Is It a Trap? A Large-scale Empirical Study And Comprehensive Assessment of Online Automated Privacy Policy Generators for Mobile Apps

Shidong Pan*
CSIRO's Data61 & ANU

Dawen Zhang
CSIRO's Data61 & ANU

Mark Staples
CSIRO's Data61

Zhenchang Xing
CSIRO's Data61 & ANU

Jieshan Chen
CSIRO's Data61

Xiwei Xu
CSIRO's Data61

Thong Hoang†
CSIRO's Data61

Abstract

Privacy regulations protect and promote the privacy of individuals by requiring mobile apps to provide a privacy policy that explains what personal information is collected and how these apps process this information. However, developers often do not have sufficient legal knowledge to create such privacy policies. Online Automated Privacy Policy Generators (APPGs) can create privacy policies, but their quality and other characteristics can vary. In this paper, we conduct the first large-scale empirical study and comprehensive assessment of APPGs for mobile apps. Specifically, we scrutinize 10 APPGs on multiple dimensions. We further perform the market penetration analysis by collecting 46,472 Android app privacy policies from Google Play, discovering that nearly 20.1% of privacy policies could be generated by existing APPGs. Lastly, we point out that generated policies in our study do not fully comply with GDPR, CCPA, or LGPD. In summary, app developers must carefully select and use the appropriate APPGs with careful consideration to avoid potential pitfalls.

1 Introduction

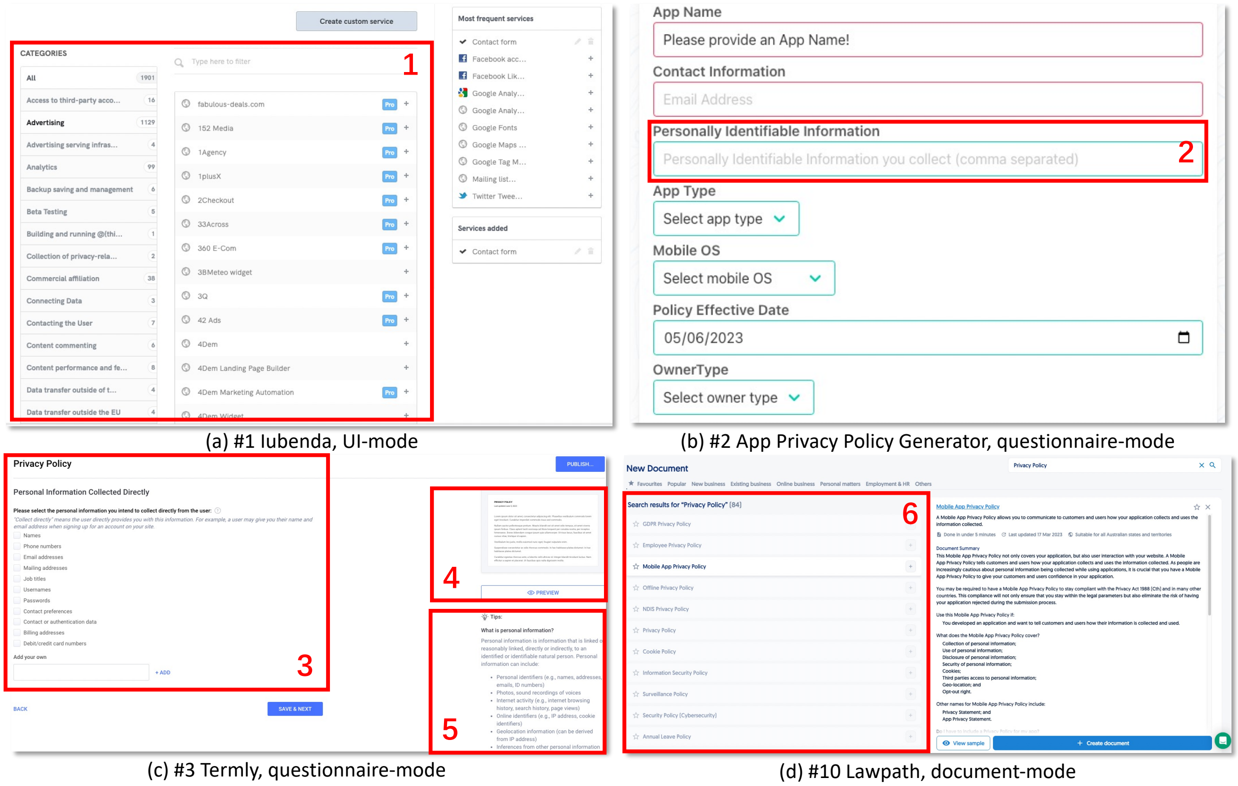
Mobile phones and apps are now a ubiquitous part of digital life. There is a large variety and volume of data collected and used by mobile apps, which inevitably brings many privacy issues [47, 49, 51, 66, 71]. Privacy policies inform users about what, why, and how their personal data are collected and used. These privacy policies have become an important element of responsible technology in mobile app ecosystems. They also form part of legal agreements for apps and services. Specifically, they are required under regulation in many jurisdictions, such as European General Data Protection Regulation (GDPR) [13], California Consumer Privacy Act (CCPA) [10], Brazilian Law of General Data Protection (LGPD) [9], Australian Privacy Principles (APP) [7], and Chinese Personal Information Protection Law (PIPL) [21]. For example, APP [Art. 1, §1.2] imposes an obligation on organisational entities to *"have a clearly expressed and up-to-date APP Privacy Policy about how the entity manages personal information."* Privacy policies must be consistent with the data collected and functions provided by service providers, and inadequate policies can create significant business and legal problems.

Developing privacy policies is a complex process, demanding both the knowledge of app features and corresponding legal requirements. While larger companies have greater resources and legal expertise to create high-quality privacy policies for their apps, most (citizen) developers do not have legal support and struggle to prepare accurate privacy policies [36, 37, 49, 58]. To develop privacy policies, developers may copy-paste-modify existing privacy policies, ad-hoc. As part of this big picture, app development by non-professional developers is growing quickly, supported by trends of using no-code/low-code automated app development tools [53] and pre-trained Large Language Models (LLMs). To address these needs, Online Automated Privacy Policy Generators (APPGs) can provide more automated solutions and more systematic support for developers to create privacy policies for their apps, rather than through ad-hoc reuse. Figure 1 provides several illustrative examples. Most APPGs are questionnaire-based tools, which work by asking app developers a series of privacy-related questions about the app, and using those answers to generate privacy policies. APPGs can create privacy policies, but their quality and other characteristics can vary and are not yet deeply understood.

As we observe in this paper, many apps fail to provide a privacy policy, perfunctorily provide only a low-quality privacy policy, or provide a privacy policy not in local language. APPGs are becoming an increasingly popular solution used by developers, but developers can be unaware of hidden issues in APPGs [61, 70]. Potential design flaws may reflect in the policies, amplify vulnerabilities, violate users' trust assumptions, and ultimately harm end-users. Thus, a comprehensive and systematic assessment of the capability and limitations of APPGs is necessary. To better understand the scale of po-

*Shidong.Pan@anu.edu.au, School of Computing, Australian National University & Software Systems Research Group, CSIRO's Data61, Australia.

†Corresponding author. James.Hoang@data61.csiro.au, Software Systems Research Group, CSIRO's Data61, Australia.



CATEGORIES

All	1901
Access to third-party acco...	16
Advertising	1129
Advertising serving infras...	4
Analytics	99
Backup saving and management	6
Beta Testing	5
Building and running @(thi...	1
Collection of privacy-rela...	2
Commercial affiliation	38
Connecting Data	3
Contacting the User	7
Content commenting	6
Content performance and fe...	8
Data transfer outside of t...	4
Data transfer outside the EU	4

Type here to filter

fabulous-deals.com Pro +
152 Media Pro +
1Agency Pro +
1plusX Pro +
2Checkout Pro +
33Across Pro +
360 E-Com Pro +
3BMeteo widget +
3Q Pro +
42 Ads Pro +
4Dem +
4Dem Landing Page Builder +
4Dem Marketing Automation Pro +
4Dem Widget +

Most frequent services

✓ Contact form
Facebook acc... +
Facebook Lik... +
Google Analy... +
Google Analy... +
Google Fonts +
Google Maps ... +
Google Tag M... +
Mailing list... +
Twitter Twee... +

Services added

✓ Contact form

(a) #1 Iubenda, UI-mode

App Name
Please provide an App Name!
Contact Information
Email Address
Personally Identifiable Information
Personally Identifiable Information you collect (comma separated)
App Type
Select app type
Mobile OS
Select mobile OS
Policy Effective Date
05/06/2023
OwnerType
Select owner type

(b) #2 App Privacy Policy Generator, questionnaire-mode

Privacy Policy

Personal Information Collected Directly

Please select the personal information you intend to collect directly from the user:
"Collect directly" means the user directly provides you with this information. For example, a user may give you their name and email address when signing up for an account on your site.

□ Names
□ Phone numbers
□ Email addresses
□ Mailing addresses
□ Job titles
□ Usernames
□ Passwords
□ Contact preferences
□ Contact or authentication data
□ Billing addresses
□ Debit/credit card numbers

Add your own

PRIVACY POLICY

Tips:

What is personal information?

Personal information is information that is linked o reasonably linked, directly or indirectly, to an identified or identifiable natural person. Personal information can include:

- Personal identifiers (e.g., names, addresses, emails, ID numbers)
- Photos, sound recordings of voices
- Internet activity (e.g., internet browsing history, search history, page views)
- Online identifiers (e.g., IP address, cookie identifiers)
- Geolocation information (can be derived from IP address)
- Inferences from other personal information

(c) #3 Termly, questionnaire-mode

New Document

Privacy Policy

Favourites Popular New business Existing business Online business Personal matters Employment & HR Others

Search results for "Privacy Policy" (84)

☆ GDPR Privacy Policy
☆ Employee Privacy Policy
☆ Mobile App Privacy Policy
☆ Offline Privacy Policy
☆ NDIS Privacy Policy
☆ Privacy Policy
☆ Cookie Policy
☆ Information Security Policy
☆ Surveillance Policy
☆ Security Policy (Cybersecurity)
☆ Annual Leave Policy

Mobile App Privacy Policy

A Mobile App Privacy Policy allows you to communicate to customers and users how your application collects and uses the information collected.

Done in under 5 minutes Last updated 17 Mar 2023 Suitable for all Australian states and territories

Document Summary
This Mobile App Privacy Policy not only covers your application, but also user interaction with your website. A Mobile App Privacy Policy tells customers and users how your application collects and uses the information collected. As people are increasingly cautious about personal information being collected while using applications, it is crucial that you have a Mobile App Privacy Policy to give your customers and users confidence in your application.

You may be required to have a Mobile App Privacy Policy to stay compliant with the Privacy Act 1988 (Cth) and in many other countries. This compliance will not only ensure that you stay within the legal parameters but also eliminate the risk of having your application rejected during the submission process.

Use this Mobile App Privacy Policy if:
You developed an application and want to tell customers and users how their information is collected and used.

What does the Mobile App Privacy Policy cover?
Collection of personal information;
Use of personal information;
Disclosure of personal information;
Security of personal information;
Cookies;
Third parties access to personal information;
Geo-location; and
Opt-out right.

Other names for Mobile App Privacy Policy include:
Privacy Statement; and
App Privacy Statement.

Do I have to include a Privacy Policy for my app?

(d) #10 Lawpath, document-mode

Figure 1: Examples (#1, #2, #3, and # 10) of APPGs. (1) The integrated UI for users to select privacy practices. (2) The completion question in the questionnaire mode APPGs. (3) The multiple-choice question in the questionnaire mode APPGs. (4) Preview of the generated privacy policy. (5) User instructions. (6) The selection page for document mode APPGs.

tential problems and their broader impact, it is worthwhile to conduct a market penetration analysis on mainstream APPGs. Moreover, by scrutinizing the features of popular APPGs, we can glean insights into the demands and preference of both the market and developers.

Numerous previous studies have analysed privacy policies [33, 47, 48, 51, 52, 56, 61, 64, 65, 71] from various perspectives. However, the majority of existing compliance analyses in relation to privacy regulations lack fine-grained scrutiny of specific clauses and requirements. APPGs often claim the generated policies are compliant with privacy regulations. A more nuanced analysis that hones in on the specific requirements stipulated by privacy regulations is important for a deeper understanding of APPGs' quality. Other issues relate to mobile apps' dependence on device permissions, for instance, *LOCATION* and *CAMERA*, to function normally. Generally, we find users are more vigilant when directly providing personal information, but underestimate the impact brought by device permission requests, which when granted by a user allow apps to continuously collect critical personal information without further user approval or consent. Therefore, it is significant to examine whether developers correctly display the needed device permissions in the generated privacy policies with APPGs for mobile apps. To evaluate current APPGs and their use, we investigate the following research questions:

RQ1: What APPGs exist for mobile apps? What are the differences between them? (Section 2)

RQ2: How many mobile apps' privacy policies could be generated by APPGs? Why are some APPGs more popular? (Section 3)

RQ3: To what extent do APPGs comply with privacy regulations in terms of specific requirements, data rights, device permission disclosure, and self-integrity? (Section 4)

An overview of our methodology is shown in Figure 2. We first collected an initial set of APPGs. After removing duplicates and filtering out irrelevant results, we identified 10 APPGs for mobile apps as our research objects. We conducted an assessment on multiple dimensions, including their features and the recognition of the extent of data use. We then conducted a large-scale market penetration analysis. Based on a random sample of apps from the Google Play app store, we successfully downloaded 46,472 privacy policies (Crawled Privacy Policy Collection) as a large-scale dataset for mobile apps. We found that 15.7% of apps do not provide a privacy policy link, 22.3% of them are low-quality privacy policies, and 20.5% of these privacy policies fail to provide a privacy

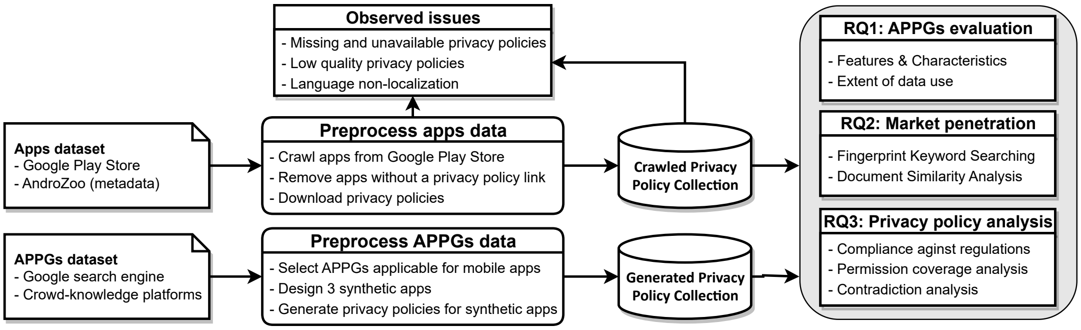

Observed issues
- Missing and unavailable privacy policies
- Low quality privacy policies
- Language non-localization
Apps dataset
- Google Play Store
- AndroZoo (metadata)
Preprocess apps data
- Crawl apps from Google Play Store
- Remove apps without a privacy policy link
- Download privacy policies
Crawled Privacy Policy Collection
APPGs dataset
- Google search engine
- Crowd-knowledge platforms
Preprocess APPGs data
- Select APPGs applicable for mobile apps
- Design 3 synthetic apps
- Generate privacy policies for synthetic apps
Generated Privacy Policy Collection
RQ1: APPGs evaluation
- Features & Characteristics
- Extent of data use
RQ2: Market penetration
- Fingerprint Keyword Searching
- Document Similarity Analysis
RQ3: Privacy policy analysis
- Compliance aginst regulations
- Permission coverage analysis
- Contradiction analysis

Figure 2: Overview of data flow in the methodology for our empirical study.

policy in English, the first language in the target market. We then generated 30 privacy policies (Generated Privacy Policy Collection) using the 10 APPGs for three synthetic apps and conducted a market penetration analysis. Our results show 20.1% app developers favorably use APPGs to generate their privacy policies, and #2 *App Privacy Policy Generator* is the most popular one, boasting a 72.7% adoption rate. In an effort to more accurately assess compliance with privacy regulations, we scrutinized the generated privacy policies. Our findings revealed a substantial level of noncompliance with privacy laws and a frequent under-claiming of data rights and highly concerning privacy practices, especially the most popular #2. In addition, our results suggest dangerous permissions are commonly missing in the generated policies of APPGs, and inappropriate APPGs selection will hinder developers' capacity to include essential device permissions. Furthermore, we found that APPGs might introduce more privacy policy contradiction issues, undermining the self-integrity.

Contributions. To our knowledge, this is the first large-scale, exhaustive empirical study of automated privacy policy generators for mobile apps. Our observations, findings, and insights will benefit stakeholders beyond just app developers, including APPG providers and privacy regulators from the APPG design perspective and legal perspective. We make the following contributions:

- We perform a systematic empirical analysis of Automated Privacy Policy Generators (APPGs), covering various aspects, such as features, characterizations, and levels of recognition of data use.
- We conduct a systematic analysis of the privacy policies of 99,149 apps in the Google Play Store. Specifically, our dataset includes 46,472 privacy policies extracted from these apps and available at [20].
- We discuss implications for APPG users, APPG providers, and privacy regulators to improve the APPG ecosystem.

2 Online Automated Privacy Policy Generators for Mobile Apps (RQ1)

This section addresses RQ1, describing what APPGs exist for mobile apps and what differences there are between them. We manually collected and identified 10 publicly available online APPGs for mobile apps. We analysed various characteristics and specifically the range of possible data uses in apps.

2.1 Collecting Online APPGs for Mobile Apps

To better understand the current status of APPGs, we identified popular APPGs for mobile apps by using the Google search engine and exploring crowd-knowledge platforms. In detail, we employed the Google search engine to broadly search APPGs for mobile apps. We acted as a hypothetical citizen developer who is trying to generate a privacy policy for their app. We created an initial set of search terms, including "*Apps Privacy Policy Generators*", "*Apps Privacy Policy Generators Online*", "*Privacy Policy Automated Generation*", "*Google Play Store Privacy Policy Generator*", "*GDPR Policy Generators*". For each search keyword, we evaluated all entries on the first result page. In addition to direct results from Google search, we also identified APPGs by using our Google search terms again in app developer crowd-knowledge platforms including Stack Overflow [27], GeeksforGeeks [12], and GitHub [14]. To evaluate search results, we manually identified and read highly related threads, and collected the mentioned APPGs. After collecting an initial set of APPGs and removing duplicates, we filtered out irrelevant APPGs based on the following two criteria: 1) APPGs that only support privacy policy generation for websites; and 2) APPGs that do not actually provide their proposed functions. For example, while some APPGs claim that they are able to generate privacy policies for mobile apps, they do not actually provide such a service.

Table 1 presents 10 identified APPGs in our dataset. "Freeware" represents whether a generator needs monetary investment to use. "Free" means the APPG provides a free version and "Paid" means the APPG provides a premium version that

Table 1: APPGs for mobile apps identified in this study.

#	Name	Freeware	Registration	Mode
1	Iubenda [16]	Free&Paid	Required	User Interface
2	App Privacy Policy Generator [5]	Free only	Not required	Questionnaire
3	Termly [28]	Free&Paid	Required	Questionnaire
4	Privacy Policies [22]	Free&Paid	Required	Questionnaire
5	App Privacy Policy [4]	Free only	Not required	Questionnaire
6	Privacy Policy Online [25]	Free only	Not required	Questionnaire
7	Terms Feed [29]	Free&Paid	Required	Questionnaire
8	Website Policies [31]	Free&Paid	Required	Questionnaire
9	Free Privacy Policy [11]	Free&Paid	Required	Questionnaire
10	Lawpath [18]	Paid only	Required	Document

unlocks more features. "Registration" indicates whether a user needs to register an account to use the APPG. All APPGs we collected are based on question answering and template completion ('boilerplate'), however, they have various different "Modes" to handle information interactions with users as we demonstrated in Figure 1. "User interface"(UI) mode means that users need to select the relative data practices through an integrated UI as à la carte manner, then answer a few other questions about the app's and developers' basic information. "Questionnaire" mode only requires users to complete a long questionnaire that covers all potential data practices and a privacy policy will be generated depending on the answers. This is the most common and popular type of APPGs (8/10). "Document" mode is common in legal consultation websites. Users need to purchase access to a privacy policy document, then answer a series of questions to complete the template. Notably, all analyses of APPGs in this paper include features in both free and paid versions. We conducted all analyses of APPGs in May 2022.

2.2 Characterization

A comprehensive assessment is conducted on identified APPGs along 10 dimensions. The first five (1-5) target practicability, user-friendliness, and convenience from potential APPG users' perspectives. The next three (6-8) cover legal compliance with privacy regulations, according to the tools' claims. The last two (9, 10) focus on the understandability of privacy policies. The results are shown in Table 2. We define these 10 dimensions as follows:

Extent of Data Recognised (1). Mobile apps normally require data from direct user input or device sensors. Some apps may also provide data to third-party services, for revenue or to enhance the user experience. Privacy regulations such as GDPR [Art. 14(1)(d)] and CCPA [§1798.130(a)(5)(B), 1798.110(c), Regs §999.308(c)(1)(d))], can require that app developers include information about all collected or shared data within their privacy policy. However, some APPGs only recognise a limited range of data practices for inclusion in privacy policies. As failing to include all data practices could be regarded as a regulatory violation in some jurisdictions, this is an important characteristic and is discussed in more detail in the next section (Section 2.3). Tentatively, we use (●) to denote that the APPG recognises a wide range of data use, (◐) for a smaller range of data use, and (○) for a few kinds of data use.

Customizability (2). Customizability addresses two concerns: can users add privacy practices that were not originally included in the APPG pipeline, and do users have the freedom to add customized clauses in the generated privacy policy document? The former can help developers to provide a more specific and accurate privacy policy; the latter can more easily accommodate developers who want to address additional concerns about users' privacy. APPGs are denoted as (●), (◐), and (○), as indications that they are positive for both, one, and zero of these questions, respectively.

User Instruction (3). The main target audience of APPGs is expected to be developers with little or no legal knowledge about privacy and data protection. Therefore it is crucial to provide sufficient user instructions to help correctly utilize the APPG. User instructions can be generally categorised into three types: providing further explanation of obscure questions (an example is shown in Figure 1), elaborating terms by listing examples, and supporting users with an interactive help center. APPGs are scored as (○) if they only provide one or nil occurrence of the above user instructions, (◐) for two to five occurrences, and (●) for more than five occurrences. Especially, APPG #1 *Iubenda* also provides an introduction video and comprehensive documentation.

Complexity (4). This dimension indicates the general learning cost for APPG users, and higher complexity reflects a higher cost. For UI-mode APPGs, users have to spend more time getting familiar with the UI and learning operation procedures, therefore, we manually set them as (●). For questionnaire-mode APPGs, we found that the complexity is related to the number of questions that users are asked in the generation process. Based on the "Statistic summary" in Table 3, if the sum of maximum multiple-choice questions and maximum completion is less than 10, the APPG is marked as (○), if it is between 10 to 50, the APPG is marked as (◐); and if it is greater than 50, the APPG is marked as (●). For document-mode APPGs, users only need to select a document from the library and fill up several questions about their basic information, thus, we manually set them as (○).

Publishing Support (5). This dimension reflects how convenient it is to deploy the generated privacy policy. Every APPG provides at least one of the following publishing options: 1) a permanently hosted website link containing the privacy policy; and 2) the generated privacy policy context in editable HTML format. In section 3.3.1, we presented that a large amount of privacy policy links in the market lead to inaccessible websites, and we believe that a permanently hosted website link provided by APPGs can mitigate this

Table 2: The characterization of 10 APPGs on 10 different dimensions.

# and APPG Name	Extent of Data Use	Customizability	User Instruction	Complexity	Publishing Support	GDPR	CCPA	LGPD	Multilingual Support	Readability
#1 Iubenda	●	◐	●	●	●	●	●	●	●	○
#2 App Privacy Policy Generator	◐	○	○	○	◐	○	○	○	○	●
#3 Termly	●	◐	●	●	●	●	●	○	○	●
#4 Privacy Policies	◐	○	○	◐	●	●	●	○	○	○
#5 App Privacy Policy	○	○	○	○	●	○	○	○	○	◐
#6 Privacy Policy Online	○	○	○	○	●	○	○	○	○	◐
#7 Terms Feed	◐	◐	○	◐	●	●	●	○	○	○
#8 Website Policies	●	●	◐	●	●	●	●	○	○	○
#9 Free Privacy Policy	◐	○	○	◐	●	●	●	○	○	○
#10 Lawpath	○	○	●	○	◐	○	○	○	○	◐

○: low level / does not support; ◐: intermediate level / partially support; ●: high level / fully support

problem to some extent. (●) indicates the APPG provides both options and (◐) means APPGs only provide option 2.

GDPR (6), CCPA (7), and LGPD (8). These three dimensions are binary, simply indicating whether the APPG claims that they provide compliance with the corresponding privacy regulations. (●) denotes support and (○) denotes non-support. No APPG claims compliance with other privacy regulations such as APP [7] and PIPL [21].

Multilingual Support (9). In Section 3.3.3, we discuss how current privacy policies suffer from the language localization problems. APPGs that provide multi-language support should benefit both app developers and app users. Additionally, in CCPA [Regs §999.308(a)(2)(d)], it stipulates that online notices should follow generally recognized industry standards, such as the W3C Web Content Accessibility Guidelines [30]. We use (●) to denote that APPG can generate privacy policies in more than one language, otherwise (○).

Readability (10). While a long privacy policy can provide more comprehensive and detailed descriptions of data practices in apps, adding unnecessary information to a privacy policy may lead to information overload [39, 63]. Although APPGs can generate more concise privacy policies that are more specifically tailored to the data and data practices relevant to each app, this does not necessarily mean that those privacy policies are more readable. Some regulations require mobile app developers to use clear and understandable language in privacy statements [46, 62]. Also, W3C Web Content Accessibility Guidelines [§3.1] requires that "make text content readable and understandable". However, those high-level principles do not give details for this requirement. Thus, we adopt the Flesch Reading-Ease Test [44] to evaluate the readability of generated privacy policy context. A higher score indicates the content is easier to read and understand, the score scales from 0 to 100. The average readability score of the privacy policies of 12 leading mobile apps [19], who have over one billion installs, is 46. We use (●) when the readability score is greater than or equal to 46, (◐) for 40 to 46, and (○) for less than 40. Specific readability scores are listed in Appendix Table 10.

Since the assessment of apps in these dimensions involved some subjective judgments, the first two authors assessed the 10 APPGs individually. For any disagreement at the table, they discussed and agreed on the same answer, and if the disagreement persisted, a third author joined the discussion to facilitate a resolution. As shown in Table 2, the level of Complexity is positively associated with the level of Extent of Data Use, Customizability, and User Instruction. Most of the APPGs (8/10) exhibit complete publishing support. As for compliance with privacy regulations, more than half the APPGs (6/10) claim that their privacy policies conform to the GDPR and CCPA, and only #1 *Iubenda* claims compliance with LGPD. Only #1 *Iubenda* supports generating privacy policies in languages other than English. All APPGs' readability scores are between 30 and 50, which indicates that readers should have at least a college-level education background to easily read the policies [44]. #2 *App Privacy Policy Generator* has the highest readability score and is the most frequently used APPG on the current market, as discussed in Section 3.5.

2.3 Extent of Data Use Recognised by APPGs

In this section, we further investigate the extent of data, data use, and data practises recognised by APPGs to support the generation of privacy policies. We identified four major aspects: the app's basic information, users' personal information, device permissions, and third-party services. We carefully gathered all the possible questions and potential options that APPG users might face. If any question or option explicitly pertains to a specific data type, permission, or third-party service, then it is classified as "recognised" (●), otherwise "absent" (○). The specific item enumeration process is available at [20].

App's basic information. Apart from user names, providing at least one communication channel to app users

Table 3: The breakdown of APPG's recognition of data use. The first row of each section is the number of APPGs as per Table 1. "Multiple-choice questions" refers to whether developers select options from listed items, and "Completion questions" means developers fill in blanks with plain text. (Selecting a date is also regarded as a completion question.) Since some questions sometimes unlock later questions, we use "Minimum" and "Maximum" to reflect the lower and upper bounds. *If date of birth is included in the APPG, then we consider that Age group can be inferred.

App's basic information	1	2	3	4	5	6	7	8	9	10
Name of the app	●	●	●	●	●	●	●	●	●	●
Name of the developer	●	●	●	●	○	○	●	●	●	●
State/Country of the developer	○	○	●	●	●	●	●	○	●	●
Physical address of the developer	●	●	●	●	○	○	●	●	●	●
Email address of the developer	●	●	●	●	●	●	●	●	●	●
Phone number of the developer	○	●	●	●	○	○	●	○	●	○
Customizable policy effective date	●	●	●	○	○	○	○	○	○	●
Users' personal information	1	2	3	4	5	6	7	8	9	10
Users' general personal information										
Name	●	●	●	●	○	○	●	●	●	○
Age group* (adulthood or underage)	●	○	●	○	○	○	○	●	○	○
Phone numbers	●	●	●	●	○	○	●	●	●	○
Email address	●	●	●	●	○	○	●	●	●	○
Residential addresses	●	●	●	●	○	○	●	○	●	○
Users' sensitive personal information										
Health data	●	○	●	○	○	○	○	●	○	○
Biometric data	●	○	●	○	○	○	○	●	○	○
Gender data	●	○	●	○	○	○	○	○	○	○
Information revealing race or ethnic origin	○	○	●	○	○	○	○	●	○	○
Government identifiers (e.g., medical card number)	●	○	●	○	○	○	○	●	○	○
Device permissions	1	2	3	4	5	6	7	8	9	10
Calendar permission	●	○	●	○	○	○	○	○	○	○
Camera permission	●	○	●	●	○	○	●	○	●	○
Contacts permission	●	○	●	●	○	○	●	●	●	○
Location permission	●	○	●	●	○	○	●	●	●	○
Microphone permission	●	○	●	○	○	○	○	○	○	○
Phone permission	●	○	●	○	○	○	○	○	○	○
Sensor permission	●	○	●	○	○	○	○	○	○	○
SMS permission	●	○	●	○	○	○	○	○	○	○
Storage permission	●	○	●	○	○	○	○	○	○	○
Selected third-party services	1	2	3	4	5	6	7	8	9	10
Facebook account access	●	●	●	○	○	○	○	●	○	○
Twitter account access	●	○	●	○	○	○	○	●	○	○
Google Analytics for Firebase	●	●	●	●	○	○	●	●	●	○
Flurry Analytics	●	●	●	●	○	●	●	●	●	○
AdMob	●	●	●	●	○	●	●	●	●	○
Statistic summary	1	2	3	4	5	6	7	8	9	10
Minimum multiple-choice questions	/	4	58	7	2	5	7	38	7	/
Maximum multiple-choice questions	/	5	79	26	2	5	26	63	26	/
Minimum completion questions	/	4	11	2	2	3	2	2	2	/
Maximum completion questions	/	4	76	22	2	3	22	13	22	/

for any potential inquiries is commonly required by privacy protection and regulation laws, and all the APPGs support this with different levels of granularity.

`Users personal information`.We manually separate this into two sub-categories: general personal information, and sensitive personal information. Most APPGs recognise general personal information, and only some APPGs (3/10) explicitly identify sensitive information.

`Device permissions`. According to the Android developer's guide [1], there are nine dangerous permission groups, namely *CALENDAR*, *CAMERA*, *CONTACTS*, *LOCATION*, *MICROPHONE*, *PHONE*, *SENSORS*, *SMS*, and *STORAGE*, on mobile phones. We expect the collection and use of data related to these permissions to be clearly declared in a privacy policy, therefore, they are set as row indices. We found that more than half of the APPGs (6/10) recognise *CAMERA*, *CONTACTS*, and *LOCATION*, and only two APPGs identify all permissions. Failing to correctly state sensor permissions in the privacy policy could lead to serious privacy issues, so we further investigated whether users can correctly utilise the APPG to state all claimed sensor permissions in Section 4.2.

`Third-party services`. Third-party services have become a significant part of mobile apps. Taking into account the diversity of service providers and their popularity, we selected five third-party services as row indices, as most APPGs either fully or partially cover them. Overall, APPGs show decent coverage among these third-party services.

`Statistic summary`. We counted the completion questions and multi-choice questions (examples are illustrated in Figure 1) that users would encounter for the questionnaire-mode APPGs. Although questions may vary dramatically, we take it as a quasi-indicator to approximately reflect users' learning and time cost to use the APPGs.

As it involves intensive manual work, to avoid the effect caused by potential human error, we employ the same strategy as introduced in the previous section. Cohen's Kappa [41] $\kappa = 0.92$ for the initial manual labelling, which is an almost perfect level of agreement. As shown in Table 3, #1, #3, and #8 are the best in terms of the recognised data use, they almost cover every data index listed in the table. #2, #4, #7, and #9 decently cover users' general personal information and third-party services, but only a few or null users' sensitive personal information and device permissions. #5, #6, and #10 merely cover some app's basic information and failed to support the rest of the categories. Above all, users need to carefully select appropriate APPGs, otherwise, the unsupported data use will be missing in the generated privacy policy.

Finding 1: APPGs are a handy solution for developers to draft privacy policies for their apps. They have various features, characterizations, and levels of recognition of data use.

3 Market Penetration of APPGs (RQ2)

The previous section carefully examined various characteristics of APPGs. In this section we consider the questions: How large is the APPG market? And are the seemingly more functional APPGs actually more popular in the market?

3.1 Status Quo of Mobile Apps' Privacy Policy

Privacy policies play an essential role in mobile app ecosystems. An app without a privacy policy is not necessarily malicious, but developers are required to provide privacy policies due to regulations in some jurisdictions. These policies are also increasingly demanded by mobile app users and marketplaces. According to the Apple App Store Developers page [3], a privacy policy is required when submitting new apps or app updates: *"By adding the following links on your product page, you can help users easily access your app's privacy policy and manage their data in your app.* ***Privacy Policy (Required):*** *The URL to your publicly accessible privacy policy."* Additionally, in Google Play Console Help Center [2], developers *"...must include a link to your privacy policy"* to prepare the app for review. Although platform principles regulate the existence and quality of privacy policies, we still discovered some critical problems in the mobile apps market. Other forms of privacy notices, such as privacy nutrition labels (PNLs) [49,54,68,69], focus more on providing succinct information to end users, potentially sacrificing information relevant or required by specific regulations or marketplace requirements.

3.2 Crawled Privacy Policy Dataset

To better understand the status and quality of existing privacy policies, we need a large-scale privacy policy dataset. However, existing datasets suffer from various problems.

The differences between datasets for previous research and our research goals. Existing datasets are either not focused on mobile apps, biased, small, or outdated. First, some prior research focuses on privacy policies across all platforms, including websites and mobile apps [33,50,61], while we only target mobile apps. Since they did not include metadata (e.g., platforms) of the collected policies, we cannot filter them out from other platforms. Second, some research only focuses on apps of certain groups, and their collected datasets are not representative enough for apps in the market. For example, one study [47] manually collected 5,684 "globally popular apps." This selection criterion leads to two critical problems: a) The top overall popular apps appear in a limited set of categories, such as *Social* and *Finance*, which can drive out apps from less popular categories such as *Parenting* or *Libraries & Demo*. b) These globally popular apps do not distribute evenly across various app categories, which may cause potential bias and unrepresentative findings. Third, some existing datasets are small and can contain potential bias. For example, APP-350 [71] only analysed 350 privacy policies, and another study [51] conducted fine-grained analysis on only 304 privacy policies. Finally, due to the rapid iteration of apps, apps' privacy policies, and APPGs, some prior datasets can quickly become outdated. In summary, to make our findings comprehensive, generalizable, and up-to-date, we needed to collect a current large-scale dataset that can assemble apps of different popularity and from various app categories.

Google Play Store. Android mobile apps are often more transparent and friendly to academic research compared to iOS mobile apps [43,45]. There are many app markets for the Android platform, but Google Play is the largest and the most accessible app market, with over two million apps, according to AppBrain [6]. Thus, we only focus on it in this study. Some Android apps are specially designed for Android smartwatches instead of mobile phones; therefore, we intentionally excluded apps belonging to categories named *Watch apps* and *Watch faces*.

Our dataset. We collected a new large-scale privacy policy dataset from existing apps on the Google Play Store. We also collected app metadata from AndroZoo [32], which is a large-scale and growing Android app collection extracted from multiple sources, including the Google Play app store. We removed duplicates and apps from other app markets and then randomly sampled 268,500 apps as our initial app dataset, which is around 10% of the whole app population. From this initial sample, we observed that some apps are either invalid or not usable. These were mainly dummy apps or student apps with a package size normally less than 10 KB, so we removed invalid apps like these. We further excluded unavailable apps that returned error messages on their Google Play page. Apps can be unavailable for many reasons including geographic differences [47] caused by government censoring, or being removed by Google because of disruptive adware, malware, restricted content, or other reasons. After these exclusions, we were left with 99,194 usable apps and their metadata. We then employed *google play scraper* [15] to obtain app information including privacy policy links, app categories, and required device permissions. Based on the available privacy policy links shown on the app's homepages, we further utilize the *BeautifulSoup* [8] and *Selenium* [26] python packages to download those websites. Eventually, we obtained 46,472 privacy policies, and we refer to this privacy policy dataset as **"Crawled Privacy Policy Collection."** We conducted the above data-gathering process in March 2022.

3.3 Observations on Status Quo

In this section, we report an analysis of app privacy policies and identify three common problems in existing apps: missing/unavailable privacy policies, low-quality privacy policies, and language non-localization problems. Although similar problems were discussed in several previous studies [47,49,71], considering the rapid development of markets, we believe it is worthwhile to revisit those critical issues.

3.3.1 Missing Links and Unavailable Privacy Policies

Providing a privacy policy is essentially required by platform principles and privacy regulations, however, we observed that 15.7% (15,572/99,194) of apps do not provide a privacy policy link on their Google Play homepage. A previous study [71] reported that this statistic was 49.1% in 2019. Seemingly, as more regulation has been introduced, more apps have provided a privacy policy. However, we found that 37.5% (37,150/99,194) of privacy policy links lead to unavailable websites with error messages such as "403 Forbidden" or "404 Not Found." This problem could be caused by potential reasons such as deliberately providing a dummy link, an outage, or the removal of the website server. Hence, in addition to simply requiring app developers to provide the privacy policy, the mobile app market also needs to frequently and regularly verify the validity of provided privacy policy links. Also, a permanent hosted privacy policy link provided by APPGs can help developers to mitigate availability issues to some extent.

3.3.2 Low-Quality Privacy Policies

Several previous studies [51,71] reported that it was not rare that some apps only provide a low-quality privacy policy, and we observe similar issues in our dataset. Firstly, some privacy policies do not contain meaningful privacy-related context, or are just dummy websites. For example, "Bway App" is a free app that provides football results, statistics, trends, and match result predictions, with over 10,000 installs; however, their privacy policy website [23] does not contain any content, and the file size of the crawled privacy policy is close to 0 KB. In addition, some are apparent "homemade" privacy policies that are too general to include essential data practices. For instance, "Easy Communication" is designed to help people with autism, cerebral palsy, dyslexia, intellectual disability, and other special needs to communicate easily. Its privacy policy [24] only has 156 words, and ambiguously mentions required device sensor permissions without any elaborations. Based on our observations and previous work, we empirically set the file size threshold as 2 KB [51], and the document length threshold as 200 words [47]. If a crawled privacy policy does not meet both criteria, it will be regarded as a low-quality privacy policy. In total, we identify 22.3% (10,375/46,472) low-quality privacy policies.

3.3.3 Language non-Localization Problem

Language localization is necessary to promote apps in a global market. Although app UI and content can be translated, app developers may neglect the privacy policy. We employed the Python package *langdetect* [17] to detect the language of the crawled privacy policy documents and found 20.5%

(9,523/46,472) of apps do not provide an English privacy policy, although these apps were released in markets where the primary language is English. Interestingly, the top five non-English languages are Spanish (15.6%), Portuguese (9.7%), German (8.8%), Korean (8.7%), and French (7.6%). This trend might be attributable to the implementation of the European GDPR and Brazilian LGPD. We observed this problem even with top apps from some large companies. For the privacy policies of the top 1,000 most installed apps (all > 500k installs) in our dataset, there are still around 14.1% that do not provide an English privacy policy. APPGs may be a solution to language non-localization problems. Only APPG #1 *Iubenda* supports multilingual generation, but some APPGs (#4, #7, and #9) have included placeholders for future options for language selection.

3.4 Synthetic Apps and Generated Privacy Policy Collection

To identify whether privacy policies could be generated by APPGs, we first need to build the ground truth with self-generated privacy policies for each APPG. As shown in Table 1, seven APPGs require users to register an account to use the service, and six of them also require users to pay subscription fees to unlock all features. We registered as required and paid subscription fees. Synthetic apps are commonly used in similar empirical studies such as [53, 61]. We designed and tailored 3 synthetic apps specifically based on APPGs' features and characteristics as summarised in Table 2 and Table 3. This was so that the boilerplates and pre-defined clauses of every APPG can be fully explored and reflected in the generated privacy policies. The functions and features of these synthetic apps are as follows:

Synthetic App 1. A toy-like app that collects only general personal information and does not need to comply with GDPR, CCPA, or LGPD.

Synthetic App 2. A social app that enables people to interact and communicate with others. Users need to create an account by providing their general personal information, and the app requires all device permissions. The app also accesses third-party services and only needs to comply with GDPR.

Synthetic App 3. A hypothetical omnipotent app that requires users to provide all general and sensitive personal information to function. The app also needs all device permission and access to third-party services. This app needs to meet the requirements of GDPR, CCPA, and LGPD.

These three synthetic apps are various in terms of functionality and sophistication, and they decently cover all data use mentioned in Table 3. Therefore, they are capable of being the boilerplate apps for APPGs' ground truth in the majority of cases. For each APPG, we created the same three custom privacy policies to test whether a privacy policy can be generated by one of analysed APPGs. Given ten APPGs and three synthetic apps, we obtained 30 privacy policies as **"Generated Privacy Policy Collection"** in total. We utilized it and "Crawled Privacy Policy Collection" to conduct the following market penetration analysis.

Table 4: Summary of market use of different APPGs.

Method	**Market Occupancy**
Fingerprint Keyword Searching	6.6% (3,066)
Document Similarity Comparison	18.1% (8,425)
Intersection	4.4% (2,042)
Union (Total)	**20.1% (9,332)**

3.5 Market Penetration Analysis

We report a market penetration analysis to highlight the extent to which APPGs are used and to provide insights about why some APPGs are more popular. Specifically, we employ two methods to detect whether a privacy policy could be created by one of these APPGs: fingerprint keyword searching and document similarity comparison.

Fingerprint Keyword Searching (FKS). APPGs offer various publishing features, including the provision of a direct link as a permanent website host or an editable HTML document. We observe that to better advertise themselves, APPGs typically embed their company names into the URLs of the generated privacy policy websites or the editable HTML document. Consequently, we construct the fingerprint keyword set grounded in this feature. To cultivate this set, we manually inspected these APPGs and the collection of generated privacy policies. To mitigate the likelihood of false positives, we verified the authenticity of all keywords individually. Specifically, we incorporated a candidate fingerprint keyword into the set only if it appeared consistently across all three ground-truth privacy policies crafted for synthetic apps. All keywords are deliberately sensitive to case and format. The full keyword list is available at [20].

Then, the collected keyword set is employed to perform FKS in the crawled privacy policy collection. For each privacy policy in our collection, if we find a match in either its website link or its document context with one of the fingerprint keywords, the policy is considered to be generated by the corresponding APPG. For example, during the examination of policy website URLs, we use the keyword "iubenda.com" to search URLs like "www.iubenda.com/123abc..." for APPG #1 *Iubenda*. Similarly, while scrutinizing policy HTML documents, identifiable phrases like "iubenda hosts this content" were utilized as markers

Document Similarity Comparison (DSC). As we mention in Section 2.2, some APPGs allow users to customize the content to some extent, and APPG users may also further adapt or polish the initial generated privacy policy in editable HTML format before publishing. Consequently, the fingerprint keywords, even sentence-level segments, are not necessarily included in a generated app privacy policy sometimes,

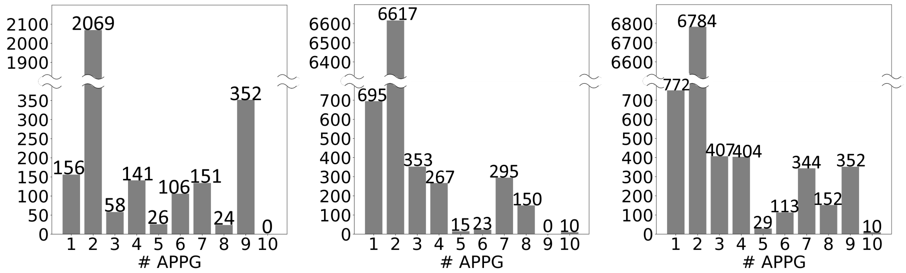


Figure 3: The number of privacy policies generated by APPGs, detected by FKS (left), DSC (mid), and both (right).

and so the searching-based method may not work in those cases. To tackle this problem, we design a semantic-based method with better robustness, called document similarity comparison.

In natural language processing (NLP) tasks, researchers adapt the concept from mathematics by treating documents as vectors, to obtain the similarity between two documents. TF-IDF [57] is a common numerical statistic designed to reflect the relative importance of words in a document corpus, and thus it can be adapted to convert documents into vectors. By doing so, the similarity between two documents can be represented as the similarity of two vectors, and we choose cosine similarity as the most common one. If the average similarity between a crawled privacy policy and three ground-truth generated privacy policies is greater than a threshold β, we consider it a match, i.e., could be created by the current APPG. If a privacy policy matches more than one APPG, we take the one with the highest similarity. To define the value of β, we first calculate the document similarities between the three generated privacy policies of each APPG as an initial value. We then randomly sampled 100 privacy policies that are generated by one of 10 APPGs based on fingerprint keyword searching and evaluate the similarities compared to the three generated privacy policies of each APPG to fine-tune the β. Based on our data, we empirically set $\beta = 0.75$.

Results. We employed the proposed methods to provide an estimation of the market penetration of APPGs. Among 46,472 valid privacy policy documents from apps at the Google Play Store, as shown in Table 4, we found 6.6% (3,066) privacy policies are highly likely generated by one of the APPGs by using fingerprint keyword searching and 18.1% (8,425) by using document similarity comparison. We also checked the intersection size, which is 4.4% (2,042), indicating that these two methods are complementary to each other. The union of the previous two methods is **20.1%** (9,332), indicating an upper bound on possible market occupancy ratio. Overall, the results show that APPGs do play a considerable role in the current mobile app market.

Figure 3 presents the market occupation for each APPG. We can see that APPG #2 *App Privacy Policy Generator* dominates as the most popular APPG, with a 72.7% adoption rate. It could be attributed to its free-of-charge and registration, low complexity, and best readability score. #1 *Iubenda* and #5 *App Privacy Policy* also take decent market shares. #10 *Lawpath* is the least used APPG in our data, perhaps related to its price to use of $199 for one privacy policy document, which is the highest among all APPGs. We notice popular APPGs generally have a low level or partial support on all data practice dimensions, therefore, we conclude that users may prefer to select easy-to-use tools rather than spend extra time learning how to master a sophisticated but more functional APPG. However, the ease of use of APPGs may come at the cost of a potentially-higher risk to breach privacy regulations.

Finding 2: The market occupancy ratio of 10 examined APPGs is around 20.1%, and #2 *App Privacy Policy Generator* is the most popular one, boasting a 72.7% adoption rate. Moreover, users tend to select easy-to-use APPGs even though at the cost of a potentially-higher risk to breach privacy regulations.

4 Assessment of Privacy Policies (RQ3)

Generators are designed to generate compliant privacy policies. This section reports a detailed assessment of generated privacy policies.

4.1 Privacy Compliance Against Regulations

APPGs need to be designed in a way that allows developers to produce compliant privacy policies. Furthermore, they must be updated to reflect evolving laws. In Table 2, six APPGs claim that they are able to generate policies in compliance with CCPA and GDPR, and only #1 *Iubenda* mentions compliance with LGPD. However, their actual compliance with generated policies may be different from what they claim. Zimmeck et al. [70] conducted two studies in May 2020 and January 2021, revealing that some APPGs had significant compliance issues, such as failing to create CCPA-compliant policies. We perform a follow-up inspection in May 2022 based on generated privacy policies by Synthetic App 3, since

Table 5: LGPD privacy policy requirements and APPGs' compliance in May 2022. Numbers indicate APPGs as per Table 1.

LGPD requirement	1	2	3	4	5	6	7	8	9	10
Identify and document the legal bases for processing personal data. [Art. 1, 2]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✓
Disclose, collect and maintain valid proof of consent. [Art. 8 §1-6]	✓	✗	✗	✗	✗	✗	✗	✓	✗	✗
Include required data collection disclosures in the privacy policy in a clear, adequate, and notable manner. [Art. 9]	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓
Disclose the essential data rights to the data subject in a clear, adequate and ostensible manner. [Art. 9, 19]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗
Appoint a data protection officer (DPO) and publicly disclosed the contact information [Art. 41 §1]	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗
Implement and disclose a security protocol (or such mechanism) to protect personal data. [Art. 46]	✓	✓	✓	✓	✗	✗	✓	✗	✓	✓
State and comply with cross-border data transfer requirements if the data controller conducts international transfer of personal data. [Art. 33]	✓	✗	✓	✓	✗	✗	✓	✗	✓	✓
Disclose the processing of children's data in a specific and highlighted consent, including types of data collected, the way it is used and data rights. [Art. 14]	✓	✓	✗	✓	✓	✓	✓	✓	✓	✗

Table 6: Tallies of the APPGs' compliance against legal requirements in privacy regulations. The individual requirements of LGPD are shown in Table 5. "N.R." stands for "no record". The enforcement date of LGPD is September 2020.

#	**GDPR**			**CCPA**			**LGPD**
	May'20	Jan'21	May'22	May'20	Jan'21	May'22	May'22
1	8/8	8/8	8/8	14/18	14/18	14/18	8/8
2	N.R.	N.R.	3/8	N.R.	N.R.	3/18	3/8
3	8/8	8/8	8/8	3/18	15/18	15/18	6/8
4	8/8	8/8	8/8	5/18	16/18	16/18	6/8
5	N.R.	N.R.	0/8	N.R.	N.R.	2/18	1/8
6	N.R.	N.R.	0/8	N.R.	N.R.	2/18	1/8
7	8/8	8/8	8/8	5/18	16/18	16/18	6/8
8	N.R.	N.R.	8/8	N.R.	N.R.	11/18	5/8
9	8/8	8/8	8/8	5/18	16/18	16/18	6/8
10	N.R.	N.R.	2/8	N.R.	N.R.	2/18	4/8

they are supposed to comply with all compliance requirements of GDPR, CCPA, and LGPD.[1] Table 6 shows tallies of generators' compliance with legal requirements in May 2020, January 2021, and May 2022 for GDPR, CCPA, and LGPD.

There is a notable discrepancy between the APPGs' claimed ability to generate policies compliant with specific regulations and their actual compliance levels. In addition, across the table, compliance with GDPR and CCPA requirements improved over time from May 2020 to May 2022 for several APPGs (#3, #4, #7, #9), demonstrating an adaptation to regulatory standards. An encouraging observation is that APPGs (#1, #3, #4, #7, #8, #9) maintained full GDPR and a majority of CCPA compliance since January 2021. This might be due to the fact that GDPR has been in effect since 2018, providing APPGs with more time to adapt their policies accordingly. Given that LGPD came into effect in September 2020, it is not unexpected that compliance levels would be relatively low in the initial stages. The findings underscore the necessity for APPGs to improve their ability to produce compliant privacy policies consistently across all major regulations. The results also call for ongoing monitoring and evaluation of APPGs to ensure they keep pace with evolving laws and maintain transparency with their users regarding their actual capabilities.

APPGs have varying levels of compliance when it comes to data collection practices. For instance, while GDPR emphasises "Categories of personal data concerned", CCPA requires to provide a detailed list of categories of personal information that should be disclosed if collected. APPGs #1, #3, #4, and #7 demonstrate compliance with GDPR's requirements, but when it comes to CCPA's detailed disclosure requirements, only APPGs #3, #7, and #9 comply. As for third-party data-sharing and data-sale practices, CCPA's emphasis is on "List of categories of personal information collected and sold" highlights this; whilst GDPR identifies "Recipients or categories of recipients of the personal data". APPGs #1, #3, #4, and #7 have clear disclosures in line with third-party disclosure, but others like #2 and #5 lag behind. Overall, APPGs #3 and #7 consistently fare well across all three regulations, indicating a comprehensive privacy policy generation. However, APPGs #2, as the most popular, display apparent gaps compared to others.

Data Right Coverage Analysis. Regulations often necessitate data rights to grant sufficient choice and control to the end users over the personal data collected by businesses. Thus, it is important to assess whether the generated policies clearly disclose data rights. For this, following [GDPR Art.13] and

[1] The compliance analysis of GDPR and CCPA of May 2020 and Jan 2021 are directly obtained from [70]. The requirements and compliance checking for GDPR, CCPA, and LGPD are curated in Appendix Table 11, Appendix Table 12 and Table 5, respectively.

[CCPA §1798.130(a)(5)(A)] [2], we extracted and summarised seven fundamental data rights. We manual examined the existence of disclosure for each data right of 10 APPGs. For comparison, we also scrutinized 12 leading mobile apps, which are mentioned in Section 2.2-Readability. These 12 leading apps are in various categories and published by different companies, with over one billion installations. Table 7 shows that almost all leading apps disclose fundamental data rights in their privacy policies. In addition, results validate the respective assertions made by APPGs, regarding their self-claimed compliance with GDPR and CCPA. Notably, #2 as the most popular APPG, does not disclose any of the data rights in the generated policy document.

Table 7: The disclosure existence of seven fundamental data rights. Numbers in the first row indicate APPGs as per Table 1, and "Apps" denotes the tallies of disclosure for 12 leading apps.

Data Right	1	2	3	4	5	6	7	8	9	10	Apps
Right to Know	✓	✗	✓	✓	✗	✗	✓	✓	✓	✓	12/12
Right to Access	✓	✗	✓	✓	✗	✗	✓	✓	✓	✓	12/12
Right to Processing	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗	12/12
Right to Restrict of Processing	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗	12/12
Right to be Forgotten	✓	✗	✓	✓	✗	✗	✓	✓	✓	✓	12/12
Right to Data Transfer	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗	12/12
Right to Lodge a Complaint	✓	✗	✓	✓	✗	✗	✓	✓	✓	✓	12/12

Privacy Practice Disclosure Analysis. In alignment with the principles of data protection and transparency, it is also imperative that privacy policies encompass disclosure of privacy practices. Based on [2] and [60], this study delineates four highly concerning privacy practices. These are: 1) *Data Encryption*: Users' data are encrypted and transferred over a secure connection. 2) *Government Requests*: It refers to the potential for government entities to request access to users' data for various reasons, including national security or criminal investigations. 3) *Data Breach Notification*: In the event of a data breach where users' information is compromised, it is essential to have a mechanism in place to notify affected users promptly. 4) *Changes Notification*: The privacy policy should mention the procedures for notifying users about significant changes to the policy, including how and when users will be informed. Then, we manually scrutinized the existence of disclosure across 10 APPGs and the same 12 leading apps. Table 8 shows that more than half APPGs disclose *Data Encryption* and *Changes Notification*. Five APPGs include *Government Requests* and only #8 includes *Data Breach Notification* in the generated policy document (*"In the event of a data breach, we will make reasonable efforts to notify affected individuals..."*). In addition, #5 and #6 do not include any privacy practice, and #8 includes all items. Furthermore, #2 as the most popular APPG, only include *Data Encryption*.

Table 8: The disclosure existence of five highly concerning privacy practices. Numbers in the first row indicate APPGs as per Table 1, and "Apps" denotes the tallies of disclosure for 12 leading apps.

Privacy Practice	1	2	3	4	5	6	7	8	9	10	Apps
Data Encryption	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	8/12
Government Requests	✓	✗	✗	✓	✗	✗	✓	✓	✓	✗	11/12
Data Breach Notification	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	1/12
Changes Notification	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗	12/12

Finding 3.1: Noncompliance with privacy laws and under-claiming issues indicate potential need for a more stringent evaluation of APPGs' capabilities.

4.2 Permissions Coverage Analysis

Device (sensor) permissions enable features of mobile apps and are normally critical in terms of personal privacy risks. Intuitively and legally, an app's privacy policy should accurately disclose its permission usages; however, it is common that developers do not list all the permissions of their app in the privacy policy [34, 42, 54, 55, 71]. Developers of mobile apps in the Google Play app store are required to clearly list the permissions they intend to obtain on their homepages[3], and we denote them as P_{claimed}. By taking the self-reported device permission usage as the ground-truth, we then assess whether the APPG's provided UIs, questionnaires, or documents can sufficiently allow the APPGs users to enter permission invocation information by their app. Notably, the self-reported device permission usages by developers may not actually reflect the actual privacy behaviour of apps.

To scrutinise how developers translate used device permissions into policies by APPGs, for each permission (*per*), we use it and Synthetic App 1 to generate a privacy policy ($PP_{App1+per}$). By comparing the differences between $PP_{App1+per}$ and PP_{App1}, we can locate and obtain the phrases/sentences each APPG uses to declare each permission in the privacy policy. For each app, by searching and counting the corresponding phrases/sentences in its privacy policy, we obtain the number of permissions displayed and denoted as $P_{\text{displayed}}$. For each app that uses the APPG, its $P_{\text{recognised}}$ is equal to the recognised dangerous permissions enumerated in Table 3. Then, we introduce two metrics for the extent to which APPGs cover claimed permissions in the generated privacy policy, as follows:

$$\text{Recognised Coverage}\,(RC) = \frac{1}{n}\sum_{1}^{n}\frac{P_{\text{displayed}}}{P_{\text{recognised}}}$$

A higher *RC* indicates the app developers use the APPG more

[2] The specifications are stated in §1798.100, §1798.105, §1798.106, §1798.110, §1798.115, and §1798.125.

[3] This feature was deprecated as Google launched the new Data Safety after we completed this work.

Table 9: Permissions coverage for APPGs.

#	APPG Name	Mode	User instruction	RC	EC
1	Iubenda	UI	●	15.6%	15.6%
3	Termly	Questionnaire	●	60.3%	60.3%
4	Privacy Policies	Questionnaire	○	85.6%	40.1%
7	Terms Feed	Questionnaire	○	52.7%	25.7%
8	Website Policies	Questionnaire	◐	60.0%	11.8%
9	Free Privacy Policy	Questionnaire	○	64.4%	25.6%

properly, despite APPGs' intrinsic design flaws. And

$$\text{Expected Coverage}\,(EC) = \frac{1}{n}\sum_{1}^{n}\frac{P_{\text{displayed}}}{P_{\text{claimed}}}$$

A higher *EC* reflects the APPG has less intrinsic design flaws about disclosing permissions in generated privacy policies.

We use the intersection of privacy policies identified by both FKS and DSC (2,042 privacy policies) because it is more certain that they are generated by one of the APPGs. Based on our observations in Section 2.3, four APPGs (#2, #5, #6, #10) do not support permission declarations, in other words, $P_{\text{recognised}} = 0$ for every app identified to use those APPGs. In addition, we find that $P_{\text{displayed}} = 0$ for four of those apps, which means that none of the developers realised that the APPG they used did not cover the permissions they claimed.

The results are shown in Table 9. First, for #1 and #3, they both can recognise all nine device permissions involved in this study, so their *RC*s are equal to *EC*s. There is a significant portion of missing permission disclosure for all APPGs, since all *EC*s are much smaller than 100%. The missing disclosure issue is not dependent on users' input, but on the design of APPGs that did not sufficiently accommodate the requirements. Second, compared to #1's and #3's consistency on *RC*s and *EC*s, for #4, #7, #8, and #9, their *EC*s are greatly smaller than *RC*s. The gap shows the importance of APPG selection, because inappropriate APPGs will hinder developers from including self-reported (claimed) permissions in their privacy policies. Third, #1 has the lowest *RC*, but it has high level user instruction; whereas #4, #7, and #9 has the relatively high *RC*s, but low level of user instruction. This counter-intuitive result indicates that current user instructions are not helpful enough to guide developers to correctly include permissions in the generated policies. In addition, #1 is the only UI-mode APPG, with the lowest *RC* (15.6%), and other comparable APPGs, whose *RC* is over 50% are all questionnaire-mode. This observation indicates that UI mode APPGs may be more challenging to be properly used by users.

Finding 3.2: A significant portion of device permissions remain under-claimed, perhaps caused by APPGs' design issues, and questionnaire mode can better guide users to include claimed device permissions compared to UI mode.

4.3 Contradiction Analysis

Previous works have discussed contradictions within a policy [34, 42]. Specifically, if affirmative and negative sentences mention the same or conflicting entities and data types in a policy, then there can be a contradiction. In this section, we compare the APPG-based and non-APPG-based privacy policies, to examine the situation of contradiction issues in both and the effect introduced by APPGs. For non-APPG-based privacy policies, we used the test set from PoliCheck [35] which contains 200 policies randomly sampled from 13K mobile apps. For APPG-based privacy policies, we use 30 policies from the "**Generated Privacy Policy Collection**" and 170 policies that are randomly sampled from the intersection list (indicated as APPG-generated by both FKS and DSC). We employ the state-of-the-art privacy policy analyser, PoliGraph [42], to conduct the following analysis.

We observed 26 contradictions in APPG-based privacy policies and 15 in non-APPG-based policies, indicating a potential tendency for APPGs to introduce more privacy policy contradiction issues. Closer examination revealed that a significant portion of these contradictions in APPG-based policies arise from conflicting statements present in different sections of the document. For example, one section of the policy explicitly states "[A condition], we sell your personal information to third parties", while an assertion in the CCPA section contradicts this by claiming "[The App] has not disclosed or sold any personal information to third parties for a business or commercial purpose in the preceding 12 months.". Therefore, it is imperative for APPG users to exercise caution and meticulously evaluate the coherence and consistency between different sections of generated privacy policies.

Finding 3.3: More privacy policy contradiction issues exist in APPG-based privacy policies.

5 Discussion and Implication

Challenges and opportunities coexist in the current APPG development. Based on our observations and study results, we summarize some findings for various roles or stakeholders in the APPG ecosystem.

App developers/APPG users. While app developers may benefit from using APPGs to create privacy policies more efficiently, they should be aware of APPGs' latent limitations. As illustrated in Section 2, APPGs have different qualities. Some do not directly target specific regulations, such as GDPR and CCPA, and some do not recognise data practices, i.e., the declaration of personal information, device permissions, and third-party services used in their mobile apps. Since app developers must make sure the privacy policies they provide are comprehensive, readable, and compliant, **it is very likely a trap** if they do not select and use appropriate APPG with careful consideration.

APPG providers. Our analysis suggests APPG providers should work on improving recognised data use, since the majority of APPGs on the market only provide a very limited scope of personal information and device permissions. The following two requirements are commonly neglected by the APPGs even though they claim to comply with CCPA: **a)** Special requirements for businesses buying, receiving, selling, or sharing personal information of 10,000,000 or more consumers in a calendar year [Regs § 999.308(c)(8), 999.317(g)(1)], and **b)** For online notices, follow generally recognized industry standards [Regs §999.308(a)(2)(d)]. Based on our market penetration analysis, we found that developers would rather use a simpler APPG (e.g., #2) than a correct one. It is worthwhile to consider this trade-off in the design and acquisition of APPGs. Our results indicate that current user instructions are not helpful enough; thus, APPG providers should work to improve user instructions and provide intuitive UIs to guide users to correctly use their tools. We have disclosed our observations and findings to APPG providers [20].

Privacy regulators. We discovered that the privacy policy of 20.1% apps could be generated by existing APPGs, which means APPGs can significantly contribute to reducing policy absence and low quality privacy policies in app markets. However, our results also highlight that the APPGs' design flaws can limit the app developers to include all essential privacy practices, data rights, and device permissions in the policies, which in turn creates a substantial risk of breaching privacy regulations. Regulators should recognize the importance of this issue and be engage with this emerging market trend. One possible approach would be to proactively provide guidance for APPGs developers, to improve compliance with privacy regulation, and to protect apps developers from potential pitfalls. Finally, we hope this study could raise broad attention to inherent design shortcomings and resulting privacy concerns of online low-code/no-code tools such as APPGs, as they are increasingly employed in practice.

6 Related Work

Privacy policy analysis. Privacy policies have been extensively analysed and discussed by the research community [33, 34, 40, 64, 71]. Wilson et al. [64] created a corpus of 115 privacy policies and 23K fine-grained data practise annotations, and revealed users' preferences on privacy policy structure and complexity. Amos et al. [33] reported that privacy compliance and readability were worse in the last 20 years, according to 130k website privacy policies. Andow et al. [34] presented *PolicyLint*, which is a privacy policy analysis tool that can identify privacy contradictions by simultaneously considering negation and varying semantic levels of data objects and entities. Bui et al. [40] proposed an automated system, dubbed *PurPliance*, that detects inconsistencies between the data-usage purposes stated in a privacy policy and the actual behaviours of an Android app.

Code-centric privacy policy auto-generation tools. Yu et al. [66, 67] developed a system named *AutoPPG*, to automatically construct readable descriptions from the source code of mobile apps, to help create privacy policies on Android. Rocky Slavin [59] developed *PoliDriod*, an Android Studio plugin that can be used to detect possible misalignments between Android API methods and privacy policies. Zimmeck et al. [70] proposed a privacy policy generator, named *PrivacyFlash*, which leverages mappings between code signatures and privacy practises expressed in policies for iOS apps. While code-centric generators might better align with apps' actual privacy behaviours compared to online APPGs, they come with significant challenges that hinder their adoption by app developers. First, they cannot ensure compliance with high-level privacy regulations, particularly non-functional requirements. Second, they present a higher entry barrier for developers due to their inherent complexity, whereas online APPGs can be easily accessed, offering a range of user instructions and publishing support as demonstrated in this study. Others [38, 65] also discussed the automated generation of privacy policies using machine learning methods.

7 Conclusion

Online Automated Privacy Policy Generators (APPGs) are broadly used by developers of mobile apps to create privacy policies to respond to regulatory requirements. This paper reports the first large-scale empirical study to comprehensively scrutinize APPGs' various features, characteristics, and extent of recognition of data use. Our market penetration analysis indicates that privacy policies of 20.1% apps could be generated by existing APPGs on the Google Play app store, and that #2 is the most popular APPG, with a 72.7% adoption rate. Our findings underline a substantial level of noncompliance with privacy laws and a frequent under-claiming of data rights and highly concerning privacy practices, especially with the most popular APPG #2. Permissions coverage analysis reveals that existing APPGs have significant issues with including all essential device permissions in the generated privacy policy, and UI mode APPGs could make it worse. Also, more contradiction issues exist in APPG-based privacy policies. In summary, for app developers, selecting and employing APPGs without careful consideration is very likely a trap, creating a substantial risk of breaching privacy regulations.

References

[1] Android developer guide. https://developer.android.com/guide/topics/permissions/overview#permission-groups, Accessed: 2022-08-26.

[2] Android developers, google play console help center. https://support.google.com/googleplay/android-developer, Accessed: 2023-09-05.

[3] App privacy details on the app store. https://developer.apple.com/app-store/app-privacy-details/, Accessed: 2022-04-30.

[4] App privacy policy. https://www.app-privacy-policy.com/, Accessed: 2022-03-28.

[5] App privacy policy generator. https://app-privacy-policy-generator.firebaseapp.com/, Accessed: 2022-03-28.

[6] Appbrain. https://www.appbrain.com/, Accessed: 2022-08-26.

[7] Australian privacy principles (APP). https://www.oaic.gov.au/privacy/australian-privacy-principles, Accessed: 2022-05-03.

[8] Beautifulsoup. https://www.crummy.com/software/BeautifulSoup/, Accessed: 2022-05-02.

[9] Brazilian general data protection law (LGPD). http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm, Retrieved: 2022-08-26.

[10] California consumer privacy act of 2018 (CCPA). https://oag.ca.gov/privacy/ccpa, Accessed: 2022-04-25.

[11] Free privacy policy. https://www.freeprivacypolicy.com/, Accessed: 2022-03-28.

[12] Geeks for geeks. https://www.geeksforgeeks.org/, Accessed: 2022-08-26.

[13] General data protection regulation (GDPR). https://gdpr-info.eu/, Retrieved: 2022-04-25.

[14] Github. https://github.com/, Accessed: 2022-08-26.

[15] google-play-scapper. https://pypi.org/project/google-play-scraper/, Accessed: 2022-04-30.

[16] Iubenda. https://www.iubenda.com/en/, Accessed: 2022-03-28.

[17] langdetect. https://pypi.org/project/langdetect/, Accessed: 2022-04-30.

[18] Lawpath. https://lawpath.com.au/, Accessed: 2022-03-28.

[19] Leading apps wikipedia. https://en.wikipedia.org/wiki/List_of_most-downloaded_Google_Play_applications, Accessed: 2022-08-26.

[20] Our dataset and repository. https://anonymous.4open.science/r/APPG-FF1C, Accessed: 2023-09-03.

[21] Personal information protection law of the people's republic of china (PIPL). http://www.npc.gov.cn/npc/c30834/202108/a8c4e3672c74491a80b53a172bb753fe.shtml, Accessed: 2022-04-25.

[22] Privacy policies. https://www.privacypolicies.com/, Accessed: 2022-03-28.

[23] The privacy policy of "bway app". http://www.e-droid.net/privacy.php?ida=1702882&idl=en, Accessed: 2022-03-28.

[24] The privacy policy of "easy communication". http://natarajanraman.in/?page_id=67, Accessed: 2022-03-28.

[25] Privacy policy online. https://www.privacypoliciesonline.com/, Accessed: 2022-03-28.

[26] Selenium. https://www.selenium.dev/, Accessed: 2022-04-28.

[27] Stack overflow. https://stackoverflow.com/, Accessed: 2022-08-26.

[28] Termly. https://termly.io/, Accessed: 2022-03-28.

[29] Terms feed. https://app.termsfeed.com/, Accessed: 2022-03-28.

[30] Web content accessibility guidelines (wcag) 2.1. https://www.w3.org/TR/WCAG21/, Accessed: 2023-06-07.

[31] Website policies. https://www.websitepolicies.com/, Accessed: 2022-03-28.

[32] Kevin Allix, Tegawendé F. Bissyandé, Jacques Klein, and Yves Le Traon. Androzoo: Collecting millions of android apps for the research community. In *Proceedings of the 13th International Conference on Mining Software Repositories*, MSR '16, pages 468–471, New York, NY, USA, 2016. ACM.

[33] Ryan Amos, Gunes Acar, Elena Lucherini, Mihir Kshirsagar, Arvind Narayanan, and Jonathan Mayer. Privacy policies over time: Curation and analysis of a million-document dataset. In *Proceedings of the Web Conference 2021*, pages 2165–2176, 2021.

[34] 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, 2019.

[35] 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, 2020.

[36] Rebecca Balebako and Lorrie Cranor. Improving app privacy: Nudging app developers to protect user privacy. *IEEE Security & Privacy*, 12(4):55–58, 2014.

[37] Rebecca Balebako, Abigail Marsh, Jialiu Lin, Jason I Hong, and Lorrie Faith Cranor. The privacy and security behaviors of smartphone app developers. 2014.

[38] Nastaran Bateni. *Automated Methods for Comparison and Generation of Privacy Policies*. PhD thesis, University of Guelph, 2021.

[39] David Bawden and Lyn Robinson. The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of information science*, 35(2):180–191, 2009.

[40] 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*, pages 2824–2843, 2021.

[41] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.

[42] Hao Cui, Rahmadi Trimananda, Athina Markopoulou, and Scott Jordan. {PoliGraph}: Automated privacy policy analysis using knowledge graphs. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 1037–1054, 2023.

[43] Sarra Habchi, Geoffrey Hecht, Romain Rouvoy, and Naouel Moha. Code smells in ios apps: How do they compare to android? In *2017 IEEE/ACM 4th International Conference on Mobile Software Engineering and Systems (MOBILESoft)*, pages 110–121. IEEE, 2017.

[44] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, Naval Technical Training Command Millington TN Research Branch, 1975.

[45] Konrad Kollnig, Anastasia Shuba, Reuben Binns, Max Van Kleek, and Nigel Shadbolt. Are iphones really better for privacy? comparative study of ios and android apps. *arXiv preprint arXiv:2109.13722*, 2021.

[46] Barbara Krumay and Jennifer Klar. Readability of privacy policies. In *IFIP Annual Conference on Data and Applications Security and Privacy*, pages 388–399. Springer, 2020.

[47] Renuka Kumar, Apurva Virkud, Ram Sundara Raman, Atul Prakash, and Roya Ensafi. A large-scale investigation into geodifferences in mobile apps. In *31st USENIX Security Symposium (USENIX Security 22)*, Boston, MA, August 2022. USENIX Association.

[48] Stephen E Levy and Carl Gutwin. Improving understanding of website privacy policies with fine-grained policy anchors. In *Proceedings of the 14th international conference on World Wide Web*, pages 480–488, 2005.

[49] 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*, pages 1–24, 2022.

[50] Frederick Liu, Shomir Wilson, Peter Story, Sebastian Zimmeck, and Norman Sadeh. Towards automatic classification of privacy policy text. *School of Computer Science Carnegie Mellon University*, 2018.

[51] Shuang Liu, Baiyang Zhao, Renjie Guo, Guozhu Meng, Fan Zhang, and Meishan Zhang. Have you been properly notified? automatic compliance analysis of privacy policy text with gdpr article 13. In *Proceedings of the Web Conference 2021*, pages 2154–2164, 2021.

[52] Razieh Nokhbeh Zaeem and K Suzanne Barber. A large publicly available corpus of website privacy policies based on dmoz. In *Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy*, pages 143–148, 2021.

[53] Marten Oltrogge, Erik Derr, Christian Stransky, Yasemin Acar, Sascha Fahl, Christian Rossow, Giancarlo Pellegrino, Sven Bugiel, and Michael Backes. The rise of the citizen developer: Assessing the security impact of online app generators. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 634–647. IEEE, 2018.

[54] Shidong Pan, Thong Hoang, Dawen Zhang, Zhenchang Xing, Xiwei Xu, Qinghua Lu, and Mark Staples. Toward the cure of privacy policy reading phobia: Automated generation of privacy nutrition labels from privacy policies. *arXiv preprint arXiv:2306.10923*, 2023.

[55] Shidong Pan, Zhen Tao, Thong Hoang, Dawen Zhang, Zhenchang Xing, Xiwei Xu, Mark Staples, and David Lo. Seeprivacy: Automated contextual privacy policy generation for mobile applications. *arXiv preprint arXiv:2307.01691*, 2023.

[56] Ellen Poplavska, Thomas B Norton, Shormir Wilson, and Norman Sadeh. From prescription to description: Mapping the gdpr to a privacy policy corpus annotation scheme. In *Legal Knowledge and Information Systems-JURIX 2020: 33rd Annual Conference*, 2020.

[57] Anand Rajaraman and Jeffrey David Ullman. *Mining of massive datasets*. Cambridge University Press, 2011.

[58] 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, 2015.

[59] Rocky Slavin. Polidroid-as. https://github.com/rslavin/PoliDroid-AS, Accessed: 2022-05-05.

[60] Anne Stopper and Jen Caltrider. See no evil: Loopholes in google's data safety labels keep companies in the clear and consumers in the dark. https://foundation.mozilla.org/en/campaigns/googles-data-safety-labels/, Accessed: 2023-09-03.

[61] Ruoxi Sun and Minhui Xue. Quality assessment of online automated privacy policy generators: an empirical study. In *Proceedings of the Evaluation and Assessment in Software Engineering*, pages 270–275. 2020.

[62] W Gregory Voss. Looking at european union data protection law reform through a different prism: The proposed eu general data protection regulation two years later. *Journal of Internet Law*, 17(9), 2014.

[63] Ari Ezra Waldman. Cognitive biases, dark patterns, and the 'privacy paradox'. *Current opinion in psychology*, 31:105–109, 2020.

[64] 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, et al. The creation and analysis of a website privacy policy corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1330–1340, 2016.

[65] 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 *CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2022.

[66] Le Yu, Xiapu Luo, Jiachi Chen, Hao Zhou, Tao Zhang, Henry Chang, and Hareton KN Leung. Ppchecker: Towards accessing the trustworthiness of android apps' privacy policies. *IEEE Transactions on Software Engineering*, 47(2):221–242, 2018.

[67] Le Yu, Tao Zhang, Xiapu Luo, and Lei Xue. Autoppg: Towards automatic generation of privacy policy for android applications. In *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 39–50, 2015.

[68] Shikun Zhang, Yuanyuan Feng, Yaxing Yao, Lorrie Faith Cranor, and Norman Sadeh. How usable are ios app privacy labels? *UMBC Faculty Collection*, 2022.

[69] Shikun Zhang and Norman Sadeh. Do privacy labels answer users' privacy questions?

[70] Sebastian Zimmeck, Rafael Goldstein, and David Baraka. Privacyflash pro: automating privacy policy generation for mobile apps. In *28th Network and Distributed System Security Symposium (NDSS 2021). NDSS*, 2021.

[71] Sebastian Zimmeck, Peter Story, Daniel Smullen, Abhilasha Ravichander, Ziqi Wang, Joel R Reidenberg, N Cameron Russell, and Norman Sadeh. Maps: Scaling privacy compliance analysis to a million apps. *Proc. Priv. Enhancing Tech.*, 2019:66, 2019.

Appendix

Readability Analysis

The readability score of privacy policies are calculated by the Flesch Reading-ease Test [44]:

$$206.835 - 1.015\left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6\left(\frac{\text{total syllables}}{\text{total words}}\right)$$

Specific readability scores are listed in the Table 10.

Table 10: Comparison of APPGs' readability scores.

#	APPG Name	Average Readability Score
1	Iubenda	37.3
2	App Privacy Policy Generator	**49.3**
3	Termly	46.0
4	Privacy Policies	37.7
5	App Privacy Policy	43.9
6	Privacy Policy Online	44.0
7	Terms Feed	37.7
8	Website Policies	36.7
9	Free Privacy Policy	37.7
10	Lawpath	41.0
/	12 leading apps	46.0

GDPR and CCPA requirements compliance

The specific requirements and compliance checking for GDPR and CCPA are tallied in Appendix Table 11 and Table 12.

Table 11: GDPR privacy policy requirements and APPGs' compliance in May 2022. Numbers indicate APPGs as per Table 1.

GDPR requirement	1	2	3	4	5	6	7	8	9	10
Identity and contact details of the data controller and their representative, if any [Art. 13(1)(a), 14(1)(a)]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗
Purposes of the processing for which the personal data are intended and legal basis for the processing [Art. 13(1)(c), 14(1)(c)]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✓
Categories of personal data concerned [Art. 14(1)(d)]	✓	✓	✓	✓	✗	✗	✓	✓	✓	✗
Recipients or categories of recipients of the personal data [Art. 13(1)(e), 14(1)(e)]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✓
Period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period [Art. 13(2)(a), 14(2)(a)]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗
Existence of the rights to request data access, rectification, erasure, and data portability as well as the rights to restrict and object to processing [Art. 13(2)(b), 14(2)(c)]	✓	✓	✓	✓	✗	✗	✓	✓	✓	✗
Right to withdraw consent for processing at any time [Art. 13(2)(c), 14(2)(d)]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗
Right to lodge complaint with a supervisory authority [Art. 13(2)(d), 14(2)(e)]	✓	✓	✓	✓	✗	✗	✓	✓	✓	✗

Table 12: CCPA privacy policy requirements and APPGs' compliance in May 2022. Numbers indicate APPGs as per Table 1.

CCPA requirement	1	2	3	4	5	6	7	8	9	10
Disclosure of right to request how personal information is collected, used, sold, disclosed for a business purpose, and shared [CCPA §1798.130(a)(5)(A), 1798.110(a), 1798.115(a), Regs § 999.308(c)(1)(a)]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✓
Disclosure of right to request deletion of personal information [CCPA §1798.105(b), 1798.130(a)(5)(A), Regs §999.308(c)(2)(a)]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗
Disclosure of whether personal information is sold and right to opt-out of sale [Regs §999.308(c)(3)(a), 999.308(c)(3)(b), 999.306]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗
Disclosure of right to not be discriminated against when requesting any rights [CCPA §1798.130(a)(5)(A), 1798.125(a), Regs §999.308(c)(4)(a)]	✗	✗	✓	✓	✗	✗	✓	✓	✓	✗
Instructions for submitting requests and link to online form or portal if offered [Regs §999.308(c)(1)(b), 999.308(c)(2)(b), 999.308(c)(2)(c)]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗
Instructions for authorized agents to make requests [Regs §999.308(c)(5)(a)]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗
Description of the process used to verify requests [Regs §999.308(c)(1)(c)]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗
List of categories of personal information collected in preceding 12 months [CCPA §1798.130(a)(5)(B), 1798.110(c), Regs §999.308(c)(1)(d)]	✗	✗	✓	✓	✗	✗	✓	✓	✓	✗
List of categories of personal information sold in preceding 12 months [CCPA §1798.130(a)(5)(C), 1798.115(c)(1), Regs §999.308(c)(1)(g)(1)]	✓	✗	✓	✓	✗	✗	✓	✗	✓	✗
List of categories of personal information disclosed for business purpose in preceding 12 months [CCPA §1798.130(a)(5)(C), 1798.115(c)(2), Regs §999.308(c)(1)(g)(1)]	✓	✗	✓	✓	✗	✗	✓	✗	✓	✗
For each personal information category, categories of third parties to whom information was disclosed or sold [Regs §999.308(c)(1)(g)(2)]	✓	✗	✓	✓	✗	✗	✓	✗	✓	✗
Categories of sources from which personal information is collected [Regs §999.308(c)(1)(e)]	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗
Business or commercial purpose for collecting or selling personal information [Regs §999.308(c)(1)(f)]	✓	✗	✓	✓	✗	✗	✓	✗	✓	✓
Whether the business has actual knowledge that it sells personal information of minors under 16 years of age and special process [Regs §999.308(c)(1)(g)(3), 999.308(c)(9)]	✓	✓	✗	✓	✓	✓	✓	✗	✓	✗
Contact information for questions or concerns [Regs §999.308(c)(6)(a)]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
Date policy was last updated [Regs §999.308(c)(7)]	✓	✓	✓	✓	✗	✗	✓	✓	✓	✗
Special requirements for businesses buying, receiving, selling, or sharing personal information of 10,000,000 or more consumers in a calendar year [Regs §999.308(c)(8), 999.317(g)(1)]	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
For online notices, follow generally recognized industry standards, such as the W3C Web Content Accessibility Guidelines, version 2.1 of June 5, 2018 [Regs §999.308(a)(2)(d)]	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗