A Decade of Privacy-Relevant Android App Reviews: Large Scale Trends

Omer Akgul
University of Maryland

Sai Teja Peddinti
Google

Nina Taft
Google

Michelle L. Mazurek
University of Maryland

Hamza Harkous
Google

Animesh Srivastava
Google

Benoit Seguin
Google

Abstract

We present an analysis of 12 million instances of privacy-relevant reviews publicly visible on the Google Play Store that span a 10 year period. By leveraging state of the art NLP techniques, we examine what users have been writing about privacy along multiple dimensions: time, countries, app types, diverse privacy topics, and even across a spectrum of emotions. We find consistent growth of privacy-relevant reviews, and explore topics that are trending (such as Data Deletion and Data Theft), as well as those on the decline (such as privacy-relevant reviews on sensitive permissions). We find that although privacy reviews come from more than 200 countries, 33 countries provide 90% of privacy reviews. We conduct a comparison across countries by examining the distribution of privacy topics a country's users write about, and find that geographic proximity is not a reliable indicator that nearby countries have similar privacy perspectives. We uncover some countries with unique patterns and explore those herein. Surprisingly, we uncover that it is not uncommon for reviews that discuss privacy to be positive (32%); many users express pleasure about privacy features within apps or privacy-focused apps. We also uncover some unexpected behaviors, such as the use of reviews to deliver privacy disclaimers to developers. Finally, we demonstrate the value of analyzing app reviews with our approach as a complement to existing methods for understanding users' perspectives about privacy.

1 Introduction

User perspectives are commonly measured through user studies (e.g., surveys, interviews, lab studies), which can provide rich data to answer focused research questions. Unfortunately, such studies do not scale beyond thousands of users, and the resulting measurements are heavily bound to the environment the studies were conducted in: user opinions may vary over time, between regions, and across different app types. Rerunning user studies to understand these differences can be prohibitively costly, both monetarily and time-wise.

In this work, we present an alternate and complementary analysis approach, with a different set of research tradeoffs. Specifically, we trade the power of surveys in having the same (targeted) questions answered by many participants for the power of ecologically valid, large-scale analysis in uncovering unanticipated insights from open-ended reviews. We believe these insights can, in turn, motivate future investigations of user-centered privacy, including survey-based studies.

Getting input about users' perspectives on privacy issues from millions of user reviews from hundreds of countries has not been feasible until recently. Advances in natural language processing (NLP) and large language models (LLMs) permit complex analysis of enormous text corpora. Here, we leverage these advances to investigate users' privacy opinions and concerns from a novel perspective: we study 12.3M (million) privacy-related reviews, extracted from $\sim$2B (billion) public reviews on Google Play spanning 10 years (Jan. 2013–Feb. 2023). These privacy reviews come from more than 200 countries or regions, 25 languages, and 160K (thousand) apps that span every Play app category.

To analyze this extensive dataset, we leverage and extend NLP techniques recently introduced in Hark [29] to automatically extract all reviews that discuss a privacy topic, assign fine-grained issue tags to each review, aggregate related issues into larger thematic clusters, and classify the "emotions" expressed in these reviews. The resulting dataset of roughly 12.3M privacy-related reviews likely constitutes the largest body of privacy feedback ever evaluated at this granularity. Using this dataset, we address the following research questions:

- **RQ1:** Which privacy issues do users raise and discuss in app reviews?
- **RQ2:** How have these privacy issues evolved over time?
- **RQ3:** How do privacy issues vary across countries?
- **RQ4:** Which types of apps have privacy reviews with strongly negative or strongly positive emotions?
- **RQ5:** How do reviews as a source of understanding privacy perspectives complement prior work?

We find that privacy reviews have grown steadily over 10 years, both in terms of absolute volume (a 4.7x increase) and when normalized for review volume (9% biannual growth). We find that themes such as `Data Deletion` are growing in importance, while reviews relating to `Excessive [Privacy-Relevant] Permissions` (a popular research topic, e.g., [11, 22, 57, 77]) have seen a significant decline.

Our broad overview across the globe also shows that geographic proximity is a weak indicator of whether nearby countries discuss similar privacy issues. We find that the countries that contribute the largest volume of privacy related reviews tend to be countries with large populations, and not necessarily countries that drive privacy regulation (such as the EU). We uncover a handful of countries (typically understudied) that discuss unique distributions of privacy topics. For example, in Türkiye, we find a significant number of reviews using almost the same quasi-legal disclaimers—implying users might be under the mistaken assumption that this bolsters privacy protection.

The app types whose privacy reviews exhibit the most strongly negative emotions (e.g. anger, annoyance, fear) are social media apps, parental control and child monitoring apps, as well as simulation games that mimic users facial and vocal expressions (leading to anxiety about surveillance). On the other hand, reviews for security and privacy apps commonly express positive emotions.

Finally, we place our results in context of related research, adding context to prior findings and identifying new areas for further study. For example, we expand on previously documented privacy concerns related to loan apps [50], finding that these concerns arise in a number countries beyond the original one identified. We also observe strongly positive reviews of many apps that claim to secure or hide content on phones, especially in the context of multi-user devices [68], raising a number of questions for future research. We demonstrate the utility of automatically distilling user feedback at scale, as a complement to other methods of understanding users' opinions and concerns.

2 Background and Related Work

Interviews and surveys are the most common method of measuring privacy attitudes. They have been broadly used to understand users' mental models of security & privacy tools [2, 8, 42], to study privacy preferences for smartphone app permissions [11, 66], to measure privacy concerns with IoT and sensors [47, 60, 73, 78], and used in methods studies [1, 64, 72]—to name only a few in this vast field. While these methods are sound and routinely used, researchers also acknowledge limitations arising from social desirability, acquiescence, and demand biases. While behavioral measurements can circumvent some of these biases [9, 20], they depend on inferring what users think rather than being able to document exactly what users' attitudes are. Moreover, recent work presents evidence that problems with validity of privacy constructs (built from surveys) may be widespread [15]. Similar methods are used to measure developers' mental models of privacy threats [48], their understanding of app privacy [48], and their responsiveness to privacy nudges [71]. Studying developers themselves is out of scope of this work.

A key difference in our work is our method: namely, recently proposed LLM techniques, specific to the app reviews context [29], enabling us to analyze inputs from millions of users. This method, complementary to those above, offers a different set of tradeoffs. While interviews allow for in-depth questioning of what users think, they typically do not scale beyond a few tens of users. Surveys can scale to thousands (but not millions) of users and enable the development of verifiable privacy scales, but are very focused; a researcher cannot learn about an issue that wasn't posed in the questionnaire. Further, recruitment across multiple geographic regions is expensive and challenging, making such studies often costly to repeat. In contrast, our approach does not support controlled experiments or hypothetical questions about potential new designs. Still, it does allow for increased ecological validity and much larger and broader samples.

In this paper, we explore what users around the globe write about privacy in Android app reviews. A few works have explored app reviews [29, 53, 54]; however, they extracted a limited set of privacy reviews because their privacy classifiers relied on keywords [54], limited heuristics for data sampling [53], or a limited privacy taxonomy [29]. These works also only included English language reviews, whereas we include 25 languages. Our set of privacy-related reviews is 1000x, 20x and 2x larger (respectively) than these earlier studies.

Many cross-country studies focus on specific aspects of privacy, such as android permissions [11], social networks [74], phone locking behavior [28], incident response [63], or how much users are willing to pay for privacy for specific types of data [59]. Beyond these focused studies, some more general cross-country studies attempt to understand the influence of factors such as culture or country on privacy attitudes or privacy preferences [14, 26]. Some studies do show differences between non-Western and Western countries in terms of misconceptions around privacy [33] practices.

A recent survey [30] notes limited geographic diversity in usable security & privacy research, with participants primarily from Western, educated, industrialized, rich, and democratic (WEIRD) societies. While this may naturally occur due to geographic and linguistic barriers, our approach—using text from >200 countries/regions and translations for 24 languages—offers an alternative. Most of the multinational surveys above include 3-7 countries; a few include 10-20. To our knowledge, our work is the first to report on data from more than 200 countries, with the top-50 explicitly compared.

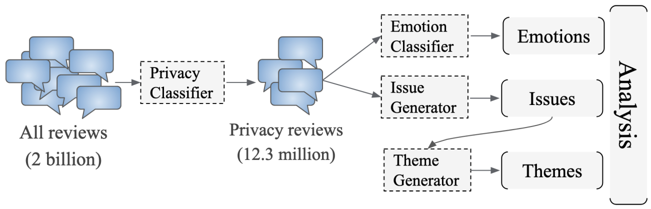

Figure 1: Our analysis pipeline overview, adapted from [29].

3 Data Analysis Pipeline

In this section, we describe our data-analysis pipeline that reuses several lessons from the Hark system [29], highlighting the modifications we made to fit the purpose of this study. We also discuss the resulting dataset and our analysis approach.

3.1 Text Analysis Pipeline

We build our analysis pipeline to leverage and extend the components in Hark, an end-to-end system for retrieval and analysis of privacy-related feedback leveraging state-of-the-art techniques in NLP [29]. An overview of our data-analysis pipeline is in Figure 1. First, the *privacy classifier* identifies the privacy-related feedback from unstructured text. The *issue generation* model takes in this privacy feedback and dynamically generates meaningful, fine-grained issues (covering both known and newly emerging issues) describing the privacy aspects discussed within each review text. The *theme creation* component groups these issues into thematic clusters and assigns a succinct title to each. Our data-analysis pipeline also includes an *emotion* classifier that dissects each review's text across 28 emotions (e.g., anger, fear, joy and confusion).

We obtained models and training data from Hark [29] and re-purposed most of the pipeline, with modifications to the privacy classifier to improve the breadth of topics identified. First, we expand the privacy taxonomy from the original paper (25 concepts) to include 89 privacy-relevant concepts.[1] The prior taxonomy already included concepts from multiple known taxonomies. However, when we asked five privacy experts (with at least 5 years of academic/industry experience in privacy research and development) to manually examine the original taxonomy [29], they observed many missing topics. For example, we added Data Hiding, Opt-out and Location concepts, and added specificity to the Anonymity and Personal Data types.

Second, we rely on these same experts, instead of crowdsourcing, to manually annotate a new training dataset for the privacy classifier, based on the expanded taxonomy and following the same *Natural Language Inference* approach and heuristics as in [29]. To best use the experts' time and generate high-quality training data without requiring multiple annotations per review, we used two labeling rounds. In the first round, we described the taxonomy to the experts and each annotator independently labeled a subset of reviews. Following the principles of *Active Learning*, we trained a classifier on the annotated reviews from the first round. The active-learning classifier is composed of Sentence-T5 frozen embeddings [55], a dense layer, and a binary classification head. The experts discussed cases where the model produced a different label from the expert as well as cases whose classification probabilities had high entropy. In the second round, all experts discussed these misclassifications or low-confidence predictions, thereby focusing their efforts on challenging examples. Overall, we generated a labeled dataset of 4.3K (nearly equal split of 2K *privacy* and 2.3K *not-privacy*) reviews, which we split into training, validation, and test sets. We fine-tuned T5-11B [61] on these datasets and use it as our privacy classifier.

Model evaluation: Our classifier has a 0.95 ROC AUC, 87% precision, and 86% recall on our new diversified test set,[2] and a similar ROC AUC of 0.88 (vs 0.92) on the original Hark test set [29]. In comparison, the privacy classifier from [29] has 0.87 ROC AUC, 89% precision, and 51% recall on our new test set, demonstrating that our classifier performs better at capturing the diversity of our taxonomy. (For additional experiments with a range of model architectures see Appendix B.) We also performed qualitative assessment of all 50 false positives/negatives of our classifier, noticing primarily issues with reviews that were short ($\leq$ 10 words) and ambiguous (missing context, causing multiple possible interpretations). For example, a short review "*Taking too much data*" was labeled as 'privacy,' when the privacy expert (based on other reviews seen) interpreted it to be a complaint about the app consuming the limited mobile data bandwidth and annotated it as 'not-privacy'; whereas an ambiguous review "*msgs have been deleted, but for some reason they remain in place*" was labeled as 'not-privacy' (could be seen as a bug in app functionality), when the privacy expert interpreted it as data deletion control not working as expected (a privacy concern). 13 of the 50 false positives/negatives were short; and the rest (though longer) were still ambiguous. Despite missing these short and ambiguous cases, our classifier had good performance.

Similar to [29], for ease of representation, we consolidated the 28 classifier-generated emotions into 8 emotions groups (plus a neutral option) based on Ekman's emotions taxonomy [21] and using Demszky et al.'s [17] grouping criteria.

3.2 Data Description

We obtained access to reviews directly from the Google Play team. Our initial dataset consists of all ~2B publicly visible Google Play app reviews for apps with >10K installs,

[1] See the extended paper for the expanded privacy taxonomy [3].

[2] We set a 0.8 prediction score threshold for best accuracy and verified results qualitatively.

spanning 10 years from Jan 2013–Feb 2023. The dataset was already anonymized (no user identifying information) and sanitized (any detected fake/spam reviews were removed). Each review is associated with the review text, its language, country, submission time, star rating, the corresponding app's package name, the app's developer-specified category information and (if available) finer-grained *app-type* information (such as 'Rideshare & Taxis' for a rideshare app, which belongs to the 'Maps & Navigation' developer-specified category). Note that we do not generally use developer-specified categories as they are too broad [57]. Instead we use the more specific app types that are displayed on Play when viewing apps.

In our dataset, ~65% of reviews are non-English. Since our classifiers work on English texts, we leverage Google's Translation API to translate reviews in 24 non-English languages to English (see Appendix A for the full list) [5]. The final set of English reviews (including translations) constitutes 98% of the initial dataset and contains 1.9B reviews for 160K apps. These reviews come from all developer-specified categories and 445 app types, with representation from more than 200 countries and territories in the world.

Applying our data-analysis pipeline to this large review dataset, we identified 12.3M privacy-related reviews, which were organized into 227 themes with at least 5K reviews. To our knowledge, this constitutes the largest privacy-review dataset ever evaluated. In the following sections, we analyze these privacy reviews in depth across dimensions.

3.3 Metrics

Hereafter, we rely on two core privacy metrics. We first define the percentage of privacy reviews (PPR):

$$PPR = \frac{Number\ of\ privacy\ reviews * 100}{Total\ number\ of\ reviews}$$

When we compute PPR over *all* 12.3M privacy reviews, we notate it as $PPR_{(all)}$. We compute $PPR_{(country)}$, where both the numerator and denominator are limited to one country. Similarly, we use $PPR_{(app-type)}$ for the fraction of reviews within a particular app type that discuss privacy.

PPR is not applicable to themes, only privacy reviews have themes. To denote the ratio of a theme across all privacy reviews, we define percentage of theme privacy reviews (PTPR):

$$PTPR = \frac{Number\ of\ privacy\ reviews\ in\ a\ theme * 100}{Total\ number\ of\ privacy\ reviews}$$

$PTPR_{(theme)}$ is often used to denote the theme of interest.

3.4 Analysis Techniques

We visualize data and report descriptive statistics to make observations. We use simple hypothesis tests, clustering, and regressions to examine trends in the underlying data.

To understand if various subsets of privacy reviews significantly increase or decrease over time, we first perform a KPSS (Kwiatkowski–Phillips–Schmidt–Shin) test to check if the PPR/PTPR time series (30 day intervals) is stationary over the long term [43].We then fit a linear regression for each non-stationary item (e.g., theme), to check if the slope estimate is statistically significant (i.e., PPR/PTPR changes over time). Items that do not meet these requirements are assumed to not consistently trend. Further, we calculate the average two-year change in PTPR/PPR over the 10 years with a sliding window stride of 14 days.[3] Average two-year change mirrors regression results for every time-series analysis we conducted, and all items with significant KPSS values also had significant slope estimates. Thus, for increased clarity, we only report average two-year change.

When elaborating on subsets of reviews (mostly obtained through the intersection of themes, country, time, and app type), we quantify the most prominent issues and qualitatively assess a random sample of reviews to confirm. Our goal is to use examples to paint a more detailed picture of the issues and themes our data-analysis pipeline generates.

3.5 Ethical considerations

App reviews are public (and are denoted as such during submission[4]), and users submit reviews with the intention of sharing their views with other potential users. However, we take additional precautions before the review data is analyzed. First, all user identifiers (such as account IDs, emails, device information, etc.) are removed from the reviews dataset. Second, only reviews from apps with at least 10K installs and at least 1K reviews are included. Third, inline with recent suggestions, we paraphrase all quotes reported in this paper [40]. Additionally, we do not disclose app package names to prevent user deanonymization when joined with other sources. Our dataset did not leave Google's premises, ensuring compliance with Google's terms of service.

Our work uncovered reviews that contained ads for spying services. We disclosed our findings to Google Play and they removed these harmful reviews before publication.

3.6 Limitations

Our study is an observational one, i.e., we don't control who leaves privacy reviews when. The resulting selection bias is common in other work [49, 67]; however, the bias we observe in this large-scale study of real-world data is likely different from the biases common in survey or interview studies, offering a different view of similar research questions.

[3] We considered annual change but found two years to be more fluctuation resistant. Note that we do not report the compounded rates.

[4] https://web.archive.org/web/20230515141330/https://play.google.com/about/comment-posting-policy/?hl=en-US

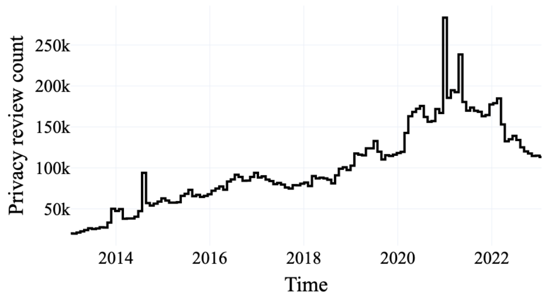

Figure 2: Privacy review counts per 30-day intervals.

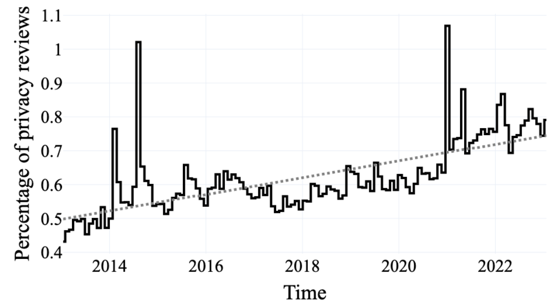

Figure 3: Percentage of privacy reviews ($PPR_{(all)}$) per 30-day intervals. Dotted line is the fitted regression.

Each review, taken by itself, is specific to the application it is left on. This might mean our dataset is too specific and not generalizable. We argue that the sheer number of apps, and observed similarities within large groups of apps, helps to smooth out this effect, and thus does reveal common themes among apps and countries. In isolation, feedback is specific; collectively, stories emerge. We rely on Google Translate APIs to translate non-English reviews. This API has been thoroughly evaluated, even on low-resource (and low-volume) languages, and has F1 quality scores of $>97\%$ for all languages with $>$2M native speakers (see Appendix E in [5]). Nonetheless, any translation errors might influence our results, though we believe the impact is minimal since we focus on 25 widely spoken languages.

Our work focuses on Android users; however, other sizeable platforms exist, raising generalizability concerns depending on the ratio of android users in a country. We note that Android is the most popular mobile operating system in the world and tends to be even more popular in non-WEIRD countries.[5] Coincidentally, these countries receive the least security & privacy research attention, some of which appear in our dataset but have received no prior privacy-focused academic interest [30], making our work instrumental in addressing this gap. We also note that, recent work has not found differences between Android and iOS users' privacy sensitivities [1], suggesting our results may provide hints beyond Android users. We leave the exploration of this hypothesis to future work.

4 Aggregate Growth

We start our dive into the 12.3M privacy reviews with an initial look at the aggregate data over the last decade. We first examine how privacy-related reviews have grown over time. The absolute volume of privacy reviews, in 30-day intervals, is shown in Figure 2. Figure 3 shows $PPR_{(all)}$ using the same intervals. Both figures trend upward over time, albeit non-uniformly, indicating that privacy reviews have increased over the last 10 years. Fitting a regression line on

[5]https://worldpopulationreview.com/country-rankings/iphone-market-share-by-country

Date	Event
February 2014	Privacy reaction to the sale of a messaging app (link)
February 2014	Surge of a “secure” messaging app (link)
February 2014	Proliferation of Talking Angela hoax (link)
August 2014	Large company forcing messenger download (link)
Jan 2021	Surge of another “secure” messaging app (link)
Jan 2021	Surge of privacy-focused search engine (link)
Jan & May 2021	Large messaging app privacy policy change (link)

Table 1: Widely discussed privacy events

$PPR_{(all)}$shows a significant increase ($R^2 = 0.44$, $p < 0.001$, KPSS $p \leq 0.05$). $PPR_{(all)}$grows from 0.5% to 0.8% (4.7$\times$ increase in volume), a 9.3% relative growth every 2 years on average. To contextualize this growth in privacy reviews, we also fit a regression line on the absolute volume of all (both privacy and not-privacy) reviews in 30-day intervals, and see that reviews have generally increased ($R^2 = 0.71$, $p < 0.001$, KPSS $p \leq 0.05$) during the same period from 6.5M to 24.5M per month (3.7$\times$ increase in volume). This shows privacy review volume is increasing faster than overall review volume.

Figure 3 exhibits notable spikes in February 2014, August 2014, January 2021, and May 2021. Analyzing reviews during these periods, we identified a few apps contributing to these spikes. These apps had well-publicized events (Table 1) that heightened user privacy concerns or increased privacy awareness, resulting in the apps seeing a 2$\times$ to 25$\times$ increase in privacy review volume. Reviews of other themes and app types stay relatively stable (see Figure 6).

Finding 1: Over the last decade, global privacy reviews have increased in absolute numbers and in PPR, exhibiting a biannual relative growth rate in PPR of 9%.

5 Trends in Privacy Themes

We now focus on our first research question (RQ1), asking which privacy themes are raised in reviews. Table 2 shows the top 20 themes by volume. We indicate whether the trend is generally increasing, decreasing or staying the same, and briefly summarize the review content. The last column states

Privacy Theme	**Trend**	**Reviews**	**Short Summary**	**# of Country Top-5 appear.**
`Data Deletion`	⇑	893K	Data deletion requests, data misuse, inability to delete data.	47
`Privacy Concerns`	⇝	541K	Vague privacy protection and concerns.	36
`Data Theft`	⇑	450K	Data stealing, sharing, and leakage; unauthorized data access	21
`Password Protection`	↓	367K	Password protection for apps and user data.	18
`Call Recording`	↑	331K	Recording of phone call conversations.	20
`Fingerprint Matters`	⇝	325K	Fingerprint scanners and their handling of the bio-metrics.	17
`Excessive Permissions`	⇓	322K	Asking for excessive privacy-sensitive permissions.	12
`Personal Information Privacy`	⇝	315K	Personal information access/usage.	4
`Location Access Concerns`	⇝	313K	Location data collection and sharing.	20
`Unneeded Camera Access`	↓	311K	App accessing camera/microphone without permission.	15
`Unneeded Access`	↓	224K	Requesting privacy-sensitive permissions (e.g., contacts).	7
`Content Hiding`	⇓	223K	Efficacy discussion of content hiding features.	11
`Unauthorized Account Access`	↑	220K	Unauthorized access to accounts, mobile devices, ...	-
`Spying Concerns`	↓	216K	Games mimicking users' considered spying. Surveillance.	5
`App Locking`	↓	205K	Privacy protections provided by app locking.	3
`Unwanted Data Collection`	⇝	200K	App collects data and sells them.	2
`Tracking Concerns`	⇝	190K	Tracking users phone number, location, and activity.	2
`Data Usage Concerns`	⇝	142K	Unwanted data usage patterns.	3
`Chats Privacy`	⇝	138K	Admiration for or the need to have private chat feature.	-
`Information Privacy`	⇝	138K	App is not upfront about its functionality.	-

Table 2: Top 20 themes in privacy reviews over the last 10 years. ⇑ / ⇓ indicates average PTPR change $> 1\%$ in two years. ↑ / ↓ indicates significant change (KPSS test $p \leq 0.05$). ⇝ indicates no detectable significant change (KPSS test $p > 0.05$).

the number of countries where this theme appears among the top 5 themes, capturing how widespread an issue is globally.

We see that Data Deletion is the top theme worldwide, that it has been increasing over the last decade, and that it is a top issue in 47 countries. We note that the top 3 privacy themes constitute 16% of all privacy reviews, the top 50 constitute 65%, and the top 200 cover 85%. From a volume perspective, this indicates that rather than a small number of dominant themes, we see a broad set of privacy topics raised across the Play store. However, from a geographic perspective, only 10 themes are a top issue in more than 10 countries (Table 2). This indicates that some themes, even if quite voluminous, may arise in a limited set of countries.

To address our second research question (RQ2), on how privacy issues evolve over time, we first determine which themes are either decreasing or increasing in significant way. In Figure 4 we plot the average 2-year change for themes among the top 20 that experience statistically significant change over the last decade (KPSS $p \leq 0.05$); stationary (non-changing) themes are excluded. The temporal evolution of five sample themes is shown in Figure 6. In Sections §5.1, §5.2, and §5.3, we continue to address RQ1 and RQ2 by focusing on the specific themes incurring long-term changes.

5.1 Themes Decreasing in Prevalence

Themes about privacy-sensitive permissions : Among the decreasing themes, (see Figure 4, left-most column), we find several themes related to privacy-sensitive permissions, namely `Excessive Permissions, Unneeded Camera Access, Unneeded [Permission] Access`. We include `Location Access Concerns` in this batch, although it has been stable over time, as concerns about the location permission are a long-standing and much written about (see §2) privacy issue.

Our data-analysis pipeline captures nuance in reviewers' discussion of permissions, sorting privacy-sensitive permissions reviews into distinct themes. For instance, within the `Excessive Permissions` theme, reviewers complain about too many privacy-sensitive permissions in general, without specifying which ones: e.g., "*what's up with the permissions requests, why don't you listen to us and keep your hands off of our data*". In contrast, reviews labeled `Unneeded Camera Access` and `Location Access Concerns` exhibit concerns about camera and location permissions respectively. `Unneeded [Permission] Access` captures all lesser mentioned privacy-sensitive permissions (47.5% contacts, 9.3% phone numbers, 7.5% call logs, etc.).

We document here, for the first time, the dominant emotions associated with privacy reviews about app permissions. Unsurprisingly, we find that negative emotions ('Anger', 'Annoyance', 'Sadness', 'Disgust', and 'Fear'), account for 45–68% of reviews most themes. (see Figure 5). Qualitative analysis shows that 'Anger' reviews commonly question the use case for privacy-sensitive permissions and express disapproval, such as: "*Horrible! it doesn't make sense to give permission for management of my email account, contacts, and images. All bad ideas.*" Among our top 20 themes, the 'Confusion' emotion appears most often with privacy-sensitive permission themes (7-14.4%). The permission-related reviews associated with 'Confusion' almost exclusively question privacy-sensitive permissions but are less confrontational (e.g., "*Why do you need identity and call information permissions?*").

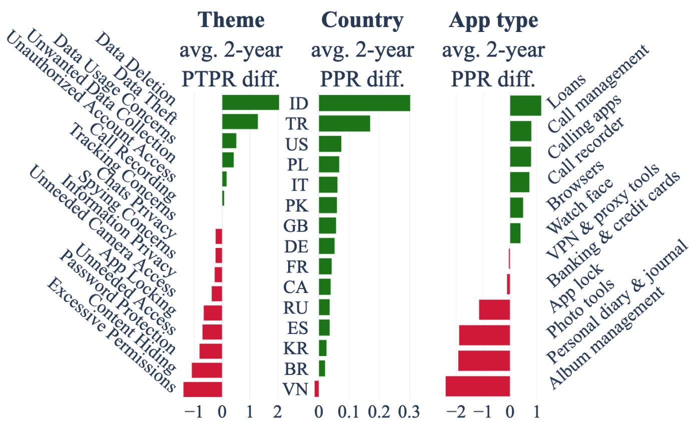

Figure 4: Avg. 2-year % point change in theme (PTPR), country (PPR), and app type (PPR). Green/red denotes increasing/decreasing trend (KPSS $p \leq 0.05$).

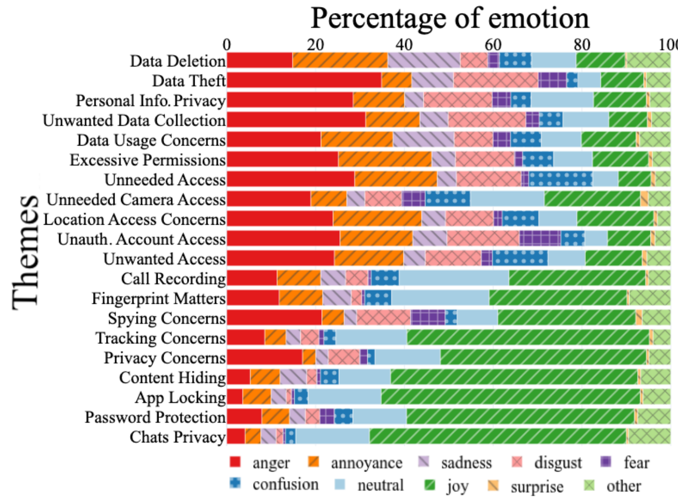

Figure 5: Percentage of emotions in top 20 themes, organized roughly by decreasing anger and increasing joy.

These themes on privacy-sensitive permissions have decreased significantly over time in PTPR (see Figure 4): going from 19% of privacy reviews to 8% between Feb 2013–Jan 2023. We hypothesize several explanations: (1) ongoing privacy enhancements in Android permissions (e.g., run-time permissions in version 6, restricting background usage in version 10, one-time grants in version 11); (2) Android's efforts in urging developers to reduce unnecessary permission requests [57]; and/or (3) the rise of other privacy concerns.

Figure 6 shows a brief increase in reviews about privacy-sensitive permissions around 2016; we speculate this relates to increased visibility of permissions with the introduction of runtime permissions in October 2015. The lag may be due to new Android versions being adopted over multiple years [46].

Finding 2: The fraction of privacy reviews related to privacy-sensitive permissions has decreased from 19% to 8%, in the last 10 years.

Themes related to device sharing: Other themes exhibiting significant decline in PTPR are `App Locking`, `Content Hiding` and `Password Protection`. These themes collectively have dropped from roughly 18% of privacy reviews in Feb 2013 to 5% as of Jan 2023.

We investigated the app types of these reviews, and found the reviews were predominantly for *App lockers/hiders*, *Album management*, and *Personal diary & journal*. Such apps typically allow protection of private content via password controls that are especially useful when a device is shared. South Asian countries contribute the most to these themes (e.g., India makes up 29.4% of `Content Hiding` reviews), and device sharing is a known challenge in these countries [68]). We speculate increasing smartphone adoption—and thus reduced device sharing—might explain the decline among south Asian countries [10]. However, more research is needed to confirm. We further explore these themes and app types in §7.2.

5.2 Themes Increasing in Prevalence

The top four themes that have shown statistically significant increases over time are: `Data Deletion`, `Data Theft`, `Data Usage Concerns`, and `Unwanted Data Collection`. These themes have risen in aggregate from 10% of privacy reviews in early Feb 2013 to 27.3% by Jan 2023. Collectively, these themes focus on PII and user-generated personal data. Our analysis methods discern the nuanced difference between PII and privacy-sensitive permission discussions.

Themes in this group capture feedback in nuanced ways. `Data Deletion`-themed reviews are almost exclusively requests to delete personal data (discussed in §5.3). In contrast, `Data Theft`-themed reviews express concern about data being stolen or an account hacked. `Data Usage Concerns` primarily refers to sharing data with third parties, and `Unwanted Data Collection` reviews express concerns with the app requesting unnecessary data (e.g., for showing ads).

These reviews occur in two styles. Many reference precise data types: national IDs (e.g., SSN, TCKN), financial data (e.g., credit cards), personally identifiable information (e.g., emails), or user-generated content (e.g., "*Requires EVERYTHING about you including your social security number to scan lottery tickets. Why wouldn't you use your local store instead of exposing yourself to a risk of data breaches?*"). Other reviews in these themes discuss personal data vaguely, without reference to specific personal data types (e.g., "*Warning! The app collects a lot of data. Personal data, device and usage-related data... The app then sells it*").

Understandably, the most common emotion with these reviews is 'Anger' (15-35% of privacy reviews per theme). `Data Theft` also contains a larger rate of 'Fear' (6.4% of privacy reviews). Interestingly, these reviews are not strictly negative; 9-12% of privacy reviews on these personal data themes ex-

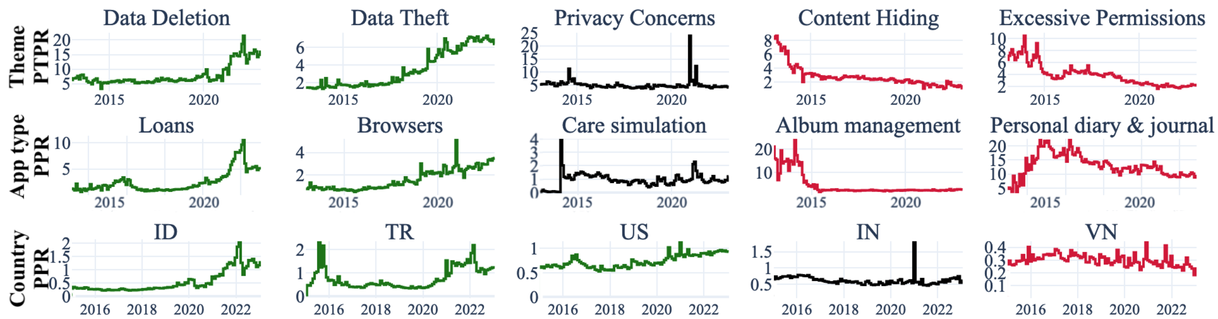

Figure 6: Selected themes, app types, and countries over time. Green/red denotes increasing/decreasing trend (KPSS $p \leq 0.05$)

hibit 'Joy.' Some express gratitude for apps that protect data, such as "*This app locks your personal stuff. Great app.*"

Finding 3: Over time, an increasing share of privacy reviews focus on personal information (e.g., PII). Our analysis uncovers privacy discussions about different types of personal data, such as privacy-sensitive permissions vs PII.

5.3 The rise in Data Deletion and Data Theft

Since `Data Deletion` and `Data Theft` are our two largest themes, and both show a statistically significant increase in volume over time, we now look at what the contributing factors may be. Our analysis enables us to break down our data along multiple dimensions, thus we looked at per-country contributions for these two topics and found that Indonesia (disproportionately) contributes 40% of the reviews in these two categories, whereas all other countries contribute less than 8% each. We also saw that Indonesia has the fastest PPR growth of any country, with 76% relative biannual growth and the third-most privacy reviews of all countries by volume. Considering only the `Data Deletion` and `Data Theft` themes for Indonesia, we next considered app types. We uncovered that 44% of these reviews came from financial loan apps, with other app types contributing less than 0.8% each.

Many `Data Theft` reviews complain that the app collects PII (social media accounts, photos of IDs), and then either loans are denied without reason, or threats are made leveraging their PII. One reviewer writes "*The app stole my data, including my facebook account. It told me that the loan process would be easier if I input all of my data. My loan was still denied even though I have a good payment history. Watch out people.*" Another reviewer says "*[COMPANY] pretends you'll get a loan with no hiccups. They say they just need your photo identity. If you're even a day late, their debt collector will harass you and threaten to leak your data (photo identity) to social media.*"

Most *Data Deletion* reviews we manually examined ask for personal information to be deleted because the user's loan was rejected. Some claim payment was requested when a loan was never applied for or received. For instance, one reviewer from Indonesia exclaims, "*Didn't receive a penny! Why do you threaten me with calling all of my contacts if I don't pay? DELETE ALL OF MY DATA PLEASE.*"

Munyendo et al. conducted a user study with 20 participants in Kenya and reported that loan apps were calling the contacts of people who had not applied for loans [50]. Our NLP analysis pipeline was able to uncover that this issue, first noted in Kenya, is in fact a more widespread problem. In addition to Indonesia, we observe many `Data Deletion` and `Data Theft` reviews on loan apps from India, Mexico and Thailand. In our data, Kenya contributes only a small portion of these reviews, which may reflect population size and/or selection bias in who leaves reviews on Google Play.

Figure 4 indicates that the loans app type is a growing concern, with the highest growth in privacy reviews among app types. In 2022, many such abusive financial loan apps were removed from the market.[6] Subsequently in 2023 Google announced a policy update prohibiting personal loan apps from accessing sensitive data such as photos and contacts.[7] Since our data collection ended in early 2023, we are unable to measure the impact of this policy.

Finding 4: We showed how our ability to break the data down along multiple dimensions (country, time, theme, app type) can be useful in understanding trends. This evaluation also illustrates how text analysis and user studies can complement each other: when a prior user study [50] identifies an unexpected issue, it can be followed up with text analysis to quickly examine the issue's geographic spread.

6 Trends in Countries

We now focus on our third research question (RQ3), examining how privacy themes discussed in reviews vary around the globe. Our 12.3M review data set includes more than 200 countries or regions. The top 10 contributing countries are the U.S. (15.4%), India (12.1%), Indonesia (10.5%), Brazil

[6] https://techcrunch.com/2022/11/18/google-clamps-down-on-illegal-loan-apps-in-kenya-nigeria/

[7] https://support.google.com/googleplay/android-developer/answer/9876821

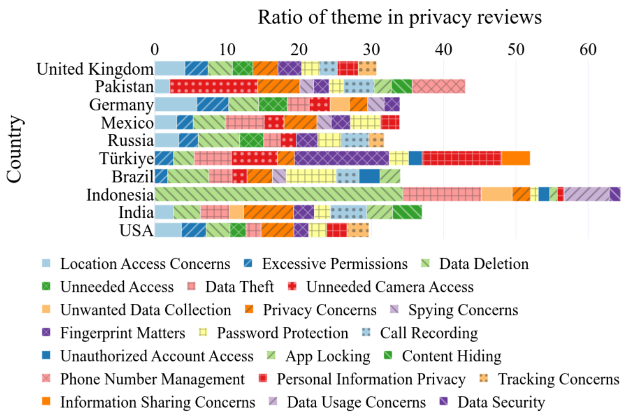

Figure 7: Top 10 themes per country (PTPR), showing variation in theme popularity across countries.

(7.7%), Türkiye (4.5%), Russia (4.4%), Mexico (4.2%), Germany (3.0%), Pakistan (2.6%), and United Kingdom (2.4%). These 10 countries together contribute 66% of privacy reviews, with the top 33 countries supplying 90%. Interestingly, 1) only two European countries are among the top 10 contributors by volume, 2) the top 33 countries come from every continent, and 3) there is a long-tail of privacy review percentages coming from more than 150 countries. Overall, while people from all corners of the world write about privacy, we find that 33 countries dominate this discussion. This distribution of privacy reviews across countries largely follows the trends in all (both privacy and not-privacy) reviews submitted, except for few minor shifts in country rankings. For instance, Brazil is second in terms of review volume, but drops to fourth with privacy reviews. Similarly, South Korea drops from eighth to 13th, while Spain jumps from 15th to 11th. Nonetheless, the top countries that contribute the highest number of reviews are also the ones that contribute the most privacy reviews.

In the rest of this section, we first cluster countries (§6.1) to understand where broad commonalities do and don't exist. We then take a closer look at countries with outlier patterns (distributions of privacy topics discussed) in §6.2, and other anomalous patterns such as high growth rates in Türkiye (§6.3) and unusual reviews about spying in Nigeria (§6.4). It is beyond the scope of this paper to carry out further detailed world-wide country comparisons.

6.1 Clustering Countries

Earlier research [13,26,33,63] has hypothesized and examined whether the influence of culture and geographic proximity changes privacy attitudes. While loose correlations are sometimes observed (e.g., [33,63]), often it is difficult to identify strong predictors because there are so many factors that influence privacy [13,26]. We explore the potential connection between geographic proximity and privacy opinions from a

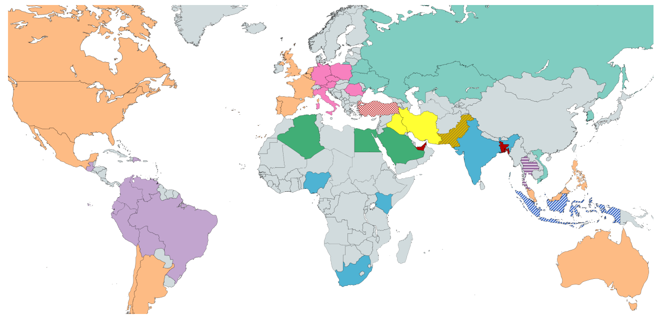

Figure 8: Countries clustered according to distribution of privacy themes discussed in app reviews.

different angle: we examine expressed opinions, from people living in a country, over a range of privacy topics. We do not study underlying factors, such as culture; instead we offer direct data summaries of user-provided texts.

We investigate the question of whether countries that are geographically close discuss similar app privacy issues, in an exploratory fashion. We cluster countries based on prevalence of privacy themes discussed.[8] As intuition for this approach, we provide a visualization in Figure 7 of the distribution of themes for the top 10 countries contributing the most privacy reviews. (We only include top 10 themes per country to facilitate readability.) We see that the U.S. and U.K. have quite similar distributions, while Indonesia, Türkiye and Pakistan have distinct patterns.

We now use these topic distributions to cluster countries. Specifically, we represent each country as a vector of size 50, where each element is the PTPR for a (country, theme) tuple, for the top 50 themes across the entire dataset. We then apply hierarchical clustering with complete linkage [51]. The resulting clusters are shown on a world map in Figure 8. A detailed breakdown can be seen in Figure 12 of the extended paper [3].

We see from this map that the question of whether geographically close countries write about the same privacy issues yields mixed results. On one hand, Russia, Ukraine, Belarus and Kazakhstan all cluster together (teal). We also see that most of the Anglo countries, namely the U.S., Canada, Australia and the U.K., are in the same (light orange) cluster. However, this light orange cluster also contains Mexico, Argentina, Spain, Portugal, and France. Interestingly, European countries divide between two clusters, with eastern European countries, Germany, and Italy in their own cluster (pink). Similarly, the Middle East and North Africa splits across three main clusters: Türkiye is alone, Iran and Iraq cluster together (yellow), and Saudi Arabia, Algeria, and Egypt cluster together (green). A detailed examination of our dendrogram shows that the Iran-Iraq cluster is surprisingly far away from

[8]While this is limited by those who chose to write reviews, it nevertheless is based on sizeable inputs from each country.

Country	Most Discussed Privacy Theme
Pakistan	`Unneeded Camera Access`
Indonesia	`Data Deletion` (mostly in loan apps)
Türkiye	`Fingerprint Matters`
Thailand	`Data Deletion` (broadly across app types)
Bangladesh & UAE	`VPN Matters`
Iran & Iraq	`Data Deletion` (comms & social media apps)

Table 3: Countries with unusual theme distributions

the Saudi Arabia-Algeria-Egypt cluster. Asian countries exhibit wide diversity. India clusters with some African countries (South Africa, Kenya, and Nigeria), and not with Pakistan or Bangladesh. Interestingly, we note that Indonesia, Pakistan, Thailand and Türkiye appear unique (cluster size 1).

Finding 5: Geographical proximity does not reliably indicate whether countries discuss the same set of privacy topics.

6.2 Unique distributions of privacy themes

As noted above, our clustering identified four countries that are alone in clusters of size 1: Indonesia, Pakistan, Thailand, and Türkiye. We also found two small clusters of only two countries: Bangladesh with the United Arab Emirates, and Iran with Iraq. In Table 3, we list the most frequently discussed issues for these countries.

Pakistan is the country with the highest rate of `Unneeded Camera Access` (12.2% of reviews from Pakistan). These reviews distinguish themselves from other countries by praising the abilities of `Hidden Camera Detection` (issue) apps, which claim to detect hidden cameras using device magnetometers (e.g., "*the app detects all hidden cameras in your vicinity, I love this app*"). Pakistan's focus on this issue could be a reflection of a higher adoption of this type of app.

The reviews from Indonesia were discussed in §5.3. The reviews from Türkiye are discussed in more depth in §6.3. Distinctively, Thailand exhibits a combination of the unique behaviors observed in reviews from Indonesia and Türkiye.

Bangladesh and UAE have a higher fraction of reviews categorized as `VPN Matters` than any other country. These reviews are overwhelmingly positive but predominantly non-specific, e.g., "*Love this app. It is a great vpn.*" We observe that 55.0% of `VPN Matters` reviews are for apps that contain 'Free' in the title. Prior work has found that "free" VPNs are often misconfigured or outright malicious [37, 58, 62]. A possible explanation for this theme's relative prevalence could be the common use of VPNs to circumvent the relatively high rate of censorship in Bangladesh and UAE [69].

For both Iran and Iraq the top theme is `Data Deletion`, with the vast majority of reviews applying to communication and social media apps (see §7 for more on apps).

6.3 Türkiye

Türkiye stands out as anomalous based on two metrics. First, its distribution of privacy topics does not cluster with any other country (cluster size 1). Second, Türkiye exhibits anomalous PPR growth over ten years. The middle column of Figure 4 shows the average 2-year change for countries, among the top 20, with statistically significant growth or decline in PPR (KPSS $p \leq 0.05$); stationary (non-changing) countries are excluded. Türkiye shows an unusually large increase, with an average 2-year relative growth of 62.1% between Feb. 2015 and Jan. 2023 (Figure 6 illustrates this). Türkiye also provides the 6th-most privacy reviews (502K) of all countries.

Figure 7 shows that Türkiye has an unusual distribution of privacy themes; `Fingerprint Matters`, `Personal Information Privacy` and `Unneeded Camera Access` appear more frequently than in other countries. The privacy reviews highlight user concerns around fingerprint collection (e.g., "*My fingerprint was scanned on [date], the company has the responsibility of my fingerprint.*") and abuse (e.g., "*If my fingerprint is used in something illegal, the app is responsible.*"). Although biometrics such as fingerprints are not shared with apps directly,[9] reviewers express concern about the potential for such sharing, e.g., when a banking app verifies authorization via a fingerprint.

In addition, we observed that 29% of privacy reviews from Türkiye include text we refer to as a "disclaimer": quasi-legal language asserting rights or claims. These disclaimers assert app developers' responsibility to safeguard information (e.g., "*This app is nice but you responsible for any inappropriate use of my personal data.*") or longer texts that threaten legal recourse for violations:

> "*This application was downloaded at [time] on [date]. I [give] no permission to share things (My photo, T.C. ID number, password, etc.) with third parties. And if such thing happens, ... the app bears sole responsibility and I will take legal action against them.*"

These disclaimer reviews appear for multiple app types, including call management/recorders, physical activity trackers, investment/cryptocurrency apps, VPNs, and even antivirus apps; many of these typically require access to sensitive permissions or resources such as app usage patterns, files, and network control. The disclaimer reviews hint at nuanced mental models in which reviewers worry that an app might violate the privacy it claims to protect.

Overall, we hypothesize that reviewers in Türkiye use these disclaimers when they are uncomfortable with an app's privacy risks but feel compelled to use it, such that the disclaimer seems like the only protective option. Anecdotally, conversations with Turkish nationals suggest similar disclaimers circulate on messaging and social media apps with privacy

[9] https://source.android.com/docs/security/features/authentication/fingerprint-hal

protection promises.[10] These disclaimers raise additional research questions: Why are these reviews seen so frequently from Türkiye? Are reviewers who post these disclaimers unusually privacy-sensitive? Do they believe the disclaimers have legal value? Do they employ other privacy protections?

Finding 6: Over 150K reviews in Türkiye post quasi-legal disclaimers asserting privacy rights. We hypothesize a large-scale misunderstanding might cause such reviews.

6.4 Nigeria

Nigeria is the only country with `Spying Concerns` as the top theme. Surprisingly, many of these reviews are associated with 'Joy,' and unlike most `Spying Concerns` reviews from other countries, they aren't about games (see § 7.1). Manual inspection reveals that reviews for tracking related apps sometimes contain ads for spying services, for example:

> "*I can't find enough words to thank [email]! I tried lots of times to spy on my spouse . . . to no avail. This guy is magic, within three hours, he gave me access the calls and messages of my spouse.*"

Issues associated with `Spying Concerns` include `Spying on Spouse`, `Spying on Partner`, `Allows Unauthorized Calls Access`, and `Allows Unauthorized Messages Access`, suggesting these ads may primarily be targeted to intimate partner abuse perpetrators. Notably, many of these review ads use similarly formatted email addresses. Although we identified this pattern in privacy reviews, a simple regular expression[11] across all reviews yielded more than 10K matches. A random sample of 100 matches yielded only three false positives. Matching reviews appear on apps with developer-specified categories such as Lifestyle, Tools, Books, Communication, and Dating. Most are from Nigeria. We disclosed this issue to Google Play, and these reviews were taken down.

A simple web search for some of the email addresses from these reviews reveals similar ads posted as comments/reviews across the web as well as posts about the (in)effectiveness of the advertised spying,[12] suggesting these services are not unique to Google Play, and may be of interest to the digital intimate partner violence research community [6, 23, 31].

Finding 7: We detected a subset of reviews offering spyware services, primarily from Nigeria. These were reported to Google Play and have been removed before publication.

7 Stark Differences Across App Types

There are 445 app types in our dataset (functional categories displayed on Play app page), with some invoking strong emotional reactions from users. We address RQ4 by focusing on app types among the top 50 (by privacy review volume) that: (1) receive a high rate of negative-emotion privacy reviews, or (2) receive overwhelmingly positive reviews. We draw connections with previous research when possible (addressing RQ5), while also identifying areas that have been under-explored in the literature.

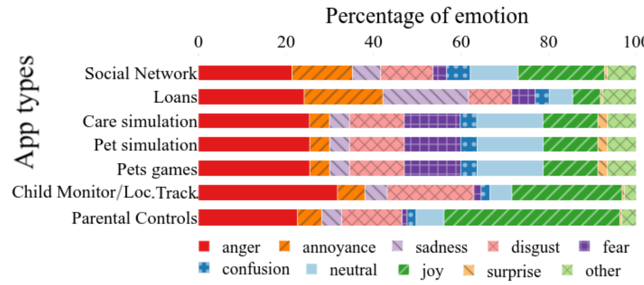

Figure 9: App types with highly negative privacy emotions.

7.1 App types with negative emotions

We define app types with strongly dominant negative emotions as those where more than 50% of privacy reviews were assigned to any of 'Anger', 'Annoyance', 'Sadness', 'Disgust', or 'Fear' (see Figure 9) Of the 25 app types that satisfy this condition, we highlight only a few below due to limited space (loan apps were already discussed in §5.3).

Social media: These apps receive the second-most privacy reviews (775K) of all app types, and ∼57% of these reviews are associated with negative emotions. Reviewers most frequently bring up `Invasion of Privacy` and `Data Deletion`. Non-specific comments, such as in `Invasion of Privacy`, include: "*You invade the privacy of people. Our privacy is sold for profits. You let fake news be posted by foreign countries.*" In `Data Deletion` reviews, users complain that they are unable to delete accounts, posted content, or the app itself (some social media apps come pre-installed and are not removable [25]). This latter feedback is more actionable for developers; for example, improving the UI to make data deletion controls more discoverable could help. The strongly negative emotions indicate users who are very frustrated with privacy properties of social apps.

Pet simulators: Several popular 'simulation' games that mimic users' facial and vocal expressions are grouped under the game categories *Care Simulation*, *Pet Simulation*, and *Pets*. These app types have a total of 199K privacy reviews, and 60% of these are associated with strong negative emotions. Reviews left on these games are often assigned themes including `Unneeded Camera Access`, `Spying Concerns`, and `Unauthorized Surveillance`. These reviews largely express concern that the 'pet' is watching the user through

[10] Similar disclaimers have circulated elsewhere for years [24, 41].
[11] Omitted here due to potential harm; contact the authors for information.
[12] https://www.scamwatcher.com/scam/view/272748

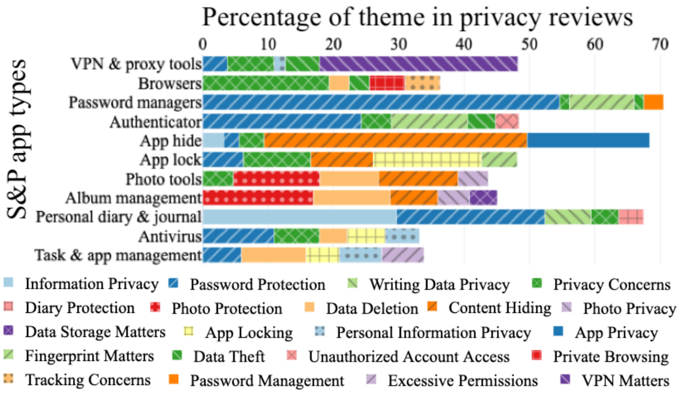

Figure 10: Top security & privacy app types themes (PTPR).

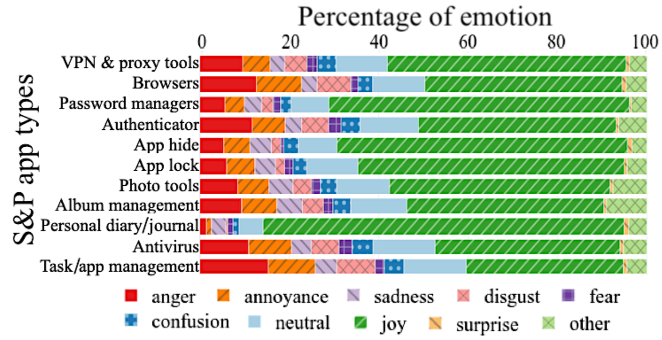

Figure 11: Top security & privacy app types with highly positive emotions.

cameras embedded in its eyes. For example: "*I don't recommend the app. She is immensely dangerous as she has cameras for eyes. She captured my picture. I was playing with it at 3am and she said she would come to my house.*" Though these reviews are based on misconceptions,[13] as these games are primarily intended for children, they often include heightened emotions. Reviews such as these were prevalent in the United States, Brazil, India, Mexico, and Italy. They first appear in early 2014 and have not diminished since.

Child monitors and location tracking: *Child Monitors & Location Tracking* and *Parental Controls* apps also receive a lot of negative feedback, with 52% of 135K privacy reviews associated with negative emotions. Generic `Privacy Concerns` is the top theme; somewhat less common, but more specific, themes include `Children Privacy Concerns` and `Monitoring Children`. Reviewers of these apps typically complain about being tracked by the app or object to the existence of this app type in general.[14] However, positive privacy reviews of child-monitoring apps are also common (36% contain 'Joy'). Positive reviews frequently praise the ability to `Track` and `Access Location`. Conflicts between children's privacy and parental supervision have been studied in a variety of contexts (e.g., [16, 76]), including a study of 736 reviews of child-monitoring apps [27]. In that study, which explicitly aimed to study reviews by children, most reviews were negative.

Other work has reported on the use of these apps for intimate partner abuse [12]. Our manual analysis did not identify any reviews explicitly acknowledging tracking adults, and we observe few reviews in related categories like `Spying concerns` or `Unauthorized surveillance`. This identifies a limitation of using our approach to complement user studies or surveys, namely that we cannot guarantee that reviews on a particular topic will be part of our data collection.

[13] https://www.usatoday.com/story/news/nation-now/2014/02/20/talking-angela-app-scare-hoax/5635337/

[14] Because these reviews may have been by children, we provide no quotes.

7.2 Privacy Protective Apps: Quite Positive!

One might intuit that when users write about privacy, they primarily complain, and hence those reviews are associated with negative emotions. Interestingly, 32% of all privacy reviews are associated with 'Joy' emotion. Often these reviews express appreciation for a particular privacy feature (e.g., private chat). To explore this further, we identified app types in the top 50 with at least 25% of reviews associated with 'Joy' (i.e., privacy positive reviews). Of the 23 app types satisfying this condition, the majority were privacy- and security-focused.

Below we summarize reviews from these privacy- and security-focused app types. Their dominant privacy themes are shown in Figure 10, and Figure 11 shows the emotion distribution. At a high level, users appreciate the (perceived) protections they receive from these apps. Our results confirm and extend prior work while also highlighting research gaps.

VPN & proxy tools: These apps received 128K privacy reviews, and 53% of these express 'Joy.' Reviewers broadly discussed `VPN Matters`, `Privacy Concerns`, and `Data Theft` themes. `VPN Matters` reviews frequently consist of short, vague endorsements (e.g., "*Great VPN, among the best out there*"). We looked at the fine grained issues within the VPN matters theme, such as, (`Best VPN`) and found the issue names similarly undescriptive. Broad affection for VPNs has been reported in prior measurements [4, 18]. Within `Privacy Concerns`, we also see fairly generic positive reviews, grouped under issues like `Privacy Protection` (e.g., "*Love the app. I get decent speeds and excellent privacy.*") and `IP Protection`. These comments are similar to terminology appearing in influencer VPN ads [4].

Differently from some prior work, reviewers in our dataset do not commonly mention benefits such as protection against internet surveillance [52, 62], utility for censorship evasion [18], or lack of server-side logging [62]. Further, while findings from prior work on VPNs tend to emphasize specific adversaries [4, 18, 62], we don't observe themes or issues with this emphasis. As expected from prior work [70], reviews commonly show confused mental models, e.g., when one user claims a VPN "*Hides my IP address from my internet ISP.*"

Browsers: These apps receive the fifth-most privacy reviews (332K) of all app types, 44% of which express 'Joy.' The majority of these privacy-positive reviews apply to smaller browsers that advertise privacy-focused design. Reviews praise `Private Browsing` modes or private browsers and built-in tracking protection (`Tracking Concerns`), e.g., "*Amazing browser for your privacy, blocking trackers. They do not keep search records is just what people need recently.*" Overall, users seem to appreciate browsers with enhanced privacy protections. Browsers are the only security & privacy app type with significantly increasing (KPSS $p \leq 0.05$) privacy reviews: a two-year average growth of 0.5% points in $PPR_{Browsers}$ (46.5% average relative growth).

App Hide and App Lock: App hiders attempt to conceal a user-selected list of apps from appearing in the list of installed apps. They generally achieve this functionality by being a *launcher*, the default navigation app of the OS. App lockers provide access control mechanisms (e.g., a password) before a user is able to open an app, provided the user is using the app locker as the launcher. Although they have related functionality, app lockers do not attempt to hide apps.

These two app types together have 148K privacy reviews, of which ∼60% express 'Joy.' Themes and issues overlap almost completely across the two app types. Reviewers of *App Lock* apps praise `Content Hiding` (e.g., "*Perfect application. It hides videos, apps, and photos.*"). Similarly, reviews for *App Hide* apps express satisfaction with their hiding functionality under the theme `App Privacy`.

Despite their popularity and users' apparent satisfaction with app lockers/hiders, surprisingly little research has examined these tools. Sambasivan et al. found important use cases for app lockers among women in South Asia [68]. Though we do not have gender information, we find that India contributed the most reviews for these apps. Kenya, Nigeria, Zambia, Venezuela, Pakistan, and Türkiye also contribute a high rate of these reviews, but use in these regions has not, to our knowledge, been studied. Other work points out that device sharing creates important threats that are not always well supported by developers and security professionals [75].

Researchers have found some app lockers may be easily circumventable [45], suggesting that users may be less protected than they believe. Other research has shown app lockers' resemblance to malware [7, 65]. However, despite wide use, there is little research analyzing the security, privacy, and usability of these apps. We argue these apps remain an interesting research topic.

Photo tools and album management: The *Photo tools* app type has 208K privacy reviews, while *Album management* has 145K. Reviews for these app types frequently express 'Joy' (>45%), often praising `Photo Protection` (e.g., "*I love this app. Photos are ALWAYS protected*") and `Content Hiding` (e.g., "*Greatest application for hiding photos or vids*") features. Unlike app lockers and hiders, these apps allow users to secure media beyond the initial lock screen.

Diary apps: Apps in this category receive a perhaps surprising amount of privacy-relevant reviews (149K). These reviews carry an exceptionally positive tone (80.7% marked with 'Joy', the highest rate among all app types) and most often fall under the themes `Information Privacy` and `Password Protection`. Users most often find these apps useful to `Keep secrets` (within the `Information Privacy` theme): "*I really like this. I can keep my secrets. Seriously, thanks.*" Some users specifically praise the ability to password-protect diary entries. These apps are highly regarded and widely used. However, we are unaware of any recent technical analyses of these tools, though some older work exists [19]. We recommend future investigations into how secure these apps are.

Antivirus & task management: *Antivirus* app type has 106K privacy reviews, of which 41% express 'Joy.' `Password Protection`, `Privacy Concerns`, `App Locking`, `Personal Information Privacy`, and `Data Deletion` are the most common themes. Users comment about a variety of issues outside the traditional antivirus app functionality, which we posit relates to antivirus apps expanding their feature offerings.[15] For example, we see praise for app locking features not only in apps categorized as app lockers, but also in general antivirus apps, many of which offer such features. We see relatively fewer mentions of malware (the primary purpose of antivirus software); examples from issues like `Phone Protection` and `Malware Protection`, include: "*This is a perfect app. It finds and removes malware very fast. Further, it detects apps that are a threat to your privacy. I 100% recommend it.*"

The *Task & app management* app type has 51k privacy reviews, of which 35% are privacy positive. These apps exhibit similar trends to antivirus apps, perhaps because of large overlap in functionality: both managers and antivirus programs offer security and performance features. However, reviews of task management apps include more reviews about `Excessive permissions`. Though both app types require sensitive permissions to function (e.g., accessing all files), task & app management users may be more hesitant to grant these permissions outside a security context: "*what is up with these crazy permissions? thanks. nope.*"

Finding 8: Within the 32% of privacy reviews that are positive, users praise privacy-protecting mechanisms (data deletion, hiding, password-protected access) for several data types (photos, diary entries, app visibility, browsing history).

[15] https://www.pcmag.com/picks/the-best-security-suites

8 Discussion & Conclusion

In examining more than a decade of Google Play app reviews, we find that privacy reviews are growing at a biannual rate of 9%. `Data Deletion` is the number one privacy issue worldwide in terms of total reviews and is a top-5 concern in 47 countries. While this issue is growing, other prevalent issues such as those around permissions (e.g., `Excessive Permissions`) are declining. We further illustrated how an emotions classifier can illuminate which app types cause the most privacy concern and which receive the most privacy praise, across an entire app store.

User research on security & privacy tools rarely focuses on what users appreciate; rather, it highlights shortcomings (e.g., [36, 44, 62, 70]). Our analysis adds perspective by unearthing several potential **privacy wins**. First, we found a large body of positive privacy reviews for app lockers and hiders, VPNs, journaling apps, and album management. Many of these app categories are understudied; future work could investigate them in more depth, including improvements to support user workflows and technical analyses of whether apps are providing the privacy users expect. Second, we see that, over time, the ratio of permission-related complaints in privacy reviews has dropped by more than half (§5.1). This steep decline suggests permissions concerns could be abating, after concerted effort to improve permissions systems. This hypothesis would be well served by a future user study.

Our analysis can also be leveraged to **identify anomalous behaviors**, such as countries with unique patterns of topic discussions (e.g., Nigeria and Türkiye), or countries with abnormally high privacy-review growth rates (e.g., Indonesia). We uncovered large groups of reviewers who use reviews to communicate with developers in ways unlikely to achieve their goals [38]. Examples include ineffective quasi-legal privacy disclaimers in Türkiye, as well as data deletion requests for specific accounts in Indonesia.

As previously noted [29], this privacy-review analysis could **help developers**. Broadly speaking, we see two types of privacy feedback. Non-specific feedback, such as reviews that say "*this is privacy invasive, do not install*," are not directly actionable, but they do offer developers an understanding of privacy sentiment, and their volume matters since they often discourage other users from installing an app. More specific feedback can be turned into actionable insights (e.g., "*please add private chat*"). Finally, privacy-positive reviews may offer developers a new kind of privacy success metric.

We argue that our analysis approach can be useful for a variety of **research purposes**. We have shown in this paper examples of: (1) *corroborating existing research*, such as the broad public support for VPNs [18] (in § 7.2), and the divergent views about child monitoring apps [16, 76]; (2) *adding context to prior work*, such as showing that privacy concerns about financial loan apps from Kenya [50] are being similarly reported in Indonesia, India, and Thailand (in § 5.3); and (3) *identifying unexpected or emerging privacy issues* that can seed future work using interviews (to enable direct discussion with users) or surveys (that can direct users to focus on specific aspects). Examples include examining if diary apps and app lockers actually deliver the privacy features offered, and whether users correctly understand the privacy offerings of those apps, in addition to exploring why permissions are a decreasing topic of discussion.

Recent advances in large language models (LLMs), such as OpenAI's GPT-4 [56] or Google's Gemini [39], have shown promising performance gains on multiple natural language tasks. However, these LLMs require significant prompt engineering (see Appendix B), have high inference costs, and cannot easily scale to large datasets, unless they have been distilled to smaller models [34]. Exploring such distillation approaches for the purpose of improving the models used in this work is a natural avenue of future work.

In summary, we show that large-scale text analysis, followed by zooming in to explore changing trends or unusual patterns, is useful to both summarize the privacy pulse as it ebbs and flows across much of the globe, as well as to surface privacy issues that may not be regularly tracked.

9 Acknowledgements

We would like to thank Kurt Thomas and Patrick Gage Kelley for their insightful comments on earlier manuscripts. We also thank the anonymous USENIX reviewers for their feedback, and the Google Play team for providing data access and removing spying reviews we identified.

References

[1] Desiree Abrokwa, Shruti Das, Omer Akgul, and Michelle L. Mazurek. Comparing security and privacy attitudes among U.S. users of different smartphone and Smart-Speaker platforms. In *17th Symposium on Usable Privacy and Security (SOUPS)*, 2021.

[2] Omer Akgul, Wei Bai, Shruti Das, and Michelle L Mazurek. Evaluating in-workflow messages for improving mental models of end-to-end encryption. In *30th USENIX Security Symposium*, 2021.

[3] Omer Akgul, Sai Teja Peddinti, Nina Taft, Michelle L. Mazurek, Hamza Harkous, Animesh Srivastava, and Benoit Seguin. A decade of privacy-relevant android app reviews: Large scale trends. *arXiv:2403.02292*, 2024.

[4] Omer Akgul, Richard Roberts, Moses Namara, Dave Levin, and Michelle L Mazurek. Investigating influencer vpn ads on youtube. In *43rd IEEE Symposium on Security and Privacy*, 2022.

[5] Ankur Bapna, Isaac Caswell, Julia Kreutzer, Orhan Firat, Daan van Esch, Aditya Siddhant, Mengmeng Niu, Pallavi Baljekar, Xavier Garcia, Wolfgang Macherey, et al. Building machine translation systems for the next thousand languages. *arXiv:2205.03983*, 2022.

[6] Rosanna Bellini. Paying the price: When intimate partners use technology for financial harm. In *CHI Conference on Human Factors in Computing Systems*, 2023.

[7] Antonio Bianchi, Jacopo Corbetta, Luca Invernizzi, Yanick Fratantonio, Christopher Kruegel, and Giovanni Vigna. What the app is that? deception and countermeasures in the android user interface. In *36th IEEE Symposium on Security and Privacy*, 2015.

[8] Veroniek Binkhorst, Tobias Fiebig, Katharina Krombholz, Wolter Pieters, and Katsiaryna Labunets. Security at the end of the tunnel: The anatomy of VPN mental models among experts and Non-Experts in a corporate context. In *31st USENIX Security Symposium*, 2022.

[9] Ron Bitton, Kobi Boymgold, Rami Puzis, and Asaf Shabtai. Evaluating the information security awareness of smartphone users. In *CHI Conference on Human Factors in Computing Systems*, 2020.

[10] Rebecca Bundhun. Why smartphones are key for india's digital ambitions. *The National*, 2022. https://www.thenationalnews.com/business/technology/2022/09/05/why-smartphones-are-key-for-indias-digital-ambitions/.

[11] Weicheng Cao, Chunqiu Xia, Sai Teja Peddinti, David Lie, Nina Taft, and Lisa M Austin. A large scale study of user behavior, expectations and engagement with android permissions. In *30th USENIX Security Symposium*, 2021.

[12] Rahul Chatterjee, Periwinkle Doerfler, Hadas Orgad, Sam Havron, Jackeline Palmer, Diana Freed, Karen Levy, Nicola Dell, Damon McCoy, and Thomas Ristenpart. The spyware used in intimate partner violence. In *39th IEEE Symposium on Security and Privacy*, 2018.

[13] Hichang Cho, Bart Knijnenburg, Alfred Kobsa, and Yao Li. Collective privacy management in social media: A cross-cultural validation. *ACM Trans. Comput.-Hum. Interact.*, June 2018.

[14] Hichang Cho, Milagros Rivera-Sánchez, and Sun Sun Lim. A multinational study on online privacy: global concerns and local responses. *New media & society*, 2009.

[15] Jessica Colnago, Lorrie Faith Cranor, Alessandro Acquisti, and Kate Hazel Stanton. Is it a concern or a preference? an investigation into the ability of privacy scales to capture and distinguish granular privacy constructs. In *18th Symposium on Usable Privacy and Security (SOUPS)*, 2022.

[16] Lorrie Faith Cranor, Adam L Durity, Abigail Marsh, and Blase Ur. Parents' and teens' perspectives on privacy in a technology-filled world. In *10th Symposium On Usable Privacy and Security (SOUPS)*, 2014.

[17] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. GoEmotions: A Dataset of Fine-Grained Emotions. In *58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.

[18] Agnieszka Dutkowska-Zuk, Austin Hounsel, Amy Morrill, Andre Xiong, Marshini Chetty, and Nick Feamster. How and why people use virtual private networks. In *31st USENIX Security Symposium*, 2022.

[19] Berni Dwan. The security implications of the digital diary. *Computer Fraud & Security*, 1997, 1997.

[20] Serge Egelman, Marian Harbach, and Eyal Peer. Behavior ever follows intention? a validation of the security behavior intentions scale (sebis). In *CHI conference on human factors in computing systems*, 2016.

[21] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6, 1992.

[22] Adrienne Porter Felt, Erika Chin, Steve Hanna, Dawn Song, and David Wagner. Android permissions demystified. In *18th ACM conference on Computer and communications security (CCS)*, 2011.

[23] Diana Freed, Jackeline Palmer, Diana Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. "a stalker's paradise": How intimate partner abusers exploit technology. In *CHI Conference on Human Factors in Computing Systems*, 2018.

[24] Zack Friedman. Stop falling for this Facebook scam. *Forbes*, August 2019. https://www.forbes.com/sites/zackfriedman/2019/08/19/facebook-scam/?sh=4b3a3abda4cf.

[25] Julien Gamba, Mohammed Rashed, Abbas Razaghpanah, Juan Tapiador, and Narseo Vallina-Rodriguez. An analysis of pre-installed android software. In *41st IEEE Symposium on Security and Privacy*, 2020.

[26] Reza Ghaiumy Anaraky, Yao Li, and Bart Knijnenburg. Difficulties of measuring culture in privacy studies. *Proceedings of the ACM Human-Computer Interaction*, 2021.

[27] Arup Kumar Ghosh, Karla Badillo-Urquiola, Shion Guha, Joseph J LaViola Jr, and Pamela J Wisniewski. Safety vs. surveillance: what children have to say about mobile apps for parental control. In *CHI Conference on Human Factors in Computing Systems*, 2018.

[28] Marian Harbach, Alexander De Luca, Nathan Malkin, and Serge Egelman. Keep on lockin'in the free world: A multi-national comparison of smartphone locking. In *CHI Conference on Human Factors in Computing Systems*, 2016.

[29] Hamza Harkous, Sai Teja Peddinti, Rishabh Khandelwal, Animesh Srivastava, and Nina Taft. Hark: A deep learning system for navigating privacy feedback at scale. In *43rd IEEE Symposium on Security and Privacy*, 2022.

[30] Ayako A. Hasegawa, Daisuke Inoue, and Mitsuaki Akiyama. How weird is usable privacy and security research? In *33rd USENIX Security Symposium*, 2024.

[31] Sam Havron, Diana Freed, Rahul Chatterjee, Damon McCoy, Nicola Dell, and Thomas Ristenpart. Clinical computer security for victims of intimate partner violence. In *28th USENIX Security Symposium*, 2019.

[32] Pengcheng He, Jianfeng Gao, and Weizhu Chen. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. *arXiv:2111.09543*, 2021.

[33] Franziska Herbert, Steffen Becker, Leonie Schaewitz, Jonas Hielscher, Marvin Kowalewski, Angela Sasse, Yasemin Acar, and Markus Dürmuth. A world full of privacy and security (mis)conceptions? findings of a representative survey in 12 countries. In *CHI Conference on Human Factors in Computing Systems*, 2023.

[34] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv:1503.02531*, 2015.

[35] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv:2106.09685*, 2021.

[36] Nicolas Huaman, Sabrina Amft, Marten Oltrogge, Yasemin Acar, and Sascha Fahl. They would do better if they worked together: The case of interaction problems between password managers and websites. In *42nd IEEE Symposium on Security and Privacy*, 2021.

[37] Muhammad Ikram, Narseo Vallina-Rodriguez, Suranga Seneviratne, Mohamed Ali Kaafar, and Vern Paxson. An analysis of the privacy and security risks of android vpn permission-enabled apps. In *ACM Internet Measurement Conference (IMC)*, 2016.

[38] Google Inc. Comment posting policy. *Google*, 2023. https://web.archive.org/web/20230515141330/https://play.google.com/about/comment-posting-policy/?hl=en-US.

[39] Google Inc. Gemini: A family of highly capable multimodal models. *arXiv:2312.11805*, 2023.

[40] Tadayoshi Kohno, Yasemin Acar, and Wulf Loh. Ethical frameworks and computer security trolley problems: Foundations for conversations. *arXiv:2302.14326*, 2023.

[41] Maya Kosoff. This Facebook 'privacy notice' is a hoax, and you shouldn't bother sharing it. *Insider*, January 2015. https://www.businessinsider.com/facebook-privacy-notice-hoax-2015-1.

[42] Katharina Krombholz, Karoline Busse, Katharina Pfeffer, Matthew Smith, and Emanuel Von Zezschwitz. "if https were secure, i wouldn't need 2fa"-end user and administrator mental models of https. In *40th IEEE Symposium on Security and Privacy*, 2019.

[43] Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54, 1992.

[44] Linda Naeun Lee, David Fifield, Nathan Malkin, Ganesh Iyer, Serge Egelman, and David A Wagner. A usability evaluation of Tor launcher. *Proceedings on Privacy Enhancing Technologies*, 2017.

[45] Ahmed Y Mahmoud and Ali Osama Mahdi. Comments on multi-window against mobile application lock. *Journal of Multidisciplinary Engineering Science Studies (JMESS)*, (5), 2016.

[46] Mehran Mahmoudi and Sarah Nadi. The android update problem: An empirical study. In *15th International Conference on Mining Software Repositories (MSR)*, 2018.

[47] Nathan Malkin, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner. Privacy attitudes of smart speaker users. *Proceedings on Privacy Enhancing Technologies*, 2019.

[48] Abraham H Mhaidli, Yixin Zou, and Florian Schaub. " we can't live without {Them!}" app developers' adoption of ad networks and their considerations of consumer risks. In *15th Symposium on Usable Privacy and Security (SOUPS)*, 2019.

[49] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and James Rosenquist. Understanding the demographics of twitter users. In *5th International*

AAAI Conference on Web and Social Media (ICWSM), 2011.

[50] Collins W Munyendo, Yasemin Acar, and Adam J Aviv. "desperate times call for desperate measures": User concerns with mobile loan apps in kenya. In *43rd IEEE Symposium on Security and Privacy*, 2022.

[51] Fionn Murtagh and Pedro Contreras. Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2012.

[52] Moses Namara, Daricia Wilkinson, Kelly Caine, and Bart P Knijnenburg. Emotional and practical considerations towards the adoption and abandonment of vpns as a privacy-enhancing technology. *Proceedings on Privacy Enhancing Technologies*, 2020.

[53] P Nema, P Anthonysamy, N Taft, and S.T. Peddinti. Analyzing user perspectives on mobile app privacy at scale. In *44th International Conference on Software Engineering (ICSE)*, 2022.

[54] Duc Cuong Nguyen, Erik Derr, Michael Backes, and Sven Bugiel. Short text, large effect: Measuring the impact of user reviews on android app security & privacy. In *40th IEEE symposium on Security and Privacy*, 2019.

[55] Jianmo Ni, Gustavo Hernandez Abrego, Noah Constant, Ji Ma, Keith Hall, Daniel Cer, and Yinfei Yang. Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models. In *Findings of the Association for Computational Linguistics (ACL)*, 2022.

[56] OpenAI. Gpt-4 technical report. *arXiv:2303.08774*, 2023.

[57] Sai Teja Peddinti, Igor Bilogrevic, Nina Taft, Martin Pelikan, Úlfar Erlingsson, Pauline Anthonysamy, and Giles Hogben. Reducing permission requests in mobile apps. In *ACM Internet Measurement Conference (IMC)*, 2019.

[58] Diego Perino, Matteo Varvello, and Claudio Soriente. Long-term measurement and analysis of the free proxy ecosystem. *ACM Transactions on the Web (TWEB)*, 2019.

[59] Jeffrey T Prince and Scott Wallsten. How much is privacy worth around the world and across platforms? *Journal of Economics & Management Strategy*, (4), 2022.

[60] Emilee Rader. Normative and Non-Social beliefs about sensor data: Implications for collective privacy management. In *18th Symposium on Usable Privacy and Security (SOUPS)*, 2022.

[61] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140), 2020.

[62] Reethika Ramesh, Anjali Vyas, and Roya Ensafi. " all of them claim to be the best": Multi-perspective study of vpn users and vpn providers. *arXiv:2208.03505*, 2022.

[63] Elissa M Redmiles. " should i worry?" a cross-cultural examination of account security incident response. In *40th IEEE Symposium on Security and Privacy*, 2019.

[64] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. How well do my results generalize? comparing security and privacy survey results from mturk, web, and telephone samples. In *40th IEEE Symposium on Security and Privacy*, 2019.

[65] Chuangang Ren, Peng Liu, and Sencun Zhu. Windowguard: Systematic protection of GUI security in android. In *Network and Distributed System Security Symposium (NDSS)*, 2017.

[66] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, Serge Egelman, et al. "won't somebody think of the children?" examining coppa compliance at scale. In *18th Privacy Enhancing Technologies Symposium (PETS)*, 2018.

[67] Derek Ruths and Jürgen Pfeffer. Social media for large studies of behavior. *Science*, 2014.

[68] Nithya Sambasivan, Garen Checkley, Amna Batool, Nova Ahmed, David Nemer, Laura Sanely Gaytán-Lugo, Tara Matthews, Sunny Consolvo, and Elizabeth Churchil. " privacy is not for me, it's for those rich women" performative privacy practices on mobile phones by women in south asia. In *14th Symposium on Usable Privacy and Security (SOUPS)*, 2018.

[69] Adrian Shahbaz, Allie Funk, and Kian Vesteinsson. Freedom on the net 2022: Countering an authoritarian overhaul of the internet. *Freedom House*, 2022. https://freedomhouse.org/sites/default/files/2022-10/FOTN2022Digital.pdf.

[70] Peter Story, Daniel Smullen, Yaxing Yao, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. Awareness, adoption, and misconceptions of web privacy tools. *Proceedings on Privacy Enhancing Technologies*, 2021.

[71] Mohammad Tahaei, Alisa Frik, and Kami Vaniea. Deciding on personalized ads: Nudging developers about user privacy. In *17th Symposium on Usable Privacy and Security (SOUPS)*, 2021.

[72] Jenny Tang, Eleanor Birrell, and Ada Lerner. Replication: How well do my results generalize now? the external validity of online privacy and security surveys. In *18th Symposium on Usable Privacy and Security (SOUPS)*, 2022.

[73] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *8th Symposium on Usable Privacy and Security (SOUPS)*, 2012.

[74] Yang Wang, Gregory Norice, and Lorrie Faith Cranor. Who is concerned about what? a study of american, chinese and indian users' privacy concerns on social network sites. In *Trust and Trustworthy Computing*, 2011.

[75] Noel Warford, Tara Matthews, Kaitlyn Yang, Omer Akgul, Sunny Consolvo, Patrick Gage Kelley, Nathan Malkin, Michelle L. Mazurek, Manya Sleeper, and Kurt Thomas. Sok: A framework for unifying at-risk user research. In *43rd IEEE Symposium on Security and Privacy*, 2022.

[76] Miranda Wei, Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. Anti-Privacy and Anti-Security advice on TikTok: Case studies of Technology-Enabled surveillance and control in intimate partner and Parent-Child relationships. In *18th Symposium on Usable Privacy and Security (SOUPS)*, 2022.

[77] Primal Wijesekera, Arjun Baokar, Ashkan Hosseini, Serge Egelman, David Wagner, and Konstantin Beznosov. Android permissions remystified: A field study on contextual integrity. In *24th USENIX Security Symposium*, 2015.

[78] Yuhang Zhao, Yaxing Yao, Jiaru Fu, and Nihan Zhou. " if sighted people know, i should be able to know:" privacy perceptions of bystanders with visual impairments around camera-based technology. In *32nd USENIX Security Symposium*, 2022.

Appendix

A Translated Languages

These 24 languages were translated (percentage of our data): Spanish (13.3%), Indonesian (9.3%), Portuguese (7.8%), Russian (5.8%), Turkish (4.2%), German (3.3%), Arabic (2.5%), French (2.3%), Italian (1.7%), Korean (1.7%), Vietnamese (1.1%), Polish (1%), Persian (0.6%), Thai (0.5%), Dutch (0.5%), Romanian (0.2%), Czech (0.2%), Hungarian (0.2%), Ukrainian (0.1%), Greek (0.1%), Chinese (0.1%), Japanese (0.1%), Malay (0.07%), Hindi (0.06%).

B Privacy Classifier Baselines

We investigated various modern transformer architectures (excluding traditional models, such as BiLSTM, which are inferior), and selected six models based on comparisons made within the recently published and extensively cited DeBERTaV3 paper [32]. In addition, we also considered BART, T5, and FLAN-T5 models for their popularity. We chose the largest available variant of DeBERTaV3, and chose other models' sizes accordingly. We use Hark's [29] T5-11B model as a comparative baseline; our T5-11B model only differs in the diverse training dataset used (§3.1). This list of models is not exhaustive, but we believe it offers a reasonable baseline.

All models were trained and tested on the datasets created in §3.1. Each model was trained with parameter optimization and 'early stopping' (using validation loss as the metric) to avoid overfitting. Each model's best trained variant (highest obtained F1) are reported in Table 4. Our findings suggest that our T5-11B model delivers the best performance across the board, with a highest ROC-AUC of 0.95 and F1 of 0.87.

In addition, we also experimented with Gemini 1.0 Pro, a state-of-the-art (closed-source) LLM [39]. Despite our prompt engineering (i.e., trying multiple prompt variants and using few-shot examples), we could not better the performance T5-11B. Our prompts described our privacy taxonomy, where each high level concept was defined using fine-grained aspects. The prompts further included instances of reviews and their labels as few-shot examples. For the best performing prompt (see the extended paper [3]), the Gemini Pro performed well on recall (0.83) but precision suffered (0.77), with a ROC-AUC of 0.87. We hypothesize that the poor precision arises because the privacy nuances captured in our carefully annotated training dataset are hard to establish using constrained prompts. Exploring LoRA techniques [35] to finetune an LLM for this task is outside the scope of this paper.

Model	Accur.	F1	P	R	ROC-AUC
Ours ($T5_{11B}$)	0.89	0.87	0.91	0.84	0.95
$DeBERTaV3_{Large}$	0.85	0.84	0.81	0.87	0.92
$T5_{Large}$	0.83	0.83	0.77	0.90	0.92
$ELECTRA_{Large}$	0.85	0.82	0.88	0.76	0.90
$FLAN-T5_{Large}$	0.84	0.81	0.85	0.77	0.91
$RoBERTa_{Large}$	0.82	0.81	0.79	0.82	0.88
$Gemini_{Pro}$	0.82	0.80	0.77	0.83	0.87
$BART_{Large}$	0.81	0.80	0.76	0.84	0.88
Hark ($T5_{11B}$)	0.79	0.77	0.73	0.81	0.87
$ALBERT_{Large}$	0.79	0.76	0.76	0.75	0.84
$BERT_{Large}$	0.78	0.75	0.74	0.77	0.84
$XLNet_{Large}$	0.74	0.72	0.67	0.79	0.82

Table 4: Privacy classifier performance. P: precision, R: recall.