children, which this research aims to extend to protect teenagers as well.

4.3 Privacy Policies vs. App Listings

Following the methodology outlined in Chapter 3.3, we were able to parse and analyse the privacy policies of applications within the dataset. As a result of complexities that arose during the mapping of words and phrases to specific data types, "Advertising Data" takes a slightly different meaning, indicating applications that collect any data for advertisement purposes. Additionally, an error occurred during the processing of email address collection, leading to its exclusion from the analysis. The error was due to the extractor mistakenly counting company email contact information as user-collected data.

The results of this analysis are shown in Table 5, alongside results obtained from the analysis of application listings for comparison. The overall average collection rate has increased by roughly 10.1% for apps targeting the teenager and children age groups and 5% for those targeting adults. This implies that there is more consistency between the privacy policy and app listing information of apps targeting adults. The processing of contact information such as physical address, phone number, and name experiences the highest overall difference, reaching 48% (specifically for apps targeting children). Interestingly, data is collected for advertising purposes in 100% of the analysed privacy policies of applications targeting teenagers. A visual overview of the discrepancies between the information disclosed on privacy policies and store listings can be found in Archive [25]. Here, we can see that financial information such as payment and credit score data appear significantly more often on privacy policies as well.

With the exception of Product Interaction and Coarse Location, all of the highlighted significant differences recorded[11] showed higher data collection rates in Privacy Policies. A closer look at the results of the LLM GPT 3.5 indicated the model had problems

[11] Using a threshold of 20%

distinguishing between the declaration of Precise and Coarse Location data collection. This inaccuracy could explain the observed difference, which disappears when the average count of both data types is considered.

5 DISCUSSION

The classifier developed to answer RQ1 and determine an app's targeted age group achieved a maximum accuracy of 93.6%, a significant improvement over the base model's peak of 66.2%. This higher accuracy shows the effectiveness of the semantic analysis approach used and indicates that analysing app listings is a viable method for determining targeted age groups.

Answering RQ2 revealed notable differences when inspecting data collection between the same applications listed on the Apple App and Google Play stores. This indicates a disparity in how apps comply with data safety regulations across platforms. On average, the same application collects more than twice the user data when downloaded from Google Play. We hypothesise this could be attributed to the App Store having stricter data safety guidelines, highlighting the need for uniform guidelines across platforms.

Significant differences have also been found in analysing data type collection across age groups, as almost every data type demonstrated higher collection rates in applications targeting teenagers. Three data types stood out: device ID, purchase history, and advertising data. The device ID is a valuable data type, as it represents a unique identifier that recognises every individual device and can be used to track user habits across platforms and apps. Purchase history can then be linked to the device ID to understand a user's buying patterns and preferences and suggest more tailored advertisements. Finally, advertising data can provide any metric evaluating the interactions and engagement the user had with the shown advertisements. The three data types have shown the highest collection rates among teenagers, implying that developers take advantage of the lack of restrictions guarding teenagers' data and target such users more than adults, even though they have the same amount of control over their privacy.

Answering RQ3 has also revealed discrepancies between data collection practices disclosed in privacy policies and those found on store listings. We expected the differences to be uniform across age groups; however, apps targeting teenagers and children are impacted the most. Unlike children, who need their parents to give consent to their personal data, teenagers are expected to grant or deny it themselves. The fact that developers disclose more information within privacy policies than they do on mobile store listings indicates a lack of transparency, which is even more concerning when directed to younger demographics that are less aware of the implications.

Related literature like Brumen et al. [7] does not find significant differences in permissions requests across the Google Play and Apple App stores. However, their results comparing privacy policies to app listings are very similar, recording even higher differences where privacy policies disclose an average of roughly 30% extra data types collected. The discrepancy in the cross-platform analysis is likely due to the different methodologies, as the research considers apps specifically picked from five store categories.

5.1 Limitations

This study has had a number of limitations that will be addressed in this section. Firstly, the classifier was trained on a relatively small dataset with only a few edge cases, which could impact its generalizability. Secondly, the methodology used to determine app age groups encountered problems labelling apps targeting entire families. For instance, mobile board games include various comments from parents mentioning having fun with their kids, where the script then classifies the app as targeting children. Answering RQ3, the use of Trafilatura to retrieve privacy policy text was not as effective as expected, as there were issues scraping websites requiring Javascript to be enabled. Despite the slower execution of roughly 30 seconds per policy, the PoliPy package developed by Samarin et al. [32] might have been a more suitable alternative. Additionally, the accuracy of the GPT LLM in recognising data types collected within the context of a given paragraph needs to be formally assessed through an annotated set of policies.

5.2 Ethical Considerations

The scrapers developed throughout this research were designed to run with the smallest impact possible on the hosts' servers. For instance, long timeouts have been implemented between request cycles to ensure that no more than 100 applications were scraped within an hour. Additionally, all of the data collected is publicly available on the developers' and hosting platforms' websites.

6 CONCLUSION

This study aimed to investigate the impact that a user's age has on the data collection practices adopted by mobile applications and their consistency with those disclosed in privacy policies. To this end, we tuned a classifier that achieved high accuracy in determining the targeted age group of mobile apps. We then documented the most collected data types across age groups and platforms and calculated their statistical significance through multiple chi-squared tests of independence. Lastly, we have proposed a novel methodology to parse privacy policies and measured notable inconsistencies between their disclosed data collection practices and those observed in their listings, especially for apps targeting teenagers and children.

These findings highlight the need for stricter enforcement of data privacy regulations and the importance of making extra distinctions between age demographics. The significant increase in data collection by apps targeting teenagers should serve to inform future revisions of any child data protection regulation, possibly changing age and permission requirements.

To address some of the limitations encountered in this study, future work can focus on labelling a larger and more comprehensive training set, possibly including listings from more mobile app stores. Additionally, the performance of the policy parsing algorithm needs to be assessed, and tools like *PI-Extract* developed by Bui et al. [9] or *Polisis* by Harkous et al. [17] can be used as an alternative in future research.

During the preparation of this work, the author used ChatGPT by OpenAI to develop and tune a classifier. Additionally, Grammarly was used to make grammatical corrections. After using these tools,

Table 5: Data collected as disclosed in Privacy Policies vs. App Listings

	Privacy Policy			App Listings			Difference *(PP - Listing)*		
Data Type	**Adults**	**Teens**	**Children**	**Adults**	**Teens**	**Children**	**Adults**	**Teens**	**Children**
Browsing History	8%	8%	4%	1%	2%	1%	7%	6%	3%
Name	58%	56%	58%	36%	24%	10%	**22%**	**32%**	**48%**
Other User Contact Info	18%	16%	1%	13%	10%	4%	5%	6%	-3%
Phone Number	42%	41%	30%	18%	5%	3%	**24%**	**36%**	**27%**
Physical Address	27%	18%	16%	7%	1%	0%	**20%**	17%	16%
Contacts	11%	12%	5%	4%	7%	1%	7%	5%	4%
Credit Info	19%	17%	15%	0%	0%	0%	19%	17%	15%
Other Financial Info	8%	10%	2%	7%	1%	0%	1%	9%	2%
Payment Info	37%	44%	39%	5%	2%	0%	**32%**	**42%**	**39%**
Fitness	2%	2%	1%	1%	1%	0%	1%	1%	1%
Health	5%	3%	2%	1%	1%	0%	4%	2%	2%
Device ID	39%	67%	50%	57%	71%	39%	-18%	-4%	11%
User ID	31%	58%	17%	45%	53%	28%	-14%	5%	-11%
Coarse Location	11%	18%	9%	30%	39%	28%	-19%	**-21%**	-19%
Precise Location	20%	17%	10%	18%	5%	1%	2%	12%	9%
Purchase History	33%	48%	36%	26%	43%	20%	7%	5%	16%
Search History	5%	5%	4%	10%	11%	3%	-5%	-6%	1%
Advertising Data	83%	100%	72%	41%	50%	31%	**42%**	**50%**	**41%**
Other Usage Data	20%	22%	33%	19%	32%	15%	1%	-10%	18%
Product Interaction	30%	40%	30%	58%	63%	51%	**-28%**	**-23%**	**-21%**
Audio Data	6%	6%	7%	6%	5%	3%	0%	1%	4%
Emails or Text Messages	52%	57%	82%	16%	13%	2%	**36%**	**44%**	**80%**
Gameplay Content	3%	15%	14%	12%	28%	17%	-9%	-13%	-3%
Other User Content	15%	19%	2%	19%	17%	4%	-4%	2%	-2%
Photos or Videos	22%	19%	17%	30%	18%	4%	-8%	1%	13%

the author reviewed and edited the content as needed and takes full responsibility for the content of the work.

REFERENCES

[1] Iman M. Almomani and Aala Al Khayer. 2020. A Comprehensive Analysis of the Android Permissions System. *IEEE Access* 8 (2020), 216671–216688. https://doi.org/10.1109/ACCESS.2020.3041432 Conference Name: IEEE Access.

[2] Apple. 2024. Age ratings - Reference - App Store Connect. https://developer.apple.com/help/app-store-connect/reference/age-ratings/

[3] Hamid Bagheri, Eunsuk Kang, Sam Malek, and Daniel Jackson. 2018. A formal approach for detection of security flaws in the android permission system. *Formal Aspects of Computing* 30, 5 (Sept. 2018), 525–544. https://doi.org/10.1007/s00165-017-0445-z

[4] Adrien Barbaresi. 2021. Trafilatura: A Web Scraping Library and Command-Line Tool for Text Discovery and Extraction. In *Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, 122–131. https://aclanthology.org/2021.acl-demo.15

[5] Anna Barzolevskaia, Enrico Branca, and Natalia Stakhanova. 2024. Measuring and Characterizing (Mis)compliance of the Android Permission System. *IEEE Transactions on Software Engineering* 50, 4 (April 2024), 742–764. https://doi.org/10.1109/TSE.2024.3362921

[6] Muhammad Ramish Bhutto. 2022. Automating Privacy Policy Extraction And Summarization. https://essay.utwente.nl/91789/ Publisher: University of Twente.

[7] Boštjan Brumen, Aljaž Zajc, and Leon Bošnjak. 2023. Permissions vs. Privacy Policies of Apps in Google Play Store and Apple App Store. In *Information Modelling and Knowledge Bases XXXIV*. IOS Press, 258–275. https://doi.org/10.3233/FAIA220507

[8] Richard Bryan. 1998. Children's Online Privacy Protection Act. https://www.ecfr.gov/current/title-16/part-312

[9] Duc Bui, Kang G. Shin, Jong-Min Choi, and Junbum Shin. 2021. Automated Extraction and Presentation of Data Practices in Privacy Policies. *Proceedings on Privacy Enhancing Technologies* 2021, 2 (April 2021), 88–110. https://doi.org/10.2478/popets-2021-0019

[10] Grazia Cecere, Fabrice Le Guel, Vincent Lefrere, Catherine E. Tucker, and Pai-Ling Yin. 2022. Privacy, Data and Competition: The Case of Apps for Young Children. https://doi.org/10.2139/ssrn.4073931

[11] Grazia Cecere, Vincent Lefrere, Fabrice Le Guel, Catherine E. Tucker, and Pai-Ling Yin. 2023. Privacy, Data and Non-regulatory Privacy Protection: The Case of Apps For Young Children. (April 2023).

[12] Laura Ceci. 2022. Annual mobile app downloads worldwide by store 2026. https://www.statista.com/statistics/1010716/apple-app-store-google-play-app-downloads-forecast/

[13] Laura Ceci. 2024. Annual number of mobile app downloads worldwide 2023. https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/

[14] Efstratios Chatzoglou, Georgios Kambourakis, and Vasileios Kouliaridis. 2021. A Multi-Tier Security Analysis of Official Car Management Apps for Android. *Future Internet* 13, 3 (March 2021), 58. https://doi.org/10.3390/fi13030058 Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[15] Data.ai. 2024. Mobile Market Data and App Insights | data.ai (fka App Annie). https://www.data.ai/en/

[16] Google. 2024. Manage target audience and app content settings. https://support.google.com/googleplay/android-developer/answer/9867159

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

[18] Fu-Hau Hsu, Nien-Chi Liu, Yan-Ling Hwang, Che-Hao Liu, Chuan-Sheng Wang, and Chang-Yi Chen. 2021. DPC:A Dynamic Permission Control Mechanism for Android Third-Party Libraries. *IEEE Transactions on Dependable and Secure Computing* 18, 4 (July 2021), 1751–1761. https://doi.org/10.1109/TDSC.2019.2937925 Conference Name: IEEE Transactions on Dependable and Secure Computing.

[19] Lindsay Jibb, Elsie Amoako, Melissa Heisey, Lily Ren, and Quinn Grundy. 2022. Data handling practices and commercial features of apps related to children: a

scoping review of content analyses. *Archives of Disease in Childhood* 107, 7 (July 2022), 665–673. https://doi.org/10.1136/archdischild-2021-323292 Publisher: BMJ Publishing Group Ltd Section: Original research.

[20] Konrad Kollnig, Anastasia Shuba, Reuben Binns, Max Van Kleek, and Nigel Shadbolt. 2021. Are iPhones Really Better for Privacy? Comparative Study of iOS and Android Apps. https://doi.org/10.2478/popets-2022-0033

[21] Minxing Liu, Haoyu Wang, Yao Guo, and Jason Hong. 2016. Identifying and Analyzing the Privacy of Apps for Kids. In *Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications*. ACM, St. Augustine Florida USA, 105–110. https://doi.org/10.1145/2873587.2873597

[22] Qian Luo, Jiajia Liu, Jiadai Wang, Yawen Tan, Yurui Cao, and Nei Kato. 2020. Automatic Content Inspection and Forensics for Children Android Apps. *IEEE Internet of Things Journal* 7, 8 (Aug. 2020), 7123–7134. https://doi.org/10.1109/JIOT.2020.2982248 Conference Name: IEEE Internet of Things Journal.

[23] Qian Luo, Yinbo Yu, Jiajia Liu, and Abderrahim Benslimane. 2022. Automatic Detection for Privacy Violations in Android Applications. *IEEE Internet of Things Journal* 9, 8 (April 2022), 6159–6172. https://doi.org/10.1109/JIOT.2021.3109785 Conference Name: IEEE Internet of Things Journal.

[24] Raj J Majethiya and Monika Shah. 2023. Comparative analysis of detecting over-claim permissions from android apps. In *2023 International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC)*. 1–8. https://doi.org/10.1109/ISACC56298.2023.10084321

[25] Adamo Mariani. 2024. App Listing vs. Privacy Policy Disclosed Data Types. https://archive.org/details/pp_vs_listing_by_age

[26] Adamo Mariani. 2024. Cross-Platform Conversion Tables. https://archive.org/details/datatype_conversion_table

[27] Adamo Mariani. 2024. Cross-Platform Permissions By Age Group. https://archive.org/details/cross_platform_permissions

[28] Marisa Meyer, Victoria Adkins, Nalingna Yuan, Heidi M. Weeks, Yung-Ju Chang, and Jenny Radesky. 2019. Advertising in Young Children's Apps: A Content Analysis. *Journal of Developmental & Behavioral Pediatrics* 40, 1 (Jan. 2019), 32. https://doi.org/10.1097/DBP.0000000000000622

[29] Ibtisam Mohamed and Dhiren Patel. 2015. Android vs iOS Security: A Comparative Study. In *2015 12th International Conference on Information Technology - New Generations*. 725–730. https://doi.org/10.1109/ITNG.2015.123

[30] Pixalate. 2020. 90% of mobile apps are for kids aged 12 and under. https://www.pixalate.com/blog/google-apple-mobile-apps-for-kids

[31] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, and Serge Egelman. 2018. "Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale. https://dspace.networks.imdea.org/handle/20.500.12761/551 Accepted: 2021-07-13T09:33:32Z.

[32] Nikita Samarin, Shayna Kothari, Zaina Siyed, Primal Wijesekera, Jordan Fischer, Chris Hoofnagle, and Serge Egelman. 2021. Investigating the Compliance of Android App Developers with the CCPA. (2021).

[33] Amirhosein Sayyadabdi, Behrouz Tork Ladani, and Bahman Zamani. 2022. Towards a Formal Approach for Detection of Vulnerabilities in the Android Permissions System. *The ISC International Journal of Information Security* 14, 3 (Oct. 2022). https://doi.org/10.22042/isecure.2022.14.3.7

[34] Monika Shah. 2022. Detecting over-claim permissions and recognising dangerous permission in Android apps. *International Journal of Information and Computer Security* (Feb. 2022). https://www.inderscienceonline.com/doi/10.1504/IJICS.2022.121298 Publisher: Inderscience Publishers (IEL).

[35] Anne Stopper and Jen Caltrider. 2023. False and Misleading Loopholes in Google's Data Safety Labels. https://foundation.mozilla.org/en/campaigns/googles-data-safety-labels/

[36] Ming-Yang Su, Sheng-Sheng Chen, Tsung-Ren Wu, Hao-Sen Chang, and You-Liang Liu. 2019. Permission Abusing by Ad Libraries of Smartphone Apps. In *2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN)*. 475–477. https://doi.org/10.1109/ICUFN.2019.8806157 ISSN: 2165-8536.

[37] Ruoxi Sun, Minhui Xue, Gareth Tyson, Shuo Wang, Seyit Camtepe, and Surya Nepal. 2023. Not Seen, Not Heard in the Digital World! Measuring Privacy Practices in Children's Apps. In *Proceedings of the ACM Web Conference 2023 (WWW '23)*. Association for Computing Machinery, New York, NY, USA, 2166–2177. https://doi.org/10.1145/3543507.3583327

[38] Luca Verderame, Davide Caputo, Andrea Romdhana, and Alessio Merlo. 2020. On the (Un)Reliability of Privacy Policies in Android Apps. In *2020 International Joint Conference on Neural Networks (IJCNN)*. 1–9. https://doi.org/10.1109/IJCNN48605.2020.9206660 ISSN: 2161-4407.

[39] Zhiqiang Wu, Hakjin Lee, and Scott Uk-Jin Lee. 2021. An Empirical Study on the Impact of Permission Smell in Android Applications. *Journal of the Korea Society of Computer and Information* 26, 6 (2021), 89–96. https://doi.org/10.9708/jksci.2021.26.06.089 Publisher: Korean Society of Computer Information.

[40] Yanjie Zhao, Tianming Liu, Haoyu Wang, Yepang Liu, John Grundy, and Li Li. 2023. Are Mobile Advertisements in Compliance with App's Age Group?. In *Proceedings of the ACM Web Conference 2023*. ACM, Austin TX USA, 3132–3141. https://doi.org/10.1145/3543507.3583534

[41] Andrei Șandor and Gabriela Tonț. 2021. Android social applications permission overview from a privacy perspective. In *2021 16th International Conference on Engineering of Modern Electric Systems (EMES)*. 1–4. https://doi.org/10.1109/EMES52337.2021.9484128