Tinker, Tailor, Trust: How Developers Create Privacy Policies With and Without AI

Shiva Mayahi
New Jersey Institute of Technology
Newark, New Jersey, USA
sm3764@njit.edu

Noura Alomar
King Saud University
Riyadh, Saudi Arabia
nnalomar@ksu.edu.sa

Nathan Malkin
New Jersey Institute of Technology
Newark, New Jersey, USA
nathan.malkin@njit.edu

Abstract

For mobile developers to comply with privacy regulations, they must create privacy policies that accurately describe their apps' data practices. This requires a complete understanding of their apps' behaviors, including those of embedded third-party SDKs. Despite the complexity of this process, little is known about how privacy policies are created and validated. To investigate, we interviewed 20 developers from around the world about their processes, also observing them use a large language model (LLM) to prepare privacy policies for their apps. We found that developers struggle with collecting information about third-party SDKs, even when they use LLMs, and feel uncertain about the legal validity of LLM outputs. Many developers do not seek legal assistance and believe that, as long as app stores accept their privacy policies, they are protected. Our findings suggest that reliance on LLMs and developers' desire to externalize validation may result in increasingly unreliable privacy policies.

CCS Concepts

• **Human-centered computing** → **Collaborative and social computing**; • **Security and privacy** → **Human and societal aspects of security and privacy**; • **Social and professional topics** → **Privacy policies**.

Keywords

mobile developer, privacy policy, LLM, AI, usable privacy

ACM Reference Format:
Shiva Mayahi, Noura Alomar, and Nathan Malkin. 2026. Tinker, Tailor, Trust: How Developers Create Privacy Policies With and Without AI. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3772318.3791323

1 Introduction

Privacy policies are a prominent and inescapable component of the modern online experience. Nearly every website and app has one. Nearly no one reads them [68]. Nonetheless, having a privacy policy is required under many regulations [2, 3] and enforced by the major app stores [18, 42].

To comply with these requirements, developers must accurately describe the data practices of their apps, including third-party software development kits (SDKs) and data (sub)processors. They must also craft the privacy statement itself, which may need to adhere to additional regional requirements, such as information or language to include. If the resulting document is incorrect or otherwise flawed, the developer may face legal consequences, thus raising the stakes of a task that otherwise shares many commonalities with requirements engineering [56]. However, this legal potential belies the reality of the privacy policies' creation. There are millions of websites and apps, many run by individuals and small teams. Is it likely that all of them seek legal counsel before publishing their privacy policies?

This disconnect between legal expectations and real-life trade-offs may explain why researchers consistently find discrepancies between apps' stated privacy policies and their actual behaviors [71, 123]. Such inconsistencies are particularly concerning because, despite these inaccuracies, privacy policies are assumed to be true and accurate by end users, app stores (who do not perform their own verification [19, 33, 61, 62]), and academics (such as those working on analyzing [71, 89] or improving the comprehension of [21, 79, 113] privacy policies).

If we are going to continue relying on privacy policies, we need to better understand how, in practice, they are being created. Moreover, a stronger understanding of the creation process can inform and support research that aims to create better tools for automating the creation of privacy policies [52, 92, 118, 122]. These goals motivated the first research question of our study:

RQ1: How do developers create privacy policies currently?

The question of how privacy policies are created, and their accuracy, accrues extra importance due to the emerging role of artificial intelligence (AI). AI tools, primarily powered by large language models (LLMs), have seen rapid adoption, especially for software development [48, 83]. LLMs are now also widely used for document generation, from co-writing creative content [30, 55, 88] to drafting legal briefs and clinical trial documents [41, 67, 110]. It therefore stands to reason that developers, already making active use of AI for coding, may also adopt it for creating privacy policies. This raises additional concerns about the documents' accuracy, since LLMs are well known for their hallucinations [29, 64, 82, 86]. On the other hand, LLMs' expansive training datasets, likely including many laws and privacy policies, may in fact be helpful for crafting better privacy policies, if used diligently (e.g., with cross-checking and verification). Thus, results depend primarily on how developers use AI tools. This too is an important open question that represents the second focus of our study:

RQ2: How do developers create privacy policies with the help of LLMs?

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791323

To answer these questions and attain detailed insights into developers' privacy policy creation processes, we conducted our study, consisting of two main components. First, we recruited 20 mobile developers with prior privacy policy creation experience from multiple regions of the world (Asia, North America, Europe, Middle East, and Africa), interviewing them about how they currently create privacy policies. We found that they rely on a variety of approaches, including automatic generators, templates, examples from competitors, and LLMs. Their workflow often involves starting with an existing privacy policy and adjusting it to the particularities of their current project. We also observed developers tailoring their policies based on industry and regional regulations.

Second, we asked developers to demonstrate creating a new privacy policy for one of their apps using an LLM (Claude Sonnet 3.5). Setting this task for all participants, regardless of whether they presently used LLMs as part of their workflows, allowed us to observe novice behaviors alongside established practices. In doing so, we identified several distinct patterns in the way participants prompted LLMs, such as providing most information upfront or preferring iterative refinement—though nearly all participants used fewer than four follow-up prompts. Most participants exhibited medium or high levels of trust towards LLM outputs and indicated that they would be comfortable publishing them with their apps without any further validation.

Our work contributes insights into the real-world processes behind privacy policy creation by individual developers and small teams. It also illuminates the way developers, who are increasingly adopting LLMs, use them for the legally-consequential process of privacy policy creation. Our observations—that developers invest relatively limited effort but exhibit high levels of trust towards the results—serve to explain prior findings about inaccuracies in privacy policies and as a caution that the veracity of privacy policies may continue to be low as AI adoption grows.

2 Related work

Privacy policies have become the primary method to inform users about how their data is collected, handled, and shared. But developers often lack the the necessary knowledge and legal expertise to create effective policies.

2.1 Privacy policy analysis and generation

Analysis. Researchers have used a variety of techniques to analyze privacy policies, including knowledge graphs [28], static analysis for compliance checking [111], and crowdsourcing-based classification [121]. More recently, several studies have explored using LLMs for policy analysis tasks [24, 38, 89, 98, 106]. Others have focused on improving privacy policy comprehension for end users through summarization and simplified presentations [23, 113, 120]. However, these analysis and comprehension tools focus on understanding existing policies rather than helping developers create new ones. While policy analysis is important, our work focuses on the creation process rather than analyzing existing documents.

Generators. Researchers have explored different methods to automate privacy policy creation for mobile apps. Yu et al. developed AutoPPG which employed static code analysis and natural language processing to characterize Android app behaviors and identify data flows not disclosed in existing privacy policies [118, 119]. Building on this work, Zimmeck et al. created PrivacyFlash Pro, combining static code analysis with questionnaires for iOS apps [122]. Jain et al. introduced PriGen, using Neural Machine Translation to convert code segments that process personal information to privacy captions [52]. They also developed approaches for detecting privacy behaviors in Android apps, directly linking code behavior to privacy notices [51, 53]. Extending this code-based approach, Campbell et al. explored how static code analysis combined with deep learning models can generate privacy policy captions that enhance GDPR compliance [108]. This study complements the research by Gardner et al. on helping iOS developers create accurate privacy labels through interactive code analysis tools [33]. Shezan et al. developed NL2GDPR, generating GDPR-compliant policies from natural language descriptions of app features [96]. More recently, Sangaroonsilp et al. created a system that prompts developers with targeted questions to allow generating more comprehensive privacy policies than existing generators [92]. Despite these advances, privacy policies are continuously found to lack accurate disclosures about apps' privacy practices [14, 72, 91].

Generation with large language models. LLMs have opened new possibilities for automated policy generation. Pan et al. created SeePrivacy, a multimodal framework that generates contextual privacy policies for mobile apps [79, 80]. By combining GUI understanding with policy analysis, their system increased users' willingness to engage with privacy information. Bateni et al. proposed using deep generative models for creating privacy policy content, training models on annotated policies to automatically generate privacy data practices [20]. Complementing this work, Malisetty et al. evaluated quantized models for IoT privacy policy language generation, studying whether they could transform policies into simpler formats [66]. However, researchers have not yet studied how LLMs are used in practice for policy creation.

Evaluation of generation tools. Researchers have shown that automated privacy policy generators (APPGs) have limitations. Sun et al. found that online APPGs often produce policies that fail to cover essential privacy items and include unnecessary statements, mainly due to their inability to analyze actual application code [103]. Pan et al. examined 10 common APPGs and analyzed over 46,000 privacy policies available on Google Play, discovering that while approximately 20% appeared to be generated by automated tools, many failed to fully comply with applicable privacy regulations [81].

Privacy policy and code consistency analysis. In another line of work, researchers investigated the extent to which privacy policies accurately described apps' actual privacy behaviors. Wang et al. developed GUILeak, which automatically detects data leakage and analyzes corresponding privacy policy claims to see if they are violated [112]. Other researchers developed PoliCheck for a similar purpose and used it to analyze a large number of applications, finding that many of them failed to disclose their privacy-sensitive data flows to users [16]. Okoyomon et al. [72] and Andow et al. [15, 72] similarly identified many apps that shared personal identifiers with third parties without appropriately disclosing such practices. To help developers fix these problems early, Li et al. created an Android Studio plugin that gives real-time feedback during development about possible privacy issues and showed that it led to improving

developers' privacy practices [60]. More recently, Xie et al. developed an LLM-based approach for checking privacy policies were compliant with applicable regulatory provisions [115]. However, there is still a need for approaches that allow connecting privacy policy creation with code development more effectively.

2.2 Privacy policy creation as requirements engineering

A number of researchers have approached privacy policy creation through the lens of requirements engineering, treating policies as outputs that should be derived systematically from software engineering artifacts and legal requirements [17, 44, 56–58, 63]. Liao et al. argued that a comprehensive approach to privacy requires aligning technical system architecture with legal institutional design [63], while Anisetti et al. proposed using machine-readable privacy certificates to match privacy requirements with service guarantees [17]. Kosenkov et al. specifically addressed the alignment of GDPR and software engineering specifications, emphasizing that consistent requirements and system specifications are essential for compliance and achieving Privacy by Design through modeling GDPR content with original legal concepts [56]. Krstic et al. proposed a model-driven development methodology that incorporates privacy policies into system design, providing semantic-preserving model transformations that generate system implementations enforcing given privacy policies by design [57].

Other researchers have focused on deriving policy documents directly from system design and code artifacts. Leicht et al. proposed a tool-supported method utilizing data-flow diagrams collected during privacy and security threat analyses to create the basic structure of privacy policies [58]. Hjerppe et al. presented a technique for constructing Layered Privacy Language policy data from web service code bases with a static analysis tool [45]. And Mohammadi et al. developed an approach for assisted generation using textual patterns to support service providers in creating comprehensible policies that comply with legislation [70]. Herwanto et al. focused on privacy requirements engineering in agile contexts, noting that most methodologies focus on waterfall approaches, and proposed an NLP-based approach to identify privacy requirements in user stories and automate the creation of privacy requirements based on derived data flow diagrams [44]. These requirements engineering approaches represent efforts to systematically connect technical system design with privacy policy creation, though they remain largely academic prototypes rather than adopted industry practices.

2.3 LLM-assisted writing

Researchers have studied how writers interact with LLMs across different document-authoring contexts [93, 97]. For instance, Reza et al. conducted a systematic review of 109 HCI papers and interviews with 15 writers, finding that desired levels of AI intervention vary across the writing process, with content-focused writers prioritizing ownership during planning while form-focused writers value control over translating and reviewing [88]. Other research has found that LLMs can improve writing productivity by 14% on average, particularly benefiting novice workers [22, 30, 65]. On the other hand, studies, such as by Kim et al., have found that writers may over-rely on the LLM, often prioritizing its emotional expressions over their own [55].

Beyond general writing, researchers have explored how LLMs can be applied to create high-stakes documents in domains such as law, finance, and healthcare [34, 90], where documents including patent claims, legal reports, and medical records demand high reliability and precision [116]. However, significant challenges emerge when LLMs generate domain-specific documents, as they frequently struggle with factuality, often generating a mix of true and false information with factuality declining in later sentences and a rise in unsupported claims [25, 99, 109, 114]. When deployed in legal or medical fields, LLMs encounter limitations due to deficiency in domain-specific knowledge [49, 50], with studies revealing that baseline models achieved only 70-80% compliance and exhibited factual errors in 18-43% of cases when drafting informed consent forms [110] and showed limited performance (scores of approximately 40% or less) in clinical thinking, logic, and transparency when generating clinical trial protocols [67].

Legal professionals showed preferences for documents perceived as crafted by humans over those believed to be AI-generated [41]. While researchers have explored hybrid architectures combining LLMs with retrieval-augmented generation to address these limitations [59, 67, 110], privacy policy creation shares characteristics with these high-stakes documents, requiring precise legal language, regulatory compliance, and accurate disclosure of technical data practices where the challenges of over-reliance, decreased ownership, and factuality issues may be particularly problematic.

2.4 User-centered studies with developers

Beyond technical approaches, empirical studies have been conducted with developers to understand their behaviors and needs. They showed that developers struggle with basic privacy requirements regardless of the support provided.

Privacy compliance challenges. Senarath and Arachchilage found that developers struggled when asked to embed privacy into application designs [94]. Building on these findings, Alhazmi and Arachchilage found that developers lack familiarity with GDPR principles and dedicate their efforts to improving app functionality [12, 13]. This finding was further validated by Horstmann et al., who asked developers to implement GDPR-compliant code, finding that most of them submitted non-compliant solutions [47].

Communication and organizational issues. In addition to individual challenges, research has revealed issues within development teams. Tahaei et al. interviewed privacy professionals, finding that they struggled with negative privacy culture, limited tool support, and technical complexity [105]. Additionally, Grover et al. highlighted that while developers shape compliance through interpretation of requirements, they often reduce compliance work to one-time projects [40]. Experiencing communication challenges between developers and privacy experts is another factor that could hinder privacy efforts within product teams [46].

Privacy labels. Recent studies showed that developers experience challenges with creating privacy labels for iOS and Android apps [54, 62]. Li et al. examined developers' experiences creating Apple's privacy nutrition labels, identifying common challenges in the creation process [62]. They also highlighted that developers

need tools with interactive functionality and code analysis to guide them through how to disclose their apps' data practices in these labels [33]. Building on these insights, Khandelwal et al. studied developers working with Google's Data Safety Section creation, highlighting their challenges and the need for better resources [54].

2.5 Research gap

Despite major progress in privacy policy automation, there is a notable absence of human-centered research examining the actual practices of developers when crafting privacy policies, particularly in the context of emerging LLM technologies. While existing studies focus on creating automated solutions, they overlook how solo developers and small teams, who often lack dedicated legal resources, navigate this complex requirement in practice. Little is known about their workflows, the tools they combine with LLMs, or their strategies for handling regional compliance variations.

3 Methods

To find out how mobile developers create privacy policies for their apps, we conducted 20 semi-structured interviews with both Android and iOS developers.

3.1 Recruitment and participants

For this study, we were particularly interested in understanding developers' perspectives on creating privacy policies. Reasoning that large organizations are more likely to delegate this responsibility to their legal team, we primarily searched for developers who worked as freelancers or with small-to-medium sized development organizations. We also required that participants had worked on at least one privacy policy in the past year.

We initially attempted to recruit through the Google Play Store by emailing more than 2,000 developers but received no responses from this channel. We then focused our efforts on LinkedIn, Codementor, and ADPlist, messaging more than 1,500 mobile developers across these platforms [4, 6, 9]. We chose this approach because we could review developers' profiles and past experiences before contacting them. We used targeted searches to identify freelancers and developers at small-to-medium sized startups, using search terms like "mobile developer" combined with "freelance," "consultant," "founder," "small business," or "startup."

Privacy regulations vary internationally, which can affect privacy policy creation approaches. For this reason, we recruited and interviewed developers from five different regions: Asia, North America, Europe, Middle East, and Africa. We were able to recruit four developers from each region, for a final total of 20 interviews, which were conducted between May and August 2025.

Our sample consisted of 17 male and 3 female developers. All participants had at least four years of professional experience, with the majority having seven or more years. All worked on Android apps, and some also had iOS experience. They were employed across a variety of industries, including heavily regulated sectors like healthcare and financial services. Their current companies ranged in size from small startups to large organizations. (Most were from smaller organizations.) All participants also reported experience using LLMs in their work. The complete breakdown of professional backgrounds can be found in Table 1 (Appendix C).

Developers who participated in our study exhibited different levels of sophistication when it comes to the legal aspects of privacy policy creation. Most (15/20) had a basic understanding of privacy policy requirements, though nonetheless were fairly confident that their existing knowledge was sufficient to craft correct policies. Four participants differentiated themselves by demonstrating greater awareness of the legal implications of privacy policies. Rather than relying on themselves, they preferred to defer to formal legal expertise, consulting experienced lawyers for legal verification or using lawyer-provided templates. One participant worked directly in a compliance department with extensive legal knowledge, representing a technical-legal hybrid approach.

3.2 Interview

Through a pre-interview questionnaire, we collected participants' specific job roles, the types of apps they were involved in developing, the regions their products primarily targeted, their familiarity with LLM tools, and whether they had used traditional or AI-assisted methods for policy creation (Appendix A). We restricted participation to developers who were 18 years or older, had direct experience creating privacy policies, and were fluent in English. Those who satisfied our inclusion criteria were invited to participate in a one-hour online interview session.

We started the interviews by asking participants about their current processes for creating privacy policies, covering:

- teams involved in such processes,
- specific tools employed for privacy policy creation (e.g., templates, software libraries, and privacy policy generators),
- regulatory requirements considered when deciding on disclosures to include in privacy policies, and
- challenges they encountered.

We then focused on investigating how LLMs might guide them throughout these processes by exploring their:

- prior experiences with LLMs for privacy policy creation,
- prompting strategies,
- approaches to validating LLM outputs, and
- circumstances that led them to seek guidance from LLMs (e.g., handling app rejections from the Google Play Store).

We also asked participants a series of questions about how they handle data safety labels; because these privacy disclosures undergo more checking by app stores, we hypothesized that developers may approach them in ways that contrast with privacy policies. Ultimately, this portion of the interview did not yield substantially unique insights, and we refrain from discussing it separately. Our complete interview protocol is included in Appendix B.

A key component of our study involved participants demonstrating how they would generate privacy policies for real applications they had developed. We asked all participants to demonstrate using an LLM for this purpose, regardless of whether they currently used LLMs as part of their workflow. While this approach meant that some demonstrations would not reflect participants' typical workflows, we felt it was valuable to collect data both on the way users with experience approach this process as well as those who have never done it before. Since usage of LLMs for privacy policy creation is highly likely to increase, we can expect many more developers to follow in their footsteps, making it important to understand early

adoption patterns, capturing how developers might approach LLMs when they first decide to try them for privacy policy creation.

For this purpose, we developed an interactive web application that allowed us to observe participants' actual LLM usage in real time rather than relying only on their descriptions of their process. Participants demonstrated their complete process by thinking aloud while creating privacy policies for one of their apps using our web application. The web app allowed participants to interact with a large language model (Claude Sonnet 3.5) through a standard chat interface. Participants received login credentials that allowed them to access the application during the interview using their own computers. The application automatically captured and saved all prompts written by participants and responses generated by the LLM. Participants could also edit and revise the generated privacy policies directly within the application. Participants' prompts were passed directly to the model without any system prompts, preprocessing, or modifications by the researchers.

We conducted two pilot interviews with developers who were previously unknown to us. (These were not included as part of our main findings.) All interviews were conducted virtually using Zoom, with participants sharing their screens during the LLM demonstration portion [11]. Interviews were recorded with audio, video, and screen capture after obtaining consent from participants. We used Zoom's automatic transcription feature to generate initial transcripts, which we then manually cleaned and corrected.

3.3 Analysis

Two researchers independently built initial codebooks based on six interviews selected for their detailed content. They coded these interviews using ATLAS.ti software for qualitative analysis [5]. Then, they merged their codebooks through meetings where they discussed differences and resolved disagreements.

We applied the finalized codebook to the remaining interviews, with all interviews receiving double-coding by both researchers. The final codebook consisted of 70 codes that captured themes related to privacy policy creation processes, influence of organizational factors and regulatory requirements on such processes, and the resources that developers consult for guidance on policy generation (e.g., LLMs and privacy policy generators). The complete codebook is available in supplementary materials.

We separately analyzed the prompts and interactions captured through our web application. For each participant, we recorded the number of iterations they performed, distinguishing between the initial prompt and subsequent refinements (reprompts). We coded the construction approach for initial prompts and, for each reprompt, the type of modification requested. Our LLM analysis codebook additionally included codes related to LLM usage, such as developers' confidence in LLM outputs, the approaches they adopted to validate such outputs, and the challenges they faced when they interacted with LLMs.

After completing all 20 interviews, we determined that saturation was reached as no new themes were emerging from the data. While our study design integrated interviews and demonstrations in single sessions, we maintained analytical separation between self-reported practices (coded and analyzed as interview data, reported in Section 4.1) and observed behaviors (coded from screen recordings and prompts, reported in Section 4.2). This separation ensures analytical clarity while enabling triangulation during interpretation.

3.4 Ethics

This study received approval from our Institutional Review Board, including multi-regional recruitment. To ensure broad compliance with regional regulations, we followed principles including data minimization, allowing participants to review and redact their data, and deleting it on demand. Participants provided informed consent through both the screening survey and verbal consent at the beginning of each interview to record audio and video. We instructed them not to share any confidential information during the interviews. Our research design did not require confidential details about participants' work. We focused on general patterns and common practices rather than company-specific information. Participants received a digital gift card worth $20 USD as compensation for participating in the study.

3.5 Limitations

As with all qualitative research, our study is intended to provide insights on the range of real-world behaviors rather than be fully generalizable. Thus, while our participants represent a variety of company sizes, industries, and countries, we do not claim that our findings apply to all tech companies and teams. We provide participant counts for key themes as additional context rather than to support quantitative conclusions.

Participants were recruited based on their interest and willingness to help rather than through systematic sampling. We made efforts to achieve gender diversity in our sample, but the final sample was male-dominated. This aligns with current industry demographics, as developer surveys show that approximately 80% of developers are male [101].

Our study design asked all participants to create a privacy policy using an LLM, even though not all currently used LLMs for this purpose in their actual work. This means that the LLM usage patterns we observed may not fully represent the behaviors of experienced LLM users or reflect participants' typical workflows. However, this approach was intentional. With LLM adoption growing, we wanted to understand not only current usage but also early adoption patterns—how developers approach these tools when they first decide to try them. The behaviors we observed from participants without prior LLM experience for privacy policy creation likely represent what many more developers will experience as they begin experimenting with these tools. While our approach created two sub-populations (those with prior LLM experience and those without), we observed that their interaction patterns were substantially similar, including prompting strategies, follow-up behaviors, and overall trust levels.

The controlled setting of policy creation in our study differs from natural workflows where developers might consult colleagues, conduct additional research, or iterate over multiple sessions. This approach trades off ecological validity in order to isolate LLM usage, which enables controlled comparison across all participants. The limited duration of a study session represents another ecological limitation. However, we imposed no time limits, and all participants stopped when satisfied with their output rather than due to time

constraints. We also explicitly asked participants what they would do differently in real-life situations.

In another trade-off, we chose to have participants work with their own existing apps rather than hypothetical scenarios. This provided authentic technical knowledge about app behaviors, third-party SDKs, and data practices, which are necessary to evaluate LLM outputs. However, this also means that participants had pre-conceived expectations, having created a privacy policy for this app at least once. Overall, our study's design choices result in policy creation happening under conditions that are in several ways different from real-life circumstances. While still illuminating LLM interaction patterns, trust calibration, and validation approaches under these specific conditions, future work should consider more ecologically valid approaches.

Finally, our study design was exploratory rather than evaluative. We examined creation methods with and without AI but did not systematically compare them to each other, nor did we assess the legal compliance or quality of resulting policies. We document current practices and identify concerning patterns but cannot make causal claims about whether LLMs improve or degrade policy quality. Such evaluative questions require future controlled studies with systematic policy quality assessment [115].

4 Results

4.1 How privacy policies are created

We asked developers about their current workflows to understand how they approach privacy policy creation.

4.1.1 Responsible parties. The first step towards a privacy policy is deciding who is responsible for creating it. For solo developers, there is only one choice, unless they outsource the development. (Among our participants, none did.) In organizations with 2–10 employees (9/20 participants), a single person also typically had the primary responsibility.

In contrast, in mid-size companies with 11–200 employees (6/20), developers often generated initial drafts before sending them "to higher management for approval" (P12). In some cases, they collaborated with product managers and product owners, defining requirements and overseeing policy development. Team leads, senior developers, and CTOs guided the process or reviewed technical aspects in startup environments.

Large organizations (4/20) showed the clearest separation between technical and legal responsibilities. P11 explained that, in their current role,[1] developers are not involved as "it's the responsibility of a different team." This is typically the legal team, with dedicated compliance professionals handling drafting, review, and multi-jurisdictional compliance. P3 described having previously worked within a large organization's structured compliance department that included a "Senior Lead Consultant, IAM Policy Manager, and Cyber Security & Information Management Officer."

Additional specialized roles included documentation reviewers, who verified compliance materials before publication (P19); marketing teams, which participated in policy reviews due to data collection communication impacts (P19); and external consultants, who provided legal verification, particularly for sensitive data applications or regional compliance requirements (P8, P14, P15, P19).

Even when other teams are responsible for writing privacy policies, developers are still involved. P20 described their company's legal team requesting detailed input from developers: "They ask us to give them the data that we gather, the channels that we transmit those data to our server…and also if we are using any third-party library in the app that may or may not gather or transmit users' sensitive data."

In consultancies and agencies that develop apps for external clients, "it's more about the client size" (P18). For example, for smaller clients, P18 would gather information, create a draft, and send it for approval, whereas "for bigger companies, they usually have technical lawyers involved. Sometimes they even provide a company-specific template. So, in many cases, my role is just to review the document rather than write it from scratch."

4.1.2 Organizational context. Beyond determining who creates privacy policies, organizational context fundamentally shapes how developers approach this responsibility, manage risk, and allocate resources to compliance work.

Risk transfer and accountability. Freelance developers and contractors often employ explicit risk transfer strategies when working with clients. P5, who runs a development company with contractors, described telling clients, "this is what I've done in the past. I'm not a lawyer by any means, so this is just my strategy." While reassuring them that "I've been doing this for over 10 years and never ran into an issue," P5 instructed clients to proceed "at your own peril."

Resource allocation decisions. Solo developers and small teams consistently justified their approaches through explicit resource trade-offs. While acknowledging the risk that "someone can take you to court if things are not according to the law" (P10), they still felt that professional legal review was financially out of reach. Developers' multifaceted roles put a pressure on their own time as well: "in a small company, I'm usually the only one handling this stuff, so I need to be really efficient. I can't spend weeks going back and forth" (P19).

Approval process variability. Larger teams and organizations tend to have more structured processes, though we observed that they can also be flexible in response to internal communication patterns and timeline pressures. P12 described the unpredictable nature of mid-sized organization workflows: "sometimes, if I'm running late and the team is okay with it, I just generate the privacy policy myself. I either publish it directly or send it to the backend team or the website team to update the privacy policy section. But sometimes I need approval first, so I'll send it by email to someone in HR or to higher-level management, like the CEO. Sometimes there's someone responsible for reviewing documentation who checks to make sure everything is correct. So it goes back and forth." The variability within a single organization suggests that company size categories may obscure important process differences.

4.1.3 Prerequisite information. During drafting the privacy policy, developers need to collect the necessary information that will go into it, which includes the data collection practices of the app and any third-party dependencies.

[1] At the time of the interview, P11 worked at a large organization, but they also had extensive hands-on experience creating privacy policies in previous roles at smaller organizations.

The majority of participants (12/20) discussed identifying and categorizing the data their applications collect, including personal information, sensitive data, and activity logs. P8 described their process: "Usually, I start by listing out the features of the app and thinking through what kinds of data we collect…Then I ask myself: where is this data stored? Who can see it? Is it encrypted? Can the user delete it? That becomes the foundation for what we need to explain in the privacy policy."

While P8's information gathering was ad hoc, other teams rely on documentation that already exists as part of their normal project management workflow. P5, for example, reported: "we have a very detailed document of everything we're doing for this project, like a 50-page document listing every single feature."

With either strategy, the detailed technical knowledge needed highlights the crucial role that developers play in privacy policy creation, as they are the ones most familiar with the app's components, such as permissions, various SDKs, and their behaviors. As P20 pointed out, "No one has a better understanding than myself about the technology that I use."

A particular challenge for developers is accounting for the behaviors of third-party dependencies. Nearly all participants (19/20) discussed investigating third-party SDKs, such as Firebase, Google Analytics, AdMob, Crashlytics, payment processors, and social login providers. Some reported obtaining this information from the source, such as P13, who advocated, "the best way…is to go to the SDK site and read their explanation of how data is collected and stored." Others relied on prior knowledge and assumptions: "it's always clear that if you're using a Facebook SDK, they're definitely going to track people" (P7). Participants also acknowledged challenges when SDKs collect data they are unaware of.

4.1.4 Creation strategies. We observed mobile developers employ four distinct approaches when creating privacy policies: modifying templates or existing documents, using generator websites, prompting LLMs, and writing from scratch.

Template reuse and modification. The most commonly used approach (10/20) involves using existing privacy policies as a starting point, then customizing them for the current project. Sources of base documents include organizational repositories, competitor policies, and sample templates found online. For example, P18 shared: "Mostly, we use standardized templates from researchers who have conducted studies on mobile applications…There are people who study mobile app development and then publish templates, either for educational purposes or general reuse." P13 described a competitive analysis approach: "I looked for apps with a high number of downloads, a large user base, and sizable in-house teams…I would copy those privacy policies, check them, and rewrite them to fit my purpose."

Some participants build upon their own previous work. P6 explained their approach of reusing "already existing policies" and resorting to online templates only when an application is "totally unique." P10 also typically uses templates, explaining: "I have a template that I've already written for other apps. Then I just fill in the information," while P12 refers to their "old privacy policies for format and structure."

Policy generator websites. The second most common approach (8/20) involves using automated policy generation websites, such as Termly, TermsFeed, iubenda [7, 8, 10, 100]. These platforms typically guide users through systematic questioning; P20 described the process: "There is a wizard for completing these steps—the data that we store, the permissions that we get, the way that users can contact, and so many different information—and they will create a template based on the information provided." Participants appreciated the systematic nature of these tools, particularly their ability to ensure comprehensive coverage of regulatory requirements, praising them as "all-in-one solutions" (P11).

AI/LLM-assisted creation. An emerging trend (4/20 participants) involves using LLMs such as ChatGPT as the primary tool for privacy policy generation: "The first step is creating a basic draft of the privacy policy using ChatGPT" (P8). P5 described a more complex approach, recommending that one "take a competitor's privacy policy, input it into an AI tool like ChatGPT or Gemini, and prompt the LLM to generate a relevant policy for their business."

Participants expressed mixed feelings about policies that were generated using LLMs. P12 noted a key limitation: "The generators usually ask more questions, questions I might miss. ChatGPT just gives me what I tell it to give me…The generators, however, ask me questions and want me to choose from options, which helps generate a better privacy policy."

From-scratch development. Just one participant described creating privacy policies from scratch. P14 reported: "I wrote the privacy policy from scratch" for their personal project and "created the first version of my privacy policy by hand." While they did not use any generators, they mentioned examining similar applications for guidance and using ChatGPT for "the last stage of writing to polish" their text. This approach required significant time investment and careful research into relevant regulatory requirements.

AI for refinement and modification. Beyond primary creation methods, some participants (7/20) use AI tools for editing and refining existing privacy policies rather than initial generation. For example, P9 noted that they had previously used templates but "more recently started using LLM tools to help refine the formatting and structure." Others similarly used "AI to check what things can be improved or not" (P17) and "pick out points that a person may not have written correctly, and also catch syntax errors" (P10). P6 also described using LLMs to "polish the document," which entailed using prompts like "Do you think I'm going to have any issues with the Play Store?" They estimated that "the likelihood of getting a rejection email [is] maybe 30% if you haven't polished it" but drops to 10% after AI feedback.

4.1.5 Factors influencing AI adoption. Among those developers who have already adopted LLMs, the shift to AI reflects specific perceived advantages, though these are tempered by persistent concerns about reliability and legal adequacy.

Why developers turn to AI. For many participants, the decision to use LLMs came down to a simple calculation: time saved outweighed other concerns. P1 estimated that "something that used to take 2 or 3 days, I can now get done in an hour or even less," while P5 described the transformation as going "from like a 12-hour task, before AI…to just doing a little bit of prompting." This speed advantage was particularly compelling for startups and solo developers operating with limited budgets, where hiring legal counsel for every policy update was impractical.

Beyond speed, developers valued LLMs for handling the tedious aspects of policy writing. Rather than staring at a blank page, P11 appreciated that "at least ChatGPT gives me a pattern, and I know what I need to change. Doing modifications is comparatively easier than doing everything from scratch." Some used AI selectively for refinement rather than initial creation. P6 explained: "The only point we use it is when we want to polish the document." This approach allowed developers to focus on substantive decisions while delegating language refinement to the AI.

LLMs also proved valuable for reactive problem-solving. When faced with app store rejections or compliance issues, several participants turned to AI for rapid interpretation and revision. P1 described using LLMs to decode technical feedback: "It was helpful, it explained more details about what was happening, what the error was, and how I could solve it." P19 had made this a standard practice: "I usually copy and paste the exact error message or feedback from Google Play into the LLM."

Factors holding back AI adoption. Despite AI's advantages, most participants maintained reservations about relying on LLMs. The most common concern involved legal validity, since AI might "miss or simplify too much" (P8). Output quality presented another barrier, with several participants feeling that "the LLM output was too general" (P8) and that "it feels synthetic" (P14).

Another concern was whether information provided by LLMs would be up-to-date. P13 was particularly wary of SDK-related advice: "LLMs use a very old database, and they're not up-to-date with the current SDK versions." P19 emphasized the stakes: "Legal baselines change frequently…and not every AI model has the latest information." For domains with rapidly evolving requirements, P11 found it appropriate to maintain skepticism: "ChatGPT is not always right, and sometimes it uses outdated information."

Some participants preferred traditional policy generators because they removed the burden of remembering what to disclose. P12 contrasted the two approaches: "I'm more comfortable resonating with the traditional models because of the flow, it asks everything in a step-by-step process, which I find better." P19 identified the core risk: "The LLM relies entirely on what I remember to tell it. That's risky because I might forget something important." While LLMs offered flexibility, P2 noted that "in the previous method [generators], we have a checklist."

Finally, some participants encountered technical frustrations that undermined their confidence in LLM workflows. P19 described a particularly problematic pattern: "The constant changing of the document when using multiple prompts. After 6-7 prompts, the entire document might change completely, wasting hours of work."

4.1.6 Tailoring targets. After the core privacy policy was complete, participants reported engaging in additional tailoring, motivated by regional and industry regulations.

Regional regulations. Regulations like GDPR, CCPA, and many other national and state laws pose various types of requirements about what should be included in privacy policies [2, 3]. Most participants showed awareness of these requirements, and some mentioned that they incorporated regulatory compliance from the project inception, ensuring their information gathering covers multiple jurisdictions: "For example, if an application is published in three different regions, we have to create three different privacy policies for those regions" (P11).

For prominent privacy regulations like GDPR, awareness extended beyond European developers to those serving European markets. A participant explained their understanding of the technical implications: "If we're making an app for Europe, their policies apply. You have to keep the data there. So if you're storing data in a database, the server must be in Europe. You can't host that data in Asia or another region" (P17).

However, developers are not always well informed. Some, like P12, learned about these regional requirements through app store rejection experiences: "I've had a few issues with GDPR before. I wasn't aware of the requirements, and we initially planned not to release the app in Europe. But later, while publishing, we ended up including the Europe region, and the app was rejected by the Google Play Store." And P3, working in South Asia, highlighted knowledge gaps in GDPR implementation: "GDPR is basically a European concept. It is not implemented in South Asia yet; it is not compulsory…But the flip side is, if you have a client who is in the USA, UK, or Europe, basically, then they will force you to enforce it." They further noted misunderstandings about GDPR requirements: "In South Asian countries like India, where I live, they just think masking the username, password, phone number, and email is enough. It definitely is not."

Audience considerations. Participants consistently acknowledged that user demographics affect their overall privacy policy approach, with one explaining: "The user and audience do affect it. And definitely, if you are making it for children, then it has a different level of requirements than if you are making it for more mature people" (P15). This consideration includes accounting for user age, which can necessitate adherence to specific regulations like COPPA or GDPR-K [1].

However, not all participants found it important to customize privacy policies based on user demographics. One participant noted there was "not a major difference" when considering user base factors (P11), and another indicated that user demographics "doesn't matter" to their approach (P20). Additionally, one participant mentioned never working on privacy policies for applications with child users (P18), suggesting that the relevance of user base considerations varies depending on participants' specific project experiences and target markets.

Industry requirements. In addition to region- and audience-based requirements, developers also need to tailor their privacy policies to the app's industry. Examples include the healthcare sector, where "HIPAA is even stricter" than GDPR (P3). Financial applications also demanded extensive documentation, as P19 observed: "I've worked on finance apps where we had to collect [Know-Your-Customer] documents—national IDs, bank statements, all that sensitive stuff. The privacy policy for those apps ends up being like 10 pages long because you have to explain every single piece of data you collect, how it's encrypted, where it's stored, who has access to it." Through our entire study, we observed that differences in the tailoring processes across different industries (e.g., consumer apps versus finance) were more significant in scope than differences due to varying regional regulations.

4.1.7 Key insights from interviews. Developers rely on diverse approaches when creating privacy policies, including templates, competitor examples, and automated generators. The creation process is shaped by organizational size: solo developers handle all aspects themselves, while larger organizations separate technical and legal responsibilities. Developers are typically the ones with the most complete technical knowledge about their apps, yet they consistently struggle with gathering accurate information about third-party SDK behaviors, even when consulting official documentation. Regional regulations and industry-specific requirements drive policy customization, though developers are not always well informed about them. Some developers already incorporate LLMs into their workflows. While attracted by dramatic time savings, they remain concerned about legal validity and outdated information. Because LLMs rely entirely on what developers remember to disclose, many still prefer the structured questioning of traditional generators.

4.2 How developers create privacy policies with AI

In the second part of our interviews, we asked each participant to create a privacy policy for their app using an LLM. Their actions during these sessions illuminate patterns in how developers initiate, structure, and refine their interactions with LLMs throughout the policy generation process.

4.2.1 Initial prompts. When developers initiated their interaction with an LLM, we observed three distinct approaches based on the amount and type of information they provided upfront. Interestingly, the participants who previously used LLMs as their primary creation tools (see §4.1.4) demonstrated each one of these strategies.

Information-heavy initial prompt (13/20). The majority of participants employed a strategy of providing all known details in their initial prompt. This front-loading strategy typically covered app functionality, relevant regulatory requirements, data handling practices, and technical implementation details all in a single comprehensive prompt. Some participants employed a strategy where their initial prompts heavily featured detailed technical specifications related to third-party SDKs used in their applications. P2, for example, directly pasted the entire manifest file containing the project's dependencies into the LLM.

LLM-as-questionnaire (5/20). Several participants explicitly requested the LLM to guide the information-gathering process by asking necessary questions, explaining, "I'm not sure what kind of information I need to share with the AI" (P1). Another participant asked the LLM to function as an "expert at privacy policy generation" and to pose "relevant questions" while indicating their desire to "iterate until satisfied with legal considerations" (P16). This approach typically resulted in the LLM providing comprehensive questionnaires covering topics such as personal information collection, device permissions, third-party services, age restrictions, data storage practices, and geographic compliance requirements.

Reference-based refinement (2/20). A smaller number of participants utilized existing competitor policies as models for the LLM. One participant demonstrated this strategy by asking the LLM to create a policy using "URLs to Fiverr, Upwork, and Freelancer privacy policies as references from multi-million dollar companies to ensure legal coverage and similarity" (P5). Another participant would first detail their "app features and third-party libraries, then use a reference URL to regenerate or refine the policy's style and precision" (P2), specifically asking the LLM to make the policy "precise and near to the reference" after providing the external URL.

4.2.2 Follow-up strategies. Developers demonstrated varied approaches to refining the LLM-generated privacy policies, with most engaging in multiple rounds of prompting before deciding that they were done.

Number of follow-up prompts. Some participants (7/20) used minimal refinement (0–1 reprompts), relying heavily on initial prompts or references and finding the results immediately satisfactory. Many participants (11/20) made 2–3 reprompts, engaging in iterative gap-filling processes where they provided comprehensive initial prompts and then added missing technical details and compliance requirements that emerged after reviewing the initial LLM output. The remaining participants (2/20) engaged in multiple rounds of detailed adjustments (4-6 reprompts), addressing similar types of issues but more thoroughly.

Areas of refinement. Participants focused their improvements on a few specific areas where they found the LLMs' outputs deficient. *Technical and implementation details* frequently required additional specification, such as clarifying how permissions are used, data storage policies, and third-party service integrations. Many developers needed to add specific SDK information that was initially overlooked in their prompts. *Compliance and legal requirements* represented another major refinement area, with participants asking the LLM to update details about age restrictions, data retention policies, and compliance with additional regulations. *Business logic and app-specific features* required frequent refinement to reflect actual app functionality. Participants added details about payment processing methods, marketing consent management, cookie and tracking clarifications, and data deletion processes that were specific to their applications' operational requirements. *Format and presentation refinements* requested by participants included specific policy structures, tone modifications, and company information. These were motivated by both regulatory requirements and organizational communication standards.

4.2.3 Trust in output. When participants told us that they were done creating the privacy policy for their app, we asked them whether they would be comfortable submitting this policy to an app store in a real-life situation, as well as to rate their comfort with the final output on a scale of 1 to 10. We found that participants' comfort correlated with the degree of validation they wanted to perform. (More on validation below, in §4.2.4.)

High trust. The plurality of participants (10/20) rated their overall comfort in the 8–10 range, demonstrating high levels of trust towards the LLM's output. Those in this group included developers who had iterated extensively with the LLM, such as P7, who said, "I felt very comfortable because I forced the LLM to ask clarifying questions, which resulted in a more thorough output that prevented overlooking details." However, the group also included many who had fewer turns but were happy with the results from their initial input: "Not really [many edits needed], because my initial prompt was just nice" (P18).

Developers in this group suggested that they would not feel the need to perform much validation. Most commonly, their validation

strategy involved "just sending the first draft to Google Play and seeing how it goes" (P1). Nonetheless, even in this group, some acknowledged limitations of their knowledge, such as P11, who rated their comfort at 8 while explaining, "I'm not a lawyer, I have very limited knowledge. So, it looks good to me, but when an expert in this field looks at it, they will see it from a different perspective."

Medium trust. A significant subset of participants (7/20) showed only moderate levels of trust (6–7) towards the LLM's output, typically seeing some value but highlighting its limitations. For example, P20 reflected: "I found the output good for small apps but wished it could be more elaborated and that the LLM had asked more detailed questions, similar to how template generators function." Critiques most often focused on the policy being generic and missing details: "The output is too generic, and it's not completely aligned with actual app behavior" (P8). These issues encouraged those in this group to seek extra validation, such as P12, who noted: "I felt the output was too concise and lacked the detailed explanations provided by other policy generators, requiring me to cross-reference with old privacy policies to ensure completeness." In addition to referencing existing policies, another strategy from a medium-trust developer involved drawing on legal assistance: "it's better to be confident in the privacy policy by showing it to a lawyer, maybe someone who is responsible for that" (P14). Participants in this group were also more likely to make substantive changes to their policy while iterating on it.

Low trust. A small number of participants (2/10) indicated minimal trust towards the LLM output (1–5), stating "I would not publish this the way it is" (P6) and citing multiple reasons. P4, for example, had concerns ranging from formatting ("Yeah, just the text structure, text with some bullet points. [...] It's not a very user-friendly kind of text.") to end-user reactions ("I think users won't like it, and they might get stressed, like, 'Can I consent to this?' or 'Is it not safe? I don't want to consent?' "). Notably, both low-trust developers worked in regulated industries and preferred validation with multiple stakeholders besides themselves.

Factors influencing trust levels. We observed several patterns in the way trust varied across participant characteristics. All three participants from large companies exhibited high comfort with the policies. A potential explanation is that they had less direct familiarity with policies and were more used to relying on others for validation. In contrast, both low-trust participants came from medium-sized organizations, which often accord greater individual responsibility. As noted above, both also worked in regulated industries (finance and healthcare). However, other participants with experience in the same industries rated their comfort significantly higher, suggesting that industry alone may be an insufficient predictor of comfort, and there are probably other factors at play, like how much regulatory support someone has at their organization or their own personal experience with compliance issues.

4.2.4 Validation approaches. While study participants did not engage in the full range of validation activities they would perform in real life, they did tell us about the strategies they would employ.

Reliance on app stores. Many developers treat platform acceptance as their primary indicator of policy adequacy. As one participant explained: "You just submit it, and as long as you don't get a notification from Google asking you to change something you didn't do, then that's fine. So, Google is the one that determines whether it's done" (P7).

Comparing to existing policies. Some developers validate their documents by comparing them to existing policies from well-known companies and competitors who, they believe, would have documents that underwent legal vetting. Some use LLMs for this aspect of the process as well: "One more step I would apply is copying and sending the LLM the version of an existing privacy policy from our competitors, and finding the points that are missing in our privacy policy. Then I need to recheck them" (P13).

Checklists. Some developers utilize checklists to ensure all required elements are included in their policies. These checklists may be internal documents developed within their organizations or publicly available guidelines from platforms like Google Play and Apple App Store.

Manual refinement and editing. Even when using LLMs, developers perform extensive manual editing to refine language, add specific details, and ensure policies feel authentic rather than artificially generated. As discussed in §4.1.1, in organizational settings, drafted policies often undergo review by multiple team members beyond the developer who created them, such as product owners, managers, or other developers. When a legal team is available or in high-stakes situations, developers may further opt for a legal consultation: "before submitting it to Google Play, I'd make sure it also goes through the legal team" (P11).

Asking LLMs about issues. Some developers feed their drafted policies back into LLMs, requesting analysis of potential problems or areas for improvement. Some participants mentioned comparing outputs from different LLMs like ChatGPT and Gemini to identify discrepancies and validate the reliability of generated content. If their app was rejected from the app store, some participants also used LLMs to help understand complex rejection messages and develop appropriate responses.

Despite their reliance on LLMs, participants acknowledged the limitations of these tools, like their propensity for invalid assumptions: "AI makes a lot of assumptions, for example, on the age here, it just assumes it's suitable for people of a certain age, when actually that would depend on specific regions" (P18). Ultimately, regardless of specific validation approach, in the absence of automated legal validation tools, there is a high degree of uncertainty about policy adequacy: "There's no proper validation to check whether the privacy policy is legally 100% perfect or not" (P16).

4.2.5 Challenges. Despite the benefits that developers reported from using LLMs for privacy policy generation, our study revealed numerous challenges and limitations that affected the quality, accuracy, and usability of the generated policies.

Outdated information. Participants reported concerns about potentially outdated information: "Sometimes they get data from older sources, not up to date. Sometimes they get fresh data, but sometimes they hallucinate as well" (P10). This issue is particularly problematic given the rapidly evolving nature of privacy regulations: "Legal baselines change frequently for different jurisdictions, and not every AI model has the latest information, which could create potential legal issues" (P13).

Incorrect assumptions. Another recurring problem involved LLMs "making assumptions about things they might not have

enough information about, or might not know. For example, regarding permissions, data security, or data storage, they might assume things are one way when they could be different" (P9). These assumptions often led to the inclusion of irrelevant or incorrect information in the generated policies.

Users grew further frustrated when they failed to get the LLM to remove the incorrect information: "It brought in a lot of unnecessary points, or data that I do not collect. In a follow-up prompt, I explicitly mentioned: remove the date of birth, remove the technical data, I do not collect that. But it did not" (P6).

Third-party integrations and SDKs. Participants reported that LLMs consistently overlooked or inadequately addressed third-party components: "A common mistake is LLMs ignoring third-party content. For example, when using Google Fonts, the AI might not automatically include details about third-party service usage" (P19). This oversight created significant gaps in policy coverage, particularly for applications with multiple integrated services.

The challenge extended beyond simple omission to uncertainty about what data third-party SDKs actually collect: "One big challenge is tracking down exactly what data third-party SDKs collect or store" (P8). As a result, this uncertainty left participants in difficult positions where they acknowledged gaps in their knowledge: "If they're doing something additional, or if they're doing something we're not 100% sure about, that, to be completely honest, I guess we don't always know" (P5).

Lack of specialized knowledge. Participants working with specialized applications found that LLMs failed to recognize critical domain-specific obligations. A developer working on a psychology application noted: "The tool must recognize the purpose of the app, and because the prompt mentioned the term 'psychology,' it should have picked up that it's a psychological tool. And it should have adhered to the client-counselor confidentiality act" (P3).

4.2.6 Opportunities. After they were finished with the policy creation, we asked participants about features they would want to see in an ideal tool for generating privacy policies.

Interactive guidance. Most participants expressed a vision that was largely similar to existing (non-LLM) privacy policy generators, in that it would take a more proactive role in gathering information through "a wizard that asks the right questions" (P19). P19 further highlighted the advantage of this approach: "The generators also force you to think through scenarios you might not consider. They have these comprehensive checklists built in with different app types. The LLM relies entirely on what I remember to tell it" (P19). However, participants still valued that LLMs offered "more custom…more customized…more convenient" solutions (P7) and wanted to see tools that combine the best of both worlds.

Updated regulatory requirements. Corresponding to concerns about missing or stale legal information, many expressed a desire for guaranteed up-to-date and vetted legal details. Some even suggested retaining an expert human in the loop: "if a professional lawyer sees this and notices that there are flaws in it, then [they would have] the LLM to fix what the lawyer observed" (P15).

Automated analysis. Participants expressed enthusiasm for automatic app analysis to identify data collection practices: "an ideal tool […] should take all of your files, one by one, and then analyze them. Then, if you click, it will generate the policy" (P17). Notably, this is the exact behavior offered by a number of academic prototypes [52, 81, 118, 119, 122]. Our interviews suggest that their features are desired but production-ready implementations are unavailable or unknown to developers.

4.2.7 Key insights from LLM demonstration sessions. Most participants front-loaded information in comprehensive initial prompts rather than engaging in extended iterative refinement, with the majority making fewer than four follow-up prompts. Despite minimal iteration, participants exhibited high trust in LLM outputs, with most rating their comfort level at 8 or above on a 10-point scale. For validation, rather than seeking legal review, participants indicated that they would rely primarily on app store acceptance, comparisons to competitor policies, or asking LLMs to review their own outputs. LLM-generated policies exhibited persistent limitations, including outdated regulatory information, incorrect assumptions about app functionality, inadequate coverage of third-party SDK behaviors, and lack of domain knowledge for regulated industries. These observations suggest a significant gap between developer trust and the actual reliability of LLM outputs.

5 Discussion

5.1 Key implications about current privacy policy creation

As the first study of how mobile developers create privacy policies with and without AI tools, our results shed light on an evolving workflow and carry important implications for researchers, platforms, and end users.

Developers create legal documents without legal assistance. Our results confirm intuition and prior research: due to limited budgets, many privacy policies are crafted without legal assistance, by individual developers or small technical teams [12, 96]. Arguably, developers do not need lawyers in order to list all their apps' data flows and behaviors. However, previous studies have found that, in practice, developers struggle with privacy requirements because they are not familiar with legal standards [12, 94, 107]. This gap between technical and legal knowledge may result in incomplete or invalid privacy policies [20, 80]. As our results illustrate, definitive resources about what to include in privacy policies may not exist or may be unknown to policy creators, who may also simply not have time to identify applicable privacy laws and understand what they require [14]. Instead, developers *infer* the requirements from previously published policies, or *defer* this determination to others like generator websites or, increasingly, LLMs.

AI *and* non-AI approaches face trade-offs. While most developers in our study still used templates and automatic generators, some already used LLMs as their primary tool, and others used AI for refinement and modification of their drafts. Each strategy has advantages and limitations. Although developers value privacy policy generators due to their systematic questioning [33, 92], studies have shown that their outputs can still be incomplete and may not meet all necessary legal requirements [80, 103]. For their part, LLMs require users to be sophisticated in their prompting, otherwise they end up with generic policies disconnected from the reality of the applications. On the other hand, LLMs offer greater flexibility and

customization, allowing users to ask follow-up questions and introduce new resources and requirements. Going forward, hybrid approaches may be the most promising.

Third-party SDK handling remains a persistent challenge. Research has consistently traced many compliance violations to third-party SDKs and their unclear data practices [16, 69, 72, 87]. Our interviews confirm that developers struggle with this issue despite concerted efforts to investigate their apps' dependencies. Unfortunately, the use of LLMs injects greater uncertainty rather than improving the situation. LLMs are likely to provide confident claims about the behavior of any specific SDK. These assertions may be out-of-date, hallucinated, or in fact completely correct—but it may be very difficult for developers to distinguish these scenarios. Ideally, any claims about SDKs should be verified; however obtaining ground truth about SDK behaviors is challenging [16, 119]. To improve the situation, SDK publishers should provide more transparent, comprehensive, and comprehensible privacy documentation. Especially if published in machine-readable formats, it could be collected and consolidated for easier automated reference. Incentives from regulators may be needed to achieve these outcomes.

Developers rely on app store validation, which is nearly nonexistent. Many developers consider app store acceptance of their privacy policy as confirmation of its legal compliance. Both Google Play and Apple's App Store do perform some automated validation of data safety labels [35, 43, 75, 76] (though even those have been found to contain inaccuracies [32]), but there is minimal evidence of similar validation for privacy policy content [31, 43]. Google Play appears to perform basic checks such as verifying privacy policy URLs exist and are accessible [73, 74, 77], and Apple checks some privacy manifest file requirements and App Tracking Transparency compliance [36, 37, 78]. However, to the best of our knowledge, neither platform conducts substantive content validation of privacy policies [26, 27, 84, 85]. (We note that our—and developers'—understanding of this is shaped by online discussions, as neither platform clearly articulates how, if at all, they validate privacy policies.) This validation gap makes developers gain confidence from platform approval while actual compliance issues persist [15, 72]. To avoid this problem, app stores should be more transparent about whether and how they validate privacy policies. If, as we suspect, the current validation is minimal, they should consider introducing more automated checks, as they already do for security issues [39, 43].

LLMs' trust may be unearned. Developers in our study exhibited high levels of trust towards the privacy policies generated by LLMs, even in their initial versions: very few performed more than one or two follow-up prompts. Those with higher trust towards the output iterated even less on the results and were more likely to say they would go ahead and submit them to the app stores. While it is possible that the LLM got everything right on the first attempt, it seems likely that errors would slip in without more thorough vetting, which few seem inclined to engage in, even outside the artificial setup in our study. The behavior we are observing is analogous to what we see with AI users in other domains: when they receive output that looks plausible and passes any immediately-available tests, they are likely to accept it without any further scrutiny [95]. This resonates with findings from other writing contexts, where users were found to give LLM outputs high ratings on perceived competence [65] and exhibit over-reliance on the models [55]. However, unlike creative or conversational writing where imperfect outputs may be acceptable or even desirable for exploration, privacy policy inaccuracies about data collection or SDK behaviors can lead to regulatory violations. While prior work on AI-assisted writing has focused on preserving author agency or enhancing creative expression [55, 88, 93, 97], privacy policy generation demands different interventions centered on factual verification, technical accuracy validation, and legal compliance checking rather than authorship or style concerns.

Privacy policies should be treated with skepticism until verified. While many people skip over privacy policies, some people do read—or at least skim—them [102], as they are typically the only reference available about a service's privacy behaviors. Privacy policies are also an important instrument for regulators and researchers trying to understand app behaviors, as evidenced by the large volume of research on analyzing, interpreting, and explaining privacy policies [20, 53, 71, 79, 98, 108, 123]. By and large, these privacy policy consumers assume that what they are reading is accurate. Yet, research suggests that this is not the case [104, 117]. Prior work has also documented the time pressures and knowledge limitations that challenge developers navigating privacy decisions [12, 13, 47]. Our study extends this understanding by shedding light directly on the privacy policy creation process. We find that it is not always a meticulous, thorough procedure undertaken by well-informed experts. For many developers, it is as simple as opening ChatGPT and asking it a few questions. That the results may be incomplete will surprise no one familiar with LLMs. But, as a field, we have yet to fully accept the implications: we cannot assume privacy policies to be correct. Until they have somehow been validated, we should approach them the way we would treat any vibe-coded artifact: with suspicion.

5.2 Design implications and future work

Our findings reveal specific challenges and opportunities for designing LLM-based privacy policy creation tools that differ from traditional generators or templates. We outline design implications grounded in observed developer behaviors and pain points.

5.2.1 Overcome LLMs' limitations. As our study demonstrates, there are clear pitfalls in relying on LLMs for privacy policy generation, yet developers still see considerable value in using them. This suggests the possibility for system designs that combine the strengths of traditional generators with the advantages of AI in order to address pain points and improve developers' experiences.

Balance structured workflow with flexible inputs. Developers valued policy generators' systematic questioning but also appreciated the flexibility and customization LLMs offered. Future tools can provide checklists and follow structured workflows while integrating more conversational interfaces. For example, they could allow natural-language inputs so that users can provide more complex inputs, such as uncertainty, edge cases, and context-dependent scenarios. The tools can also allow users to ask follow-up questions based on the suggestions provided. This would preserve the comprehensiveness of traditional generators while maintaining conversational flexibility.

Enable retrieval of up-to-date information. Stale information is a large issue for those using LLMs for privacy policy generation, with the two primary areas of concern being SDK behaviors and the specifics of policies and regulations. However, systems can overcome this limitation by removing the reliance on the information encoded in the models. Instead, techniques like retrieval-augmented generation and web search can be used to query online sources, such as documentation websites and online repositories, in order to ensure that the tools can draw on up-to-date facts.

5.2.2 Support developer workflows. Currently, developers create privacy policies through a combination of tools and techniques that draw on disparate sources. They also consistently point to time as a limiting factor. There is an opportunity to make the process more efficient while also increasing the reliability of the end result.

Achieve a stronger understanding of the app's codebase. As we observed, the first step for developers is typically to get a handle on their own app's behaviors. This is something LLM-driven coding agents, such as Claude Code and OpenAI's Codex, are becoming increasingly proficient at. They can therefore be used to streamline the initial information-gathering phase to collect inputs for privacy policy generation. This approach can be used to augment—or even just to trigger—more principled app analysis frameworks developed by prior research [20, 66, 79, 118, 119, 122].

Streamline common scenarios. Our findings identify several workflows common to many developers, such as comparing candidate policies to previously published examples (from the developers themselves or from more established corporations). Design workflows could explicitly scaffold this process by building in semantic similarity comparisons to templates and already-published policies. Another manual workflow that can be automated is cross-comparing the recommendations of different tools, such as different LLMs and generators. LLMs can also explain why something should be included, not just that it should be, providing an educational component that may improve future policy creation. This can be further enhanced with built-in checks for common scenarios and proactive checks for inconsistencies.

5.2.3 Address research gaps. As initial exploratory research, our study leaves many open questions for future research to investigate.

Establish accuracy and compliance tests and benchmarks. A major problem for developers as well as researchers is that it is difficult to determine whether a particular privacy policy accurately reflects an app's behavior, and even harder to ascertain whether it complies with relevant regulations. More automated means of verifying these two properties would enhance the efficacy of both LLM-driven and non-AI tools. It would also unlock new research directions, such as quantifying the accuracy gap between LLM-generated policies and actual app behaviors across different levels of developer experience and company sizes, evaluating the effectiveness of different validation strategies beyond app store submission, and measuring improvements in LLM-generated privacy policies over time.

Continue researching the privacy policy creation context. While we observed that LLM adoption for privacy policy creation is emerging but not yet dominant, we did not thoroughly investigate what drives adoption decisions, as well as the reasons behind varying trust levels that developers exhibited. Understanding these factors would not only lead to better privacy policy creation tools but can also help developers appropriately assess their own confidence in LLM outputs. Thinking critically about when AI can be trusted is fast becoming an essential skill in the modern world. For developers, it can also lead to more trustworthy privacy policies.

6 Conclusion

We studied 20 mobile app developers from around the world and with a variety of backgrounds, each of whom had prior experience creating privacy policies. We found that they frequently create privacy policies without legal assistance and rely on templates, generators, and, increasingly, LLMs to fulfill legal requirements they may not fully understand. They rarely write policies from scratch; instead, they start either with existing policies (from similar services or their own prior projects) or with the output of a generator website or LLM. They then modify the draft based on the distinct features of the current project. At this stage too they may reference competitors' policies or prompt LLMs for assistance and modifications. Once the privacy policy has been written, validation approaches center on app store acceptance rather than legal verification. Platforms and SDK publishers could assist developers by providing transparent validation processes and machine-readable privacy documentation, and the research community could support them by developing hybrid tools that combine the systematic questioning of generators with the flexibility of LLMs.

Acknowledgments

We would like to thank our study participants for generously sharing their time, views, and valuable experiences.

References

[1] 1998. Children's Online Privacy Protection Act of 1998. https://www.govinfo.gov/app/details/USCODE-2011-title15/USCODE-2011-title15-chap91.
[2] 2016. Regulation (EU) 2016/679 (General Data Protection Regulation). https://eur-lex.europa.eu/eli/reg/2016/679/oj, L119/1–88 pages.
[3] 2018. California Consumer Privacy Act. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5.
[4] 2025. ADPList. https://adplist.org/.
[5] 2025. ATLAS.Ti. https://atlasti.com.
[6] 2025. Codementor. https://www.codementor.io/.
[7] 2025. Free Privacy Policy. https://www.freeprivacypolicy.com/.
[8] 2025. GetTerms. https://getterms.io/.
[9] 2025. LinkedIn. https://www.linkedin.com/.
[10] 2025. Privacy Policy Generator. https://www.privacypolicies.com/.
[11] 2025. Zoom. https://www.zoom.com/.
[12] Abdulrahman Alhazmi and Nalin Asanka Gamagedara Arachchilage. 2020. Why Are Developers Struggling to Put GDPR into Practice When Developing Privacy-Preserving Software Systems?. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*. USENIX Association. https://www.usenix.org/conference/soups2020/presentation/alhazmi.
[13] Abdulrahman Alhazmi and Nalin Asanka Gamagedara Arachchilage. 2021. I'm All Ears! Listening to Software Developers on Putting GDPR Principles into Software Development Practice. *Personal and Ubiquitous Computing* 25, 5 (Oct. 2021), 879–892. doi:10.1007/s00779-021-01544-1
[14] Noura Alomar and Serge Egelman. 2022. Developers Say the Darnedest Things: Privacy Compliance Processes Followed by Developers of Child-Directed Apps. *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2022, 4 (Oct. 2022), 250–273. doi:10.56553/popets-2022-0108
[15] Benjamin Andow, Samin Yaseer Mahmud, Wenyu Wang, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Tao Xie. 2019. PolicyLint: Investigating Internal Privacy Policy Contradictions on Google Play. In *28th USENIX Security Symposium (USENIX Security 19)*. USENIX Association, 585–602. https://www.usenix.org/conference/usenixsecurity19/presentation/andow.
[16] Benjamin Andow, Samin Yaseer Mahmud, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Serge Egelman. 2020. Actions Speak Louder than Words: Entity-Sensitive Privacy Policy and Data Flow Analysis with

PoliCheck. In *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, 985–1002. https://www.usenix.org/conference/usenixsecurity20/presentation/andow.

[17] Marco Anisetti, Claudio A. Ardagna, Michele Bezzi, Ernesto Damiani, and Antonino Sabetta. 2013. Machine-Readable Privacy Certificates for Services. In *On the Move to Meaningful Internet Systems: OTM 2013 Conferences*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Robert Meersman, Hervé Panetto, Tharam Dillon, Johann Eder, Zohra Bellahsene, Norbert Ritter, Pieter De Leenheer, and Deijing Dou (Eds.). Vol. 8185. Springer, Berlin, Heidelberg, 434–450. doi:10.1007/978-3-642-41030-7_31

[18] Apple. 2025. App Privacy Details - App Store. https://developer.apple.com/app-store/app-privacy-details/.

[19] Rawan Baalous, Alanoud Althobaiti, Dareen Alyoubi, Rama Alzahrani, and Mona Aljohani. 2025. Detecting the Inconsistency between Android Apps' Data Collection and Google Play's Data Safety Using Static Analysis. *Cybernetics and Information Technologies* 25, 1 (March 2025), 110–125. doi:10.2478/cait-2025-0007

[20] Nastaran Bateni and Rozita Dara. 2021. Automated Generation of Privacy Policy Using Deep Models. In *2021 IEEE International Symposium on Technology and Society (ISTAS)*. IEEE, 1–6. doi:10.1109/ISTAS52410.2021.9629155

[21] Wasja Brunotte, Larissa Chazette, Lukas Kohler, Jil Klunder, and Kurt Schneider. 2022. What About My Privacy? Helping Users Understand Online Privacy Policies. In *Proceedings of the International Conference on Software and System Processes and International Conference on Global Software Engineering*. ACM, 56–65. doi:10.1145/3529320.3529327

[22] Erik Brynjolfsson, Danielle Li, and Lindsey Raymond. 2023. *Generative AI at Work*. Technical Report w31161. National Bureau of Economic Research, Cambridge, MA. doi:10.3386/w31161

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

[24] Yuxin Chen, Peng Tang, Weidong Qiu, and Shujun Li. 2025. Using LLMs for Automated Privacy Policy Analysis: Prompt Engineering, Fine-Tuning and Explainability. doi:10.48550/ARXIV.2503.16516

[25] Cheng-Han Chiang and Hung-yi Lee. 2024. Merging Facts, Crafting Fallacies: Evaluating the Contradictory Nature of Aggregated Factual Claims in Long-Form Generations. In *Findings of the Association for Computational Linguistics ACL 2024*. Association for Computational Linguistics, 2734–2751. doi:10.18653/v1/2024.findings-acl.160

[26] Google Play Developer Community. 2023. App Rejected Due to User Data Privacy Policy Multiple Times. https://support.google.com/googleplay/android-developer/thread/249431313/app-rejected-due-to-user-data-privacy-policy-multiple-times?hl=en.

[27] Google Play Developer Community. 2024. App Rejected for Incorrect Privacy Policy, But I Checked 20 Times. https://support.google.com/googleplay/android-developer/thread/277005092/app-rejected-for-incorrect-privacy-policy-but-i-checked-20-times?hl=en.

[28] Hao Cui, Rahmadi Trimananda, Athina Markopoulou, and Scott Jordan. 2023. PoliGraph : Automated Privacy Policy Analysis Using Knowledge Graphs. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, 1037–1054. https://www.usenix.org/conference/usenixsecurity23/presentation/cui.

[29] Matthew Dahl, Varun Magesh, Mirac Suzgun, and Daniel E Ho. 2024. Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models. *Journal of Legal Analysis* 16, 1 (Jan. 2024), 64–93. doi:10.1093/jla/laae003 https://academic.oup.com/jla/article/16/1/64/7699227.

[30] Paramveer S. Dhillon, Somayeh Molaei, Jiaqi Li, Maximilian Golub, Shaochun Zheng, and Lionel Peter Robert. 2024. Shaping Human-AI Collaboration: Varied Scaffolding Levels in Co-writing with Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. doi:10.1145/3613904.3642134 https://dl.acm.org/doi/10.1145/3613904.3642134.

[31] Mozilla Foundation. 2023. False and Misleading Loopholes in Google's Data Safety Labels. https://www.mozillafoundation.org/en/campaigns/googles-data-safety-labels/.

[32] Mozilla Foundation. 2023. Privacy Not Included: A Buyer's Guide for Connected Products. https://www.mozillafoundation.org/en/privacynotincluded/articles/mozilla-study-data-privacy-labels-for-most-top-apps-in-google-play-store-are-false-or-misleading/.

[33] Jack Gardner, Yuanyuan Feng, Kayla Reiman, Zhi Lin, Akshath Jain, and Norman Sadeh. 2022. Helping Mobile Application Developers Create Accurate Privacy Labels. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, Genoa, Italy, 212–230. doi:10.1109/EuroSPW55150.2022.00028 https://ieeexplore.ieee.org/document/9799337/.

[34] Geetanjali Garg and Shobha Bhatt. 2025. Generative Large Language Models in Clinical, Legal and Financial Domains. In *Transformative Natural Language Processing*, Akshi Kumar and Saurabh Raj Sangwan (Eds.). Springer Nature Switzerland, Cham, 205–221. doi:10.1007/978-3-031-88988-2_9 https://link.springer.com/10.1007/978-3-031-88988-2_9.

[35] GitHub. 2024. App Rejected by Google Play Because "Data Safety Section: Location Data Type - Approximate Location". https://github.com/commons-app/apps-android-commons/issues/5708.

[36] GitHub. 2025. Apple App Store Rejecting Apps Because of "(ITMS-91061) Missing Privacy Manifest". https://github.com/python/cpython/issues/132006.

[37] GitHub. 2025. Request to Add Privacy Manifest File for App Store Compliance. https://github.com/jdg/MBProgressHUD/issues/668.

[38] Arda Goknil, Femke B. Gelderblom, Simeon Tverdal, Shukun Tokas, and Hui Song. 2024. Privacy Policy Analysis through Prompt Engineering for LLMs. doi:10.48550/ARXIV.2409.14879

[39] Google. 2025. 6 Ways Google Play Helps Keep You Safe. https://blog.google/products/google-play/keeping-google-play-safe-2025/.

[40] Rohan Grover. 2024. Encoding Privacy: Sociotechnical Dynamics of Data Protection Compliance Work. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–13. doi:10.1145/3613904.3642872

[41] Jakub Harasta, Tereza Novotná, and Jaromir Savelka. 2024. It Cannot Be Right If It Was Written by AI: On Lawyers' Preferences of Documents Perceived as Authored by an LLM vs a Human. *Artificial Intelligence and Law* (Dec. 2024). doi:10.1007/s10506-024-09422-w

[42] Google Play Console Help. 2025. Prepare Your App for Review. https://support.google.com/googleplay/android-developer/answer/9859455.

[43] Google Play Console Help. 2025. Provide Information for Google Play's Data Safety Section. https://support.google.com/googleplay/android-developer/answer/10787469?hl=en.

[44] Guntur Budi Herwanto, Gerald Quirchmayr, and A. Min Tjoa. 2024. Leveraging NLP Techniques for Privacy Requirements Engineering in User Stories. *IEEE Access* 12 (2024), 22167–22189. doi:10.1109/ACCESS.2024.3364533

[45] Kalle Hjerppe, Jukka Ruohonen, and Ville Leppänen. 2023. Extracting LPL Privacy Policy Purposes from Annotated Web Service Source Code. *Software and Systems Modeling* 22, 1 (Feb. 2023), 331–349. doi:10.1007/s10270-022-00998-y

[46] Stefan Albert Horstmann, Samuel Domiks, Marco Gutfleisch, Mindy Tran, Yasemin Acar, Veelasha Moonsamy, and Alena Naiakshina. 2024. "Those Things Are Written by Lawyers, and Programmers Are Reading That." Mapping the Communication Gap Between Software Developers and Privacy Experts. *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2024, 1 (Jan. 2024), 151–170. doi:10.56553/popets-2024-0010

[47] Stefan Albert Horstmann, Sandy Hong, David Klein, Raphael Serafini, Martin Degeling, Martin Johns, Veelasha Moonsamy, and Alena Naiakshina. 2025. "Sorry for Bugging You so Much." Exploring Developers' Behavior Towards Privacy-Compliant Implementation. In *2025 IEEE Symposium on Security and Privacy (SP)*. IEEE, 1215–1233. doi:10.1109/SP61157.2025.00146

[48] Xinyi Hou, Yanjie Zhao, Yue Liu, Zhou Yang, Kailong Wang, Li Li, Xiapu Luo, David Lo, John Grundy, and Haoyu Wang. 2024. Large Language Models for Software Engineering: A Systematic Literature Review. *ACM Transactions on Software Engineering and Methodology* 33, 8 (Nov. 2024), 1–79. doi:10.1145/3695988

[49] Quzhe Huang, Mingxu Tao, Chen Zhang, Zhenwei An, Cong Jiang, Zhibin Chen, Zirui Wu, and Yansong Feng. 2023. Lawyer LLaMA Technical Report. doi:10.48550/ARXIV.2305.15062

[50] Weijing Huang, Xianfeng Liao, Zhiqiang Xie, Jiang Qian, Bojin Zhuang, Shaojun Wang, and Jing Xiao. 2020. Generating Reasonable Legal Text through the Combination of Language Modeling and Question Answering. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, 3687–3693. doi:10.24963/ijcai.2020/510

[51] Vijayanta Jain. 2022. Creating Consistent Privacy Notices by Translating Code Segments into Privacy Captions. In *2022 IEEE 30th International Requirements Engineering Conference (RE)*. IEEE, 201–206. doi:10.1109/RE54965.2022.00024

[52] Vijayanta Jain, Sanonda Datta Gupta, Sepideh Ghanavati, and Sai Teja Peddinti. 2021. PriGen: Towards Automated Translation of Android Applications' Code to Privacy Captions. In *Research Challenges in Information Science*, Samira Cherfi, Anna Perini, and Selmin Nurcan (Eds.). Vol. 415. Springer International Publishing, 142–151. doi:10.1007/978-3-030-75018-3_9

[53] Vijayanta Jain, Sanonda Datta Gupta, Sepideh Ghanavati, Sai Teja Peddinti, and Collin McMillan. 2022. PAcT: Detecting and Classifying Privacy Behavior of Android Applications. In *Proceedings of the 15th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. ACM, 104–118. doi:10.1145/3507657.3528543

[54] Rishabh Khandelwal, Asmit Nayak, Paul Chung, and Kassem Fawaz. 2024. Unpacking Privacy Labels: A Measurement and Developer Perspective on Google's Data Safety Section. In *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association, 2831–2848. https://www.usenix.org/conference/usenixsecurity24/presentation/khandelwal.

[55] Taewan Kim, Donghoon Shin, Young-Ho Kim, and Hwajung Hong. 2024. DiaryMate: Understanding User Perceptions and Experience in Human-AI Collaboration for Personal Journaling. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–15. doi:10.1145/3613904.3642693

[56] Oleksandr Kosenkov, Ehsan Zabardast, Davide Fucci, Daniel Mendez, and Michael Unterkalmsteiner. 2026. Privacy by Design: Aligning GDPR and Software Engineering Specifications with a Requirements Engineering Approach. *Information and Software Technology* 190 (Feb. 2026). doi:10.1016/j.infsof.2025.107946

[57] Srdan Krstic, Hoang Nguyen, and David Basin. 2024. Model-Driven Privacy. *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2024, 1 (Jan. 2024), 314–329. doi:10.56553/popets-2024-0018

[58] Jens Leicht, Marvin Wagner, and Maritta Heisel. 2024. Creating Privacy Policies from Data-Flow Diagrams. In *Computer Security. ESORICS 2023 International Workshops*, Sokratis Katsikas, Frédéric Cuppens, Nora Cuppens-Boulahia, Costas Lambrinoudakis, Joaquin Garcia-Alfaro, Guillermo Navarro-Arribas, Pantaleone Nespoli, Christos Kalloniatis, John Mylopoulos, Annie Antón, and Stefanos Gritzalis (Eds.). Vol. 14398. Springer, 433–453. doi:10.1007/978-3-031-54204-6_26

[59] Haitao Li, Jiaying Ye, Yiran Hu, Jia Chen, Qingyao Ai, Yueyue Wu, Junjie Chen, Yifan Chen, Cheng Luo, Quan Zhou, and Yiqun Liu. 2025. CaseGen: A Benchmark for Multi-Stage Legal Case Documents Generation. doi:10.48550/ARXIV.2502.17943

[60] Tianshi Li, Yuvraj Agarwal, and Jason I. Hong. 2018. Coconut: An IDE Plugin for Developing Privacy-Friendly Apps. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (Dec. 2018), 1–35. doi:10.1145/3287056

[61] Tianshi Li, Lorrie Faith Cranor, Yuvraj Agarwal, and Jason I. Hong. 2024. Matcha: An IDE Plugin for Creating Accurate Privacy Nutrition Labels. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 1 (March 2024), 1–38. doi:10.1145/3643544

[62] Tianshi Li, Kayla Reiman, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I. Hong. 2022. Understanding Challenges for Developers to Create Accurate Privacy Nutrition Labels. In *CHI Conference on Human Factors in Computing Systems*. ACM, 1–24. doi:10.1145/3491102.3502012

[63] Kevin Liao, Shreya Thipireddy, and Daniel Weitzner. 2025. Data Traceability for Privacy Alignment. doi:10.48550/ARXIV.2503.09823

[64] Fang Liu, Yang Liu, Lin Shi, Houkun Huang, Ruifeng Wang, Zhen Yang, Li Zhang, Zhongqi Li, and Yuchi Ma. 2024. Exploring and Evaluating Hallucinations in LLM-Powered Code Generation. doi:10.48550/ARXIV.2404.00971

[65] Teresa Luther, Joachim Kimmerle, and Ulrike Cress. 2024. Teaming Up with an AI: Exploring Human–AI Collaboration in a Writing Scenario with ChatGPT. *AI* 5, 3 (Aug. 2024), 1357–1376. doi:10.3390/ai5030065

[66] Bhavani Malisetty and Alfredo J. Perez. 2024. Evaluating Quantized Llama 2 Models for IoT Privacy Policy Language Generation. *Future Internet* 16, 7 (June 2024), 224. doi:10.3390/fi16070224

[67] Nigel Markey, Ilyass El-Mansouri, Gaetan Rensonnet, Casper Van Langen, and Christoph Meier. 2025. From RAGs to Riches: Utilizing Large Language Models to Write Documents for Clinical Trials. *Clinical Trials* 22, 5 (Oct. 2025), 626–631. doi:10.1177/17407745251320806

[68] Colleen McClain, Michelle Faverio, Monica Anderson, and Eugenie Park. 2023. How Americans View Data Privacy. https://www.pewresearch.org/internet/2023/10/18/how-americans-view-data-privacy/.

[69] Abraham H. Mhaidli, Yixin Zou, and Florian Schaub. 2019. "We Can't Live Without Them!" App Developers' Adoption of Ad Networks and Their Considerations of Consumer Risks. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, 225–244. https://www.usenix.org/conference/soups2019/presentation/mhaidli.

[70] Nazila Mohammadi, Jens Leicht, Ludger Goeke, and Maritta Heisel. 2020. Assisted Generation of Privacy Policies Using Textual Patterns:. In *Proceedings of the 15th International Conference on Evaluation of Novel Approaches to Software Engineering*. SCITEPRESS - Science and Technology Publications, 347–358. doi:10.5220/0009371103470358

[71] Gabriel Morales, Pragyan K C, Sadia Jahan, Mitra Bokaei Hosseini, and Rocky Slavin. 2024. A Large Language Model Approach to Code and Privacy Policy Alignment. In *2024 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 79–90. doi:10.1109/SANER60148.2024.00016

[72] Ehimare Okoyomon, Nikita Samarin, Primal Wijesekera, Amit Elazari Bar On, Narseo Vallina-Rodriguez, Irwin Reyes, Álvaro Feal, and Serge Egelman. 2019. On The Ridiculousness of Notice and Consent: Contradictions in App Privacy Policies. In *Workshop on Technology and Consumer Protection (ConPro 2019), in Conjunction with the 39th IEEE Symposium on Security and Privacy*. https://dspace.networks.imdea.org/handle/20.500.12761/690.

[73] Stack Overflow. 2020. Android - Google Play Console App Rejection "APK Requires Valid Privacy Policy and Prominent Disclosure". https://stackoverflow.com/questions/64929059/google-play-console-app-rejection-apk-requires-valid-privacy-policy-and-prominen.

[74] Stack Overflow. 2021. Android - Google Rejected App Due to "Privacy Policy Link Invalid or Missing". https://stackoverflow.com/questions/66038709/google-rejected-app-due-to-privacy-policy-link-invalid-or-missing.

[75] Stack Overflow. 2021. App Store Rejection - Guideline 5.1.2 - Legal - Privacy - Data Use and Sharing. https://stackoverflow.com/questions/67351633/app-store-rejection-guideline-5-1-2-legal-privacy-data-use-and-sharing.

[76] Stack Overflow. 2021. iOS - App Tracking Transparency Privacy Checkboxes and App Store Release Rejection. https://stackoverflow.com/questions/68137030/app-tracking-transparency-privacy-checkboxes-and-app-store-release-rejection.

[77] Stack Overflow. 2023. How to Address Google User Data Privacy Policy Issue? https://stackoverflow.com/questions/75492066/how-to-address-google-user-data-privacy-policy-issue.

[78] Stack Overflow. 2025. iOS - Manually Created Privacy Manifest Appears to Be Ignored. https://stackoverflow.com/questions/79444233/manually-created-privacy-manifest-appears-to-be-ignored.

[79] Shidong Pan, Zhen Tao, Thong Hoang, Dawen Zhang, Tianshi Li, Zhenchang Xing, Xiwei Xu, Mark Staples, Thierry Rakotoarivelo, and David Lo. 2024. A NEW HOPE: Contextual Privacy Policies for Mobile Applications and An Approach Toward Automated Generation. In *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association, 5699–5716. https://www.usenix.org/conference/usenixsecurity24/presentation/pan-shidong-hope.

[80] Shidong Pan, Zhen Tao, Thong Hoang, Dawen Zhang, Zhenchang Xing, Xiwei Xu, Mark Staples, and David Lo. 2023. SeePrivacy: Automated Contextual Privacy Policy Generation for Mobile Applications. doi:10.48550/ARXIV.2307.01691

[81] Shidong Pan, Dawen Zhang, Mark Staples, Zhenchang Xing, Jieshan Chen, Xiwei Xu, and Thong Hoang. 2024. Is It a Trap? A Large-scale Empirical Study And Comprehensive Assessment of Online Automated Privacy Policy Generators for Mobile Apps. In *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association, 5681–5698. https://www.usenix.org/conference/usenixsecurity24/presentation/pan-shidong-trap.

[82] Gabrijela Perković, Antun Drobnjak, and Ivica Botički. 2024. Hallucinations in LLMs: Understanding and Addressing Challenges. In *2024 47th MIPRO ICT and Electronics Convention (MIPRO)*. IEEE, 2084–2088. doi:10.1109/MIPRO60963.2024.10569238

[83] Sanka Rasnayaka, Guanlin Wang, Ridwan Shariffdeen, and Ganesh Neelakanta Iyer. 2024. An Empirical Study on Usage and Perceptions of LLMs in a Software Engineering Project. In *Proceedings of the 1st International Workshop on Large Language Models for Code*. ACM, 111–118. doi:10.1145/3643795.3648379

[84] Reddit. 2024. Privacy Policy Explicitly Referenced in Apps. https://www.reddit.com/r/androiddev/comments/1e6y51p/privacy_policy_explicitly_referenced_in_apps/.

[85] Reddit. 2024. What Are the Requirements for a Privacy Policy Page of Your App? https://www.reddit.com/r/iOSProgramming/comments/1b7es5g/what_are_the_requirements_for_a_privacy_policy/.

[86] G. Pradeep Reddy, Y. V. Pavan Kumar, and K. Purna Prakash. 2024. Hallucinations in Large Language Models (LLMs). In *2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)*. IEEE, 1–6. doi:10.1109/eStream61684.2024.10542617

[87] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, and Serge Egelman. 2018. "Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale. *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2018, 3 (June 2018), 63–83. doi:10.1515/popets-2018-0021

[88] Mohi Reza, Jeb Thomas-Mitchell, Peter Dushniku, Nathan Laundry, Joseph Jay Williams, and Anastasia Kuzminykh. 2025. Co-Writing with AI, on Human Terms: Aligning Research with User Demands Across the Writing Process. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (Oct. 2025), 1–37. doi:10.1145/3757566

[89] David Rodriguez, Ian Yang, Jose M. Del Alamo, and Norman Sadeh. 2024. Large Language Models: A New Approach for Privacy Policy Analysis at Scale. *Computing* 106, 12 (Dec. 2024), 3879–3903. doi:10.1007/s00607-024-01331-9

[90] Yasmeen Saleh, Manar Abu Talib, Qassim Nasir, and Fatima Dakalbab. 2025. Evaluating Large Language Models: A Systematic Review of Efficiency, Applications, and Future Directions. *Frontiers in Computer Science* 7 (May 2025), 1523699. doi:10.3389/fcomp.2025.1523699

[91] Nikita Samarin, Shayna Kothari, Zaina Siyed, Oscar Bjorkman, Reena Yuan, Primal Wijesekera, Noura Alomar, Jordan Fischer, Chris Hoofnagle, and Serge Egelman. 2023. Lessons in VCR Repair: Compliance of Android App Developers with the California Consumer Privacy Act (CCPA). *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2023, 3 (July 2023), 103–121. doi:10.56553/popets-2023-0072

[92] Pattaraporn Sangaroonsilp, Hoa Khanh Dam, Omar Haggag, and John Grundy. 2024. Interactive GDPR-Compliant Privacy Policy Generation for Software Applications. doi:10.48550/ARXIV.2410.03069

[93] Steven Schirra, Sasha G Volkov, and Frank Bentley. 2025. "It's Something to Polish Your Own Thoughts, Rather than Create Thoughts for You": Understanding Participants' Use of Chatbots and LLMs During Online Research Participation. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors*

in Computing Systems. ACM, 1–6. doi:10.1145/3706599.3720027
[94] Awanthika Senarath and Nalin A. G. Arachchilage. 2018. Why Developers Cannot Embed Privacy into Software Systems?: An Empirical Investigation. In *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*. ACM, 211–216. doi:10.1145/3210459.3210484
[95] Raphael Serafini, Asli Yardim, and Alena Naiakshina. 2025. Exploring the Impact of Intervention Methods on Developers' Security Behavior in a Manipulated ChatGPT Study. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, 1–26. doi:10.1145/3706598.3713989
[96] Faysal Hossain Shezan, Yingjie Lao, Minlong Peng, Xin Wang, Mingming Sun, and Ping Li. 2022. NL2GDPR: Automatically Develop GDPR Compliant Android Application Features from Natural Language. In *2022 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 1–9. doi:10.1109/CNS56114.2022.10273858
[97] Momin N Siddiqui, Roy D Pea, and Hari Subramonyam. 2025. Script&Shift: A Layered Interface Paradigm for Integrating Content Development and Rhetorical Strategy with LLM Writing Assistants. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, 1–19. doi:10.1145/3706598.3714119
[98] Bhanuka Silva, Dishanika Denipitiyage, Suranga Seneviratne, Anirban Mahanti, and Aruna Seneviratne. 2024. Entailment-Driven Privacy Policy Classification with LLMs. In *2024 Conference on Building a Secure & Empowered Cyberspace (BuildSEC)*. IEEE, 8–15. doi:10.1109/BuildSEC64048.2024.00010
[99] Ava Spataru, Eric Hambro, Elena Voita, and Nicola Cancedda. 2024. Know When To Stop: A Study of Semantic Drift in Text Generation. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. Association for Computational Linguistics, 3656–3671. doi:10.18653/v1/2024.naacl-long.202
[100] Nishant Srivastava. 2025. App Privacy Policy Generator. https://app-privacy-policy-generator.nisrulz.com/.
[101] Statista. 2024. Developer Gender Distribution Worldwide 2024. https://www.statista.com/statistics/1446245/worldwide-developer-gender-distribution/.
[102] Nili Steinfeld. 2016. "I Agree to the Terms and Conditions": (How) Do Users Read Privacy Policies Online? An Eye-Tracking Experiment. *Computers in Human Behavior* 55 (Feb. 2016), 992–1000. doi:10.1016/j.chb.2015.09.038
[103] Ruoxi Sun and Minhui Xue. 2020. Quality Assessment of Online Automated Privacy Policy Generators: An Empirical Study. In *Proceedings of the Evaluation and Assessment in Software Engineering*. ACM, 270–275. doi:10.1145/3383219.3383247
[104] Yi Ping Sun. 2018. *Investigating the Effectiveness of Android Privacy Policies*. Ph. D. Dissertation. University of Toronto (Canada). https://utoronto.scholaris.ca/bitstreams/1e569d53-7f2b-4061-920d-a059998a40b4/download.
[105] Mohammad Tahaei, Alisa Frik, and Kami Vaniea. 2021. Privacy Champions in Software Teams: Understanding Their Motivations, Strategies, and Challenges. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, 1–15. doi:10.1145/3411764.3445768
[106] Chenhao Tang, Zhengliang Liu, Chong Ma, Zihao Wu, Yiwei Li, Wei Liu, Dajiang Zhu, Quanzheng Li, Xiang Li, Tianming Liu, and Lei Fan. 2023. PolicyGPT: Automated Analysis of Privacy Policies with Large Language Models. doi:10.48550/ARXIV.2309.10238
[107] Zhen Tao, Shidong Pan, Zhenchang Xing, Xiaoyu Sun, Omar Haggag, John Grundy, Jingjie Li, and Liming Zhu. 2025. Privacy Bills of Materials (PriBOM): A Transparent Privacy Information Inventory for Collaborative Privacy Notice Generation in Mobile App Development. *Proceedings on Privacy Enhancing Technologies (PoPETs)* 2025, 4 (Oct. 2025), 392–409. doi:10.56553/popets-2025-0136
[108] Trudy-Ann Campbell, Samson Eromonsei, and Olusegun Afolabi. 2024. Efficient Compliance with GDPR through Automating Privacy Policy Captions in Web and Mobile Application. *World Journal of Advanced Engineering Technology and Sciences* 12, 2 (July 2024), 446–467. doi:10.30574/wjaets.2024.12.2.0317
[109] Lifu Tu, Rui Meng, Shafiq Joty, Yingbo Zhou, and Semih Yavuz. 2025. Investigating Factuality in Long-Form Text Generation: The Roles of Self-Known and Self-Unknownan. In *Proceedings of the 2nd Workshop on Uncertainty-Aware NLP (UncertaiNLP 2025)*. Association for Computational Linguistics, 322–336. doi:10.18653/v1/2025.uncertainlp-main.27
[110] Julian Wang and Victor Xiaoqi Wang. 2025. Assessing Consistency and Reproducibility in the Outputs of Large Language Models: Evidence Across Diverse Finance and Accounting Tasks. doi:10.2139/ssrn.5189069
[111] Lun Wang, Usmann Khan, Joseph Near, Qi Pang, Jithendaraa Subramanian, Neel Somani, Peng Gao, Andrew Low, and Dawn Song. 2022. {PrivGuard}: Privacy Regulation Compliance Made Easier. In *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, 3753–3770. https://www.usenix.org/conference/usenixsecurity22/presentation/wang-lun.
[112] Xiaoyin Wang, Xue Qin, Mitra Bokaei Hosseini, Rocky Slavin, Travis D. Breaux, and Jianwei Niu. 2018. GUILeak: Tracing Privacy Policy Claims on User Input Data for Android Applications. In *Proceedings of the 40th International Conference on Software Engineering*. ACM, 37–47. doi:10.1145/3180155.3180196
[113] Justin Woodring, Katherine Perez, and Aisha Ali-Gombe. 2024. Enhancing Privacy Policy Comprehension through Privacify: A User-Centric Approach Using Advanced Language Models. *Computers & Security* 145 (Oct. 2024). doi:10.1016/j.cose.2024.103997
[114] Siwei Wu, Yizhi Li, Xingwei Qu, Rishi Ravikumar, Yucheng Li, Tyler Loakman, Shanghaoran Quan, Xiaoyong Wei, Riza Batista-Navarro, and Chenghua Lin. 2025. LongEval: A Comprehensive Analysis of Long-Text Generation Through a Plan-based Paradigm. doi:10.48550/ARXIV.2502.19103
[115] Qinge Xie, Karthik Ramakrishnan, and Frank Li. 2025. Evaluating Privacy Policies under Modern Privacy Laws At Scale: An LLM-Based Automated Approach. In *34th USENIX Security Symposium (USENIX Security 25)*. USENIX Association, 5797–5816. https://www.usenix.org/conference/usenixsecurity25/presentation/xie.
[116] Yongmin Yoo, Qiongkai Xu, and Longbing Cao. 2025. PatentScore: Multi-dimensional Evaluation of LLM-Generated Patent Claims. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 30715–30734. doi:10.18653/v1/2025.emnlp-main.1564
[117] Le Yu, Xiapu Luo, Xule Liu, and Tao Zhang. 2016. Can We Trust the Privacy Policies of Android Apps?. In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 538–549. doi:10.1109/DSN.2016.55
[118] Le Yu, Tao Zhang, Xiapu Luo, and Lei Xue. 2015. 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*. ACM, 39–50. doi:10.1145/2808117.2808125
[119] Le Yu, Tao Zhang, Xiapu Luo, Lei Xue, and Henry Chang. 2017. Toward Automatically Generating Privacy Policy for Android Apps. *IEEE Transactions on Information Forensics and Security* 12, 4 (April 2017), 865–880. doi:10.1109/TIFS.2016.2639339
[120] Shuning Zhang, Xin Yi, Shixuan Li, Haobin Xing, and Hewu Li. 2025. PrivCAPTCHA: Interactive CAPTCHA to Facilitate Effective Comprehension of APP Privacy Policy. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, 1–20. doi:10.1145/3706598.3713928
[121] Sebastian Zimmeck and Steven M. Bellovin. 2014. Privee: An Architecture for Automatically Analyzing Web Privacy Policies. In *23rd USENIX Security Symposium (USENIX Security 14)*. USENIX Association, 1–16. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/zimmeck.
[122] Sebastian Zimmeck, Rafael Goldstein, and David Baraka. 2021. PrivacyFlash Pro: Automating Privacy Policy Generation for Mobile Apps. In *Proceedings 2021 Network and Distributed System Security Symposium*. Internet Society. doi:10.14722/ndss.2021.24100
[123] Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven M. Bellovin, and Joel Reidenberg. 2017. Automated Analysis of Privacy Requirements for Mobile Apps. In *Proceedings 2017 Network and Distributed System Security Symposium*. Internet Society. doi:10.14722/ndss.2017.23034

A Screening survey

[After participants reviewed the study information and consent form and agreed to participate in the study.]

(1) Are you 18 years of age or older?
• Yes • No
(2) Do you meet the following conditions? (Check all that apply)
• Professional experience in creating privacy policies for Android apps • Fluent in English • Currently use or have used LLMs for policy generation • Have created privacy policies through traditional methods without LLMs • None of the above
(3) Are you able to commit up to one hour (via video conference) to participate in this study?
• Yes • No
(4) How many privacy policies have you worked on in the past year?
• None • 1-2 • 3-5 • 6-10 • More than 10
(5) How do you typically handle privacy policy development? (Select all that apply)
• I work on them myself • I work with a team • I delegate the task to others • Other (please specify)

(6) What methods do you use to create privacy policies? (Select all that apply)
• Writing from scratch • Adapting my own prior privacy policies • Adapting templates provided by my organization • Adapting publicly available templates • Adapting policies that have been published by other organizations • Using Large Language Models (like ChatGPT, Claude, etc) • Working with legal or privacy professional(s) • Other (please specify)
(7) Please briefly describe your typical process for creating or updating a privacy policy. (optional)
(8) How long does it typically take you to create a privacy policy?
(9) What types of Android apps do you typically create privacy policies for? (Select all that apply)
• Games • Social Media • Productivity • Entertainment • Education • Health & Fitness • Finance • Other (please specify)
(10) What regions do your products typically serve?
• The entire world • North America • Central and South America • Europe • Asia • Africa • Oceania
(11) What best describes your current position?
• Android Developer • Mobile Developer (Android + iOS) • Full-Stack Developer • Technical Lead • UX/UI Designer • Product Manager • Engineering Manager • Security Engineer • Privacy Engineer • Security or Privacy Manager • Compliance Professional • Product Counsel or other legal position • Other (please specify)
(12) What's the approximate size of your current company?
• Just me • 2-10 employees • 11-50 employees • 51-200 employees • 201-1000 employees • More than 1000 employees
(13) What industry sectors do you work in? (Select all that apply)
• Technology/Software • E-commerce • Healthcare • Finance/Banking • Consumer Services • Education • Industrial & Manufacturing • Government & Public • Other (please specify)
(14) How many years of professional experience do you have in your current job type?
• Less than 1 year • 1-3 years • 4-6 years • 7-10 years • More than 10 years
(15) Please provide the URL(s) of apps you have created privacy policies for
(16) Country? (Free text)
(17) Would you be interested in participating in a 60-minute interview to share your experience with privacy policy development?
• Yes • No

B Interview guide

[After the interviewer introduced themselves, explained what the research was about, and confirmed verbal consent from participants.]
Let's start with a bit about your background:

(1) Could you tell me about your role and background in creating privacy policies, including your technical and legal experience (if any)?

I'd like to understand your approach to privacy policy creation:

(2) What is your current process for creating privacy policies?
(3) Do you use any existing materials as a starting point? Like templates or policies you find online?
(4) How do you determine what needs to be included in a policy?

Let's talk about how you customize existing policies:

(5) Do you use any policy template libraries? For example, sites like Termly, TermsFeed, or iubenda, or maybe privacy policy generators from industry associations.

About your team structure:

(6) How does your company size influence your approach to privacy policy creation?
(7) What's your team structure for privacy policy development?
(8) What about the geographic regions you operate in - do different privacy laws like GDPR or CCPA change your approach?
(9) Does your user base (like having kids as users or operating in heavily regulated industries) affect your privacy policy approach?

Let's discuss your workflow and tools:

(10) What tools do you use in your privacy policy workflow? How do you use them?
(11) Do you do anything to validate the compliance of the policies you have generated?
If Yes: How do you handle it? What tools do you use?

I'm sure you're familiar with large language models like ChatGPT. Some people are using them to create privacy policies. In the next part of our interview, I want to ask you to try using an LLM to create a privacy policy, to understand what that process might look like.
The privacy policy you'll be putting together should be for a specific, real app that you've created.
Before we jump into the tool, I'd like to learn a bit more about the app you chose:

(12) Could you tell me the name of the app and what it does?
(13) Could you show me the app on the Google Play Store/App Store? I'd love to take a quick look at it.
(14) What third-party SDKs are you using in that app?
For example, any crash reporting tools, analytics platforms, advertising networks, or any other third-party services your app might be using?

For this exercise, let's use the app you picked and prepared information for.

(15) Could you walk me through your complete process of creating a privacy policy using LLMs - from initial information gathering to final review?
What I'd like you to do is:
• Think aloud as you work. Tell me what you're thinking and why you're making certain choices
• Show me how you'd typically write your prompts
• Explain any adjustments you make along the way
I'd like you to keep working on the policy until you're happy enough with it that you would use it if you were really submitting the app to the Play Store. And if you reach a stopping point for another reason, you can let me know too.
Do you have any questions before we start?

Thanks for showing me your process.
I have some follow-up questions for you.

(16) Would you be comfortable submitting this policy to the Play Store in a real-life situation?
- If so, why? How do you know it's ready?
- If not, why not? What would you want to do before submitting the policy?

(17) How did the experience align with your expectations?
(18) If you had to rate your comfort level with the final output on a scale of 1-10, where would it fall? Why?
(19) Did you find yourself wanting to make a lot of edits to what the LLM produced? What kinds of changes?
(20) If you could change one thing about how the LLM handled the privacy policy, what would it be?

Let's discuss any difficulties you've encountered:

(21) Overall what challenges and limitations have you encountered when using LLMs for privacy policies?
(22) Have you encountered any significant errors or inaccuracies? How did you handle them?
(23) What differences did you notice between your traditional method and using LLMs?
(24) Imagine there was an ideal tool for generating privacy policies. What specific features would it have?

One of the things you need to do when releasing an app on Google Play is filling out those Data Safety Labels - you know, that section where you declare what data your app collects and how it's used. I wanted to ask you some questions about this process.

(25) Could you describe a typical workflow when you're preparing the Data Safety Labels manually?
(26) What were some of the main challenges you faced when doing this?
(27) Has it ever happened that Google Play rejected or flagged your Data Safety Labels? What happened?
(28) Did you ever need to look up information when filling out the Data Safety Labels? What was it?

I'd like to understand how you might work with LLMs for creating Data Safety Labels. Let's try a quick exercise with a specific scenario:

(29) Imagine you're about to submit your app to the Play Store and need to complete the 'Data Sharing' section of the Data Safety Labels. Your app uses [the specific SDK they previously mentioned] and you need to determine what to disclose about third-party sharing.
Could you walk me through how you would use this LLM right now to help with that process?
Please think out loud as you work, and explain:
- How you're gathering the necessary information about data collection
- How you're approaching the LLM to generate the labels
- Any challenges you encounter

(30) If Google Play flags issues with your data safety declarations, do you turn to LLMs to help understand or fix the problem?
(31) Let's say you submitted your app and got rejected because Google Play detected that your app collects approximate location data, but you didn't declare this in your Data Safety form. They specifically mentioned 'Location Data Type - Approximate Location' as missing. How would you use an LLM to help resolve this specific issue? Could you show me how you'd prompt it to get better information about your app's location data practices?
(32) How do you generally approach reviewing and handling the outputs you receive from LLMs? Do you have any verification or quality assessment process you typically follow?

And about your LLM workflow:

(33) What specific LLM tools do you use daily?
(34) Why did you choose these particular LLMs?

Thank you so much for sharing your experience and insights with me today. Your input has been really valuable.

(35) Do you have any other thoughts about using LLMs for privacy or compliance?

C Participants

Table 1: Background of study participants.

ID	Industry	Years of Experience	Current Company Size (No. of Employees)	Region	No. of Privacy Policies Created (Past Year)
P01	Staffing & Recruiting, Retail	7–10	>1000	Middle East	3–5
P02	Tech/Software	4–6	2–10	Asia	3–5
P03	Tech/Software, E-commerce, Healthcare, Education, Government, Non Profit, MNC	7–10	2–10	Asia	6–10
P04	Tech/Software, Finance/Banking	4–6	201–1000	Middle East	3–5
P05	Tech/Software, E-commerce, Healthcare, Finance/Banking, Education	>10	11–50	North America	>10
P06	Healthcare, Finance/Banking	4–6	51–200	Africa	3–5
P07	Tech/Software	4–6	2–10	Africa	3–5
P08	Healthcare, Fitness	4–6	2–10	Europe	3–5
P09	Tech/Software, Education	4–6	2–10	North America	1–2
P10	Tech/Software, E-commerce, Healthcare, Consumer Services, Education, Industrial & Manufacturing	4–6	11–50	Asia	3–5
P11	Tech/Software	7–10	>1000	Europe	6–10
P12	Tech/Software, Education	4–6	11–50	North America	6–10
P13	Tech/Software, Healthcare, Finance/Banking	>10	2–10	Europe	3–5
P14	Tech/Software, Consumer Services, Education, Industrial & Manufacturing	7–10	2–10	Europe	1–2
P15	Tech/Software, E-commerce, Healthcare, Finance/Banking, Consumer Services, Education, Government & Public	7–10	2–10	Africa	>10
P16	Tech/Software, E-commerce, Healthcare, Consumer Services, Education	>10	11–50	Asia	3–5
P17	Tech/Software, E-commerce, Healthcare	4–6	2–10	Middle East	3–5
P18	Tech/Software, Consumer Service	7–10	>1000	Africa	6–10
P19	Tech/Software	7–10	11–50	Middle East	3–5
P20	Tech/Software	>10	>1000	North America	1–2