
# Evaluating Privacy Perceptions, Experience, and Behavior of Software Development Teams

Maxwell Prybylo and Sara Haghighi, *University of Maine;* Sai Teja Peddinti, *Google;* Sepideh Ghanavati, *University of Maine*

https://www.usenix.org/conference/soups2024/presentation/prybylo

This paper is included in the Proceedings of the Twentieth Symposium on Usable Privacy and Security.

August 12–13, 2024 • Philadelphia, PA, USA

978-1-939133-42-7

Open access to the Proceedings of the Twentieth Symposium on Usable Privacy and Security is sponsored by USENIX.

# Evaluating Privacy Perceptions, Experience, and Behavior of Software Development Teams

Maxwell Prybylo
*University of Maine*

Sara Haghighi
*University of Maine*

Sai Teja Peddinti
*Google*

Sepideh Ghanavati
*University of Maine*

## Abstract

With the increase in the number of privacy regulations, small development teams are forced to make privacy decisions on their own. In this paper, we conduct a mixed-method survey study, including statistical and qualitative analysis, to evaluate the privacy perceptions, practices, and knowledge of members involved in various phases of the Software Development Life Cycle (SDLC). Our survey includes 362 participants from 23 countries, encompassing roles such as product managers, developers, and testers. Our results show diverse definitions of privacy across SDLC roles, emphasizing the need for a holistic privacy approach throughout SDLC. We find that software teams, regardless of their region, are less familiar with privacy concepts (such as anonymization), relying on self-teaching and forums. Most participants are more familiar with GDPR and HIPAA than other regulations, with multi-jurisdictional compliance being their primary concern. Our results advocate the need for role-dependent solutions to address the privacy challenges, and we highlight research directions and educational takeaways to help improve privacy-aware SDLC.

## 1 Introduction

With the vast increase in privacy violations in the US and around the world [95], many countries have adopted new privacy regulations [66], such as the European General Data Protection Regulation (GDPR) [27]. With these new regulations, developers are under increased scrutiny while implementing privacy engineering solutions throughout the Software Development Life Cycle (SDLC) or face financial penalties. Many mobile apps are initially developed by a small team of independent developers with limited privacy expertise or access to legal/policy resources to make privacy decisions [6, 7, 63]. Research shows that this lack of access to privacy expertise leads to challenges in creating concise, accurate and consistent privacy policies [11, 13, 52, 65, 72, 77, 78, 96, 99], implementing privacy concepts throughout the SDLC - from early analysis to testing [26, 40, 63, 86], and distinguishing between privacy and security approaches, tools and regulations [7, 9, 21, 36, 40, 81, 82, 89].

In recent years, several approaches, including Privacy by Design (PbD), have been introduced to help developers incorporate privacy rules throughout the SDLC [17, 20, 31–33, 44, 45, 53, 55, 65, 80, 99]. However, few works examined the implementation of these solutions from the developers' perspective and their impact on privacy practices. Most studies focus on only a limited group of developers and overlook the broader SDLC roles and the unique challenges faced by each role (e.g., product manager when defining privacy requirements or the QAs when identifying privacy leaks) [21, 48, 51, 61, 91]. They also do not examine how factors such as legal expertise, regulations, and regional differences influence software teams' privacy perceptions and practices.

In this paper, we conduct a large mixed-method survey study on Prolific with 189 participants located in the US and 173 participants located in 22 other countries (in total 362), who are involved in various roles in the SDLC – including administrators (e.g., scrum masters, product managers), development and Quality Assurance (QA) teams, and information security/privacy experts. The non-US participants are located in EU+UK (132), South Africa (21), Mexico (15), Canada (3), and South America (2). Our goal is to identify the current state of privacy comprehension, practices, and behaviors in various SDLC roles, and the privacy gaps that have yet to be addressed. Our survey comprises of three parts: pre-screening questions (e.g., describing their product/customers), generic questions regarding participants' demographics (education, role, company size, etc.); and role-specific questions to examine their perceptions, experiences, and behaviors. We combine


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2024.
August 11–13, 2024, Philadelphia, PA, United States.

the participants' demographics (e.g., location data) with the role-specific responses to help determine:

– **RQ1:** Are there any differences in privacy perceptions among various roles, locations, and other demographics?

– **RQ2:** Does access to privacy experts (e.g., a Chief Privacy Officer - CPO) impact privacy perceptions and practices?

– **RQ3:** How do privacy practices and experiences vary according to SDLC roles, locations, and other demographics?

– **RQ4:** What is the degree of familiarity of different roles regarding privacy concepts, approaches, tools, and regulations?

To the best of our knowledge, this is the first study to conduct such a holistic evaluation based on the roles in the SDLC.

Our results show that participants have diverse perceptions/definitions of privacy, showcasing the need for a refined approach to privacy in SDLC. Scrum masters, product managers, and information security/privacy experts define privacy more in terms of limited disclosure, while developers and QAs perceive privacy as control over personal information. Our study finds a lack of adoption of most PbD strategies and other privacy techniques, such as Privacy Enhancing Technologies (PETs) and Privacy Impact Assessment (PIA), in SDLC. Most QA members rely on legal/privacy experts to protect users' data, and they lack privacy knowledge and expertise. Members of software teams are generally self-taught regarding privacy concepts, and most are not familiar with regulations that exist in the US, such as the California Consumer Privacy Act (CCPA) [35] and the Children's Online Privacy Protection Rule (COPPA) [29]. We also find that software teams face challenges in both understanding and adhering to privacy regulations, especially across multiple jurisdictions. These findings highlight the need for more privacy-focused education and training. Comparing regional-specific trends regarding the use of PETs, the creation of PIAs, or the presence of a CPO, we did not observe any differences among our participants, regardless of their location. This shows that privacy practices are primarily determined by the culture of the organization and are not influenced by various regulations across regions [5]. Our results highlight the challenges faced in various SDLC roles and advocate the need for role-dependent solutions to address them. Based on these findings, we outline research directions and educational takeaways to help improve privacy-aware SDLC.

## 2 Related Work

Understanding developers' privacy expertise and concerns has been explored in research through user studies with developers and analysis of developers' forums [21, 48, 51, 61, 91]. Tahaei et al. [85] and Horstmann et al. [48] conducted interviews with developers and privacy experts and identified factors such as poor privacy culture, tensions between privacy and other business rules, lack of proper communication between privacy experts and developers, lack of standardized privacy tools, and mismatch between the technical expertise of developers and privacy experts that impact how developers implement privacy. They also emphasize the role of privacy champions to minimize such barriers. In 2014 (i.e., pre-GDPR and CCPA), Balebako et al. [7] examined how app developers make privacy and security decisions and revealed that smaller companies exhibit fewer positive privacy and security behaviors. Their research emphasizes the need for simplified, cost-effective privacy tools such as privacy checklists, especially for small firms. Other studies [3, 48, 51, 57] with practitioners and developers highlight that while regulations impact practitioners' behaviors and corporate cultures, the developers and practitioners mostly rely on app markets to spot privacy issues, and they struggle with implementing and maintaining privacy labels, as well as leveraging third-party tools to maintain compliance.

The analysis of Stack Overflow (SO) posts shows that developers frequently query regarding PbD, compliance, and confidentiality [84, 91]. Delile et al. [24] compared privacy questions on SO with responses generated by ChatGPT to identify whether ChatGPT could be used as an alternative tool. Their results indicate that, in ∼30% of cases SO is more accurate than ChatGPT. Li et al. [59] and Parsons et al. [70] studied posts on several Reddit forums and identified that most discussions on personal data usage occur in response to external events such as Android OS changes or privacy laws.

These studies pinpoint developers' challenges in correctly implementing privacy requirements and maintaining compliance. Our work complements these efforts; however, it is the first study to assess privacy perceptions, practices, and knowledge of members of software teams involved in various roles in SDLC through a large-scale mixed-method approach. Prior work focused only on developers (i.e., programmers) in the US and a few countries, whereas we studied members from various SDLC roles (including product managers, QA, etc.) spanning 23 countries. In our work, we investigate how factors such as organizational aspects (e.g., the presence of a CPO) and participants' demographics (e.g., role, education, and location) impact privacy perceptions, experience, and behaviors of software teams. We also explore how frequently developers use online forums for privacy-related queries.

## 3 Study Design

In this paper, we aim to understand how members of software teams in small, medium-sized, and large companies (i.e., with <20, 21–100, and 100+ employees), implement privacy in their software applications and examine their level of privacy comprehension, expertise, practices, and behaviors based on various demographics (such as roles, location, education level, etc.). For this purpose, we first conducted a pilot study to evaluate our survey design and then a large-scale study with members of software teams in 23 countries. Our pilot study was completed in January 2023, while our large-scale study was done between February–April 2023.

Table 1: Breakdown of Participants Roles

| Role | Count |
|---|---|
| AD: Admin., Product Manager, Scrum Master | 70 |
| SD: Software Designer, Architect, Developer | 198 |
| QA: Software Tester, Quality Assurance Eng. | 40 |
| ISec: Information Security/Privacy Expert | 54 |
| **Total** | **362** |

## 3.1 Survey Tools

**Survey Creation** We utilized *Qualtrics* for survey creation, a platform supported by our university. Using Qualtrics, we customized our survey to individualize questions based on the participants' role (Q9) as defined in Table 1. For example, we asked developers about familiarity with PbD (Q39) and their use of forums such as Reddit (Q32), while information security members were asked about the management of access control, encryption algorithms, and certificates (Q60-Q62).

**Survey Platform Selection** We conducted the large-scale survey using Qualtrics integration on the Prolific [68, 90] platform, since it provides a higher pay rate and allows selecting from a more specific pool of participants with basic programming knowledge, in our case - software teams.

Tahaei et al. and Kaur et al. [56, 90] recommend using Prolific and MTurk for large-scale surveys. Although pre-screening via programming questions is recommended [23, 56, 74, 90], it has limitations: (a) overusing such questions could lead to automatically responding *correctly* without paying attention to the questions [90]; (b) in studies such as ours where the software teams include a variety of roles (e.g., product manager, QA, etc.) as well as with specific programming skills (e.g., JavaScript developers), having programming knowledge questions may bias the participants' pools towards more experienced developers in larger companies with traditional programming knowledge, preventing recruiting *novice* developers and those in other SDLC roles; and (c) AI tools like ChatGPT [15] are widely accessible and can handle code-based questions. Thus, these questions are no longer a strong barrier to screen participants. Our analysis of Danilova et al.'s [23] pre-screening questions with GPT-3.5 shows that the tool can answer the questions with 95% accuracy. We discuss how we mitigate these issues below and in Section 4.

**Conducting the Survey** Prolific maintains a pool of active participants who are regularly screened and vetted by the platform. In our survey, we decided on the sample size based on Prolific's guidelines (a minimum of 300 for a representative sample). We initially pre-screened the Prolific participants based on the following requirements: (a) to be at least 18 years old, (b) fluent in English, and (c) working in industries such as Graphic Design, Information Services, Data Processing, Product Development, Software, Video Games, etc. We used their industry (rather than their role) as a filter, since Prolific does not allow selecting participants based on role. We paid an average of $25.17/hr to those who completed the survey. After the initial pre-screening, we recruited 686 participants across both US and non-US pools. Out of the 686 participants who started the survey, 14 did not give their consent, and 295 did not finish the survey; hence, they were excluded from our analysis. We then conducted another filtering process to ensure that the participants work in the software industry and, in fact, have software development experience. We asked them, "Q4. In short, tell us about your product and who your customers are." We manually evaluated their responses and cross-checked them with Q6 (their post-secondary degree) and Q9 (their roles). We found that most of them are involved in software development activities such as *"I make a productivity app for Mac & Windows to record & share the user's screen."* We eliminated 15 participants as we could not verify their involvement in SDLC; for example, those with responses as *"NA"* or *"I sell home decor items. My customers are primarily women."* Following these steps, we ended up with a total of 362 participants for our final count.

## 3.2 Pilot Survey: University Students

Our pilot survey participants were our university's graduate students (who mostly have industry experience through internships and part-/full-time jobs) over the age of 18 from the disciplines of Computing and Information Science, Electrical and Computer Engineering, and Business, who had experience in software development, IT, or related fields. To maintain their anonymity, we did not collect any personally identifiable information such as their contact, names, or company names.

The goal of the pilot study was to gather initial insights and feedback before the deployment of our main study on Prolific. Upon the IRB approval, we launched the survey using Qualtrics. The survey consisted of 40 questions, including 13 short and 27 multiple-choice questions, which were derived based on our informal discussions with developers in small companies and prior gaps in research. We estimated that the survey takes ∼30-40 minutes to complete. Every participant was presented with the same set of questions regardless of their role on a software team. We used the responses to improve our large-scale survey (i.e., Subsection 3.3).

We received 45 responses but most were incomplete due to the survey's length and the diversity of questions. After discussing the study with the participants, we revised and shortened the survey based on participants' role in the SDLC.

## 3.3 Software Teams Survey

The main feedback we received from the pilot study was that the survey required too much time to complete (∼27 minutes). To address this limitation and to focus on capturing participants' perspectives related to their SDLC roles, we separated the survey questions according to the roles. This shortened the survey duration by 12 minutes and enhanced

the quality of the responses we received. We first asked all participants the same set of 10 questions that are partly related to demographics and the degree of privacy understanding. We then divided the remainder of the questions into four groups, one for each role defined in Table 1. Our breakdown loosely follows the SDLC phases, but we separated the Information Security/Privacy (ISec) roles from the Software Developer (SD) roles to evaluate the significance of security or privacy knowledge in our survey. Although "Others" role was an option, none of the participants selected it.

### 3.4 Survey Questions

The survey includes a mix of demographic, perception, experiential, and behavioral questions which are crafted based on our RQs (see Section 1) and the challenges identified in prior research regarding creating privacy-preserving applications, such as understanding privacy concepts [2, 9, 40], knowledge of regulations and establishing compliance [4, 28, 33], creating consistent and accurate privacy policies [12, 34, 52, 60, 69, 71, 72, 77, 96, 99], knowledge of privacy approaches and existing tools [16, 38–40]. The complete list of questions (except questions 1-3, which are the required Prolific identification questions and our consent form) is found here.[1]

*Demographic questions* collect basic information about the participants, such as age, education, their SDLC role, and the company size; e.g., "What areas/roles of the development team are you currently involved with?".

*Perception questions* aim to understand participants' perceptions toward privacy; e.g., "How do you define privacy?".

*Experiential questions* ask about their experience with privacy challenges and tasks; e.g., "What was the process for the Privacy Impact Assessment, and who was involved?".

*Behavioral questions* ask about the participants' behaviors and knowledge related to privacy; e.g., "List any privacy-by-design strategies you have used or know."

## 4 Ethics & Limitations

**Ethical Considerations** This research adheres to our university's ethical guidelines and was conducted with our Institutional Review Board (IRB)'s approval. All participants agreed to a thorough consent form that included information about the investigators, the risks, benefits, compensation, and confidentiality. All participants were informed about their voluntary participation, maintaining their right to withdraw at any time. No personally identifiable information was collected, and measures were in place to ensure the anonymity, confidentiality, and security of responses. The contact information of all investigators and the IRB team was also included. No participants contacted the investigators or the IRB about the study or the compensation.

[1] Survey questions: http://tinyurl.com/2p9n49e4

**Limitations** Like most survey studies, our analysis is based on participant self-report data and is affected by self-report bias, recall bias, and social desirability bias. Participants were informed during consent that the survey pertained to privacy due to our institutions' IRB requirement. This may introduce priming and self-selection biases. There is also recruitment bias as the Prolific user base may not fully represent the diverse population of SDLC individuals. We used multiple screening questions to ensure that recruited participants have experience in software development activities (Section 3.1). We adopted a conservative process to remove participants for whom we could not verify their SDLC involvement, however, we may have removed a few professionals. We also asked follow-up and write-in questions to ensure the multiple-choice questions were backed up with written facts. To mitigate the potential for survey responses being generated by AI tools like ChatGPT [15], we minimized open-ended questions in favor of multiple-choice formats and carefully scrutinized the write-in responses to remove those that appeared AI-generated. Short responses with typos and errors suggested that our responses were not AI-generated. Despite our efforts, AI-generated responses could affect the study's outcomes.

We carefully framed our questions so as not to prompt biased responses. However, we could not avoid one leading question that asks about the confidence in their companies' privacy and security measures. We aimed to reduce the bias by providing four options instead of a 'yes' and 'no', with the option to not answer. Additionally, the question follows their own definition of privacy, further helping minimize bias. We employ statistical analyses (like the chi-square test [37]) to ensure the broad applicability of our findings. To control for Type I errors in the presence of multiple hypothesis tests, we report our results after employing Bonferroni correction.

## 5 Study Analysis Process

Our survey results are organized around our research questions (RQs, see Section 1), focusing on various areas of privacy within the SDLC and across different roles. Our RQs examine the perceptions held, privacy experience and challenges, and privacy behaviors while considering the demographic breakdown (see Section 3.4) to provide additional context and to allow for a more nuanced understanding of the data. Our analysis follows a mixed-method approach, encompassing both quantitative and qualitative methodologies.

**Qualitative Analysis** We evaluate the descriptive and open-ended questions through open coding procedures and iterative processes. However, in our analysis, we used taxonomies and categories based on the current literature to classify the responses. For the open-ended question regarding the *definition of privacy*, the first two authors, independently, classified 50 responses based on the taxonomy of privacy introduced by Solove [79] and the examples and hypotheses from [41, 50]. Similarly, for the *usage of PETs*, we used PETs categories

from the literature [22, 62, 76]. The first two authors independently assigned categories for the first 25 responses. They then discussed their results, resolved the discrepancies, and created a guideline (see Appendix K). They continued with the rest of the responses, evaluated the agreements and resolved the disagreements in another round of discussion. Lastly, a third privacy expert examined the results to ensure their correctness and completeness. For the non-subjective descriptive questions e.g., *which PbD strategies they use*, one author categorized them based on the current literature, (e.g., privacy by design strategies [47], phases and roles in the SDLC [73] for PIAs) and the second author reviewed them for correctness.

**Quantitative Analysis** For the questions where our goal is to understand if a correlation exists between the demographics and the privacy-related perception, experience, and knowledge, we conducted statistical analyses. We used the Chi-Squared test [37] to determine whether there is a significant correlation between two categorical variables. For questions where the responses are on a Likert scale, we used the Kruskal-Wallace test [14]. For *perception*, *experience*, and *behavioral* questions, we hypothesize from our RQs that the size of the company, the presence of a CPO or a similar role, the education level, roles, and participants' location may impact their confidence in privacy/security measures, various privacy practices (such as the creation of PIA or privacy policies), and their familiarity with PETs, regulations, and usage of forums. To control Type I errors and avoid false positives, we use Bonferroni correction [75]. Since Bonferroni correction is very conservative and may increase Type II errors, we discuss the results with respect to $alpha = 0.05$ as well as the adjusted value (i.e., $\frac{0.05}{24} = 0.0021$, for our 24 statistical tests).

# 6 Findings

## 6.1 Survey Demographics

In our main study, we received a total of 362 responses (after filtering - see Section 3.1). 189 participants reside in the US and the other 173 come from 22 other countries (see Section 1). Table 1 shows a breakdown of participants' roles, with the majority (∼55%) in SD roles. As shown in Appendix B - Table 10, most participants identify as male, are below the age of 45, and have completed their BSc., with ∼61% in Computer Science (CS), Information Technology (IT), Data Science (DS), and Electrical & Computer Engineering (ECE) majors. This value includes the answers to "Others, please specify". Among those with a Business degree, 61% are in AD (e.g., product manager), and 28% are in SD roles. Among those in the "Other" degree category, 48% identified as SD, 19.5% as QA, 21.0% as AD, and 11.5% as ISec. The company sizes of <100 and 100+ employees are distributed almost equally.

## 6.2 Perceptions of Privacy

We seek to understand software teams' privacy comprehension by examining how they define privacy, their confidence in their company's practices, and if these differ based on roles or organization differences (i.e., RQ1&2).

### 6.2.1 Definition of Privacy

One of our key questions is, "How do you define privacy?". The responses were diverse, showing differing perceptions. Some participants defined privacy in terms of data security, highlighting the need to protect user data from unauthorized access. For example, one participant explained that *"It involves implementing measures to safeguard sensitive information, such as encryption, access controls, and data anonymization"*. Others described privacy from a user rights perspective: *"I define privacy as the ability to control all that is related to my information and to keep it from reaching someone who is unauthorized"*. Few responses incorporated legal compliance, with one participant defining privacy as: *"This involves being compliant with regulations and ensuring all data is protected with a least-privilege access model with ownership of the different part data sources with assigned data stewards"*.

To categorize the diverse definitions of privacy, we utilized Solove's taxonomy [79], that breaks down privacy into various categories based on the types of harm of a privacy breach. We chose Solove's taxonomy for two key reasons: (a) it provides a structured and detailed approach to understanding and analyzing definitions of privacy, which is essential with our wide range of definitions and perspectives; (b) it has been widely recognized and used in privacy research [8, 10, 46, 100]. We followed an open coding procedure to map the provided definitions with the taxonomy, as described in Section 5. Multiple classes for each definition were also possible. Figure 1 shows the mapping. (For a breakdown of Solove's Taxonomy see Appendix C - Table 11; the 'Blackmail' category did not apply to any participant's definitions.)

Figure 1 shows that the top frequently occurring categories are 'Disclosure', 'Increased Accessibility', and 'Insecurity'.

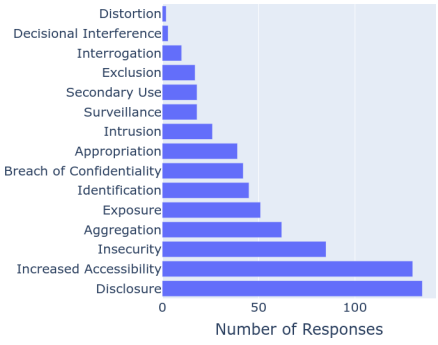


Figure 1: Privacy Definitions based on Solove's Taxonomy

Table 2: Distribution of Participants' Confidence

| Role | Yes | No | Unsure | PnS |
|---|---|---|---|---|
| **AD** | 53 (75.7%) | 6 (8.6%) | 11 (15.7%) | 0 |
| **SD** | 149 (75.3%) | 8 (4%) | 35 (17.7%) | 6 (3%) |
| **QA** | 25 (63%) | 2 (5%) | 12 (30%) | 1 (2%) |
| **ISec** | 44 (81.5%) | 2 (3.7%) | 8 (14.8%) | 0 |
| **Total** | **271** | **18** | **66** | **7** |

This result indicates that most participants either consider the traditional definition of privacy as *control over personal information* or perceive privacy in terms of *security*. For 'Disclosure', one participant highlights the importance of transparency and clear communication about data collection purposes and user control: *"Privacy is the assurance that all data belonging to an individual will be disclosed to others only with that individual's consent, for uses understood and approved by that individual."* For 'Increased Accessibility', a participant who works with genetic data underscores the need for controlled access to such information, only granting access if needed: *"The users' ability to define who can access their data and even in that what kind of data can be accessed. As I work in genetic data from patients in my line of work, the clinical information is always controlled access and only researchers working on the particular project can gain access on a need-to-know basis."*

We further examined how privacy perceptions differ across various roles. Almost 50% of participants in AD or ISec roles define privacy as 'Disclosure', while QA and SD roles mostly consider privacy as 'Increased Accessibility', which is related to *access control*. ISec roles mentioned 'Aggregation' more frequently than other roles, which is an anonymization technique used only in privacy rather than security.

> The variety in our participants' definitions of privacy shows the complexity of privacy perceptions, and the need for a holistic approach that covers a variety of aspects of privacy throughout the SDLC.

#### 6.2.2 Confidence in Security and Privacy Measures

We asked participants about their confidence in the privacy and security measures implemented in their organization. Table 2 shows the distribution of the participants and their responses. Note that 'PnS' stands for 'prefer not to say'. In all roles, most participants are confident in their company's security and privacy measures. Interestingly, ISec members are the most confident while the QA members are the most uncertain. This can be either due to QA members considering privacy and security as an afterthought [40], thus ignoring these requirements, or because they encounter more non-compliance instances during testing than any other roles.

We analyzed whether there is a correlation between participants' confidence in security and privacy measures and their demographic factors, such as the company's size (**H1a**), participants' roles (**H1b**), education level (**H1c**), and the presence of CPO or a similar position (**H1d**) (see Appendix D and Table 12 for more details). As shown in Table 12, with Bonferroni adjustment ($\frac{0.05}{24} = 0.0021$), we cannot reject the null hypothesis for **H1a**, **H1b**, and **H1c** ($p-value = 0.494, 0.654$ and 0.570); thus, we find no correlation between confidence in security and privacy measures and a company's size, participants' roles, or education levels. However, with a $p-value = 0.0007$ for **H1d**, we can reject the null hypothesis and say there is a correlation between the presence of a CPO (or similar position) and confidence in privacy and security measures. We further evaluate whether the existence of a CPO could lead to positive privacy outcomes in Subsection 6.2.3.

Table 3: Distribution of Company Size vs Existence of a CPO

| Company Size | Yes | No | Unsure | Others |
|---|---|---|---|---|
| **0–20** | 31.5% | 51.5% | 15.7% | 1.4% |
| **21–100** | 46.1% | 29.2% | 21.6% | 3.1% |
| **100+** | 47.3% | 34.9% | 17.8% | 0% |

We asked the ISec members specific questions regarding their company's security and privacy measures. When asked *"whether their company conducts security audits for third-party software used in their products"*, slightly more than half (∼56%) said 'Yes' while a large number (∼38%) were 'Unsure'. This is alarming since research shows a large number of third-party software and libraries include security and privacy vulnerabilities [1, 42, 97]. However, when we asked *"whether their company securely manages encryption keys and implements encryption algorithms and access control policies"*, more than 70% responded 'Yes' – which highlights inconsistencies in privacy practices even among experts.

> A CPO is important in fostering employees' confidence in the privacy and security measures of an organization.

#### 6.2.3 Presence of a Chief Privacy Officer (CPO)

To evaluate the impact of a CPO or other similar roles on privacy practices, we focus on the AD and SD roles, who are the majority of our participants (i.e., 268 (74%)). We did not include ISec and QA teams to avoid any response bias, due to their active privacy role in the company. We asked them "Do you have a Privacy Officer or similar position in your company?". Table 13 in Appendix E shows the distribution. Interestingly, only slightly more participants responded 'Yes' (42.6%) than 'No' (38.4%). ∼18% were 'Unsure', which may indicate the lack of proper communication among employees regarding the company's privacy practices and the purpose of a CPO. 1.1% responded 'Other', which included a legal team or a CTO. Among those that said 'Unsure', 23% are in AD and 77% are in SD roles which may indicate CPO members communicate more with the management team (i.e., AD).

We investigated whether the larger companies have a CPO. Table 3 shows the distribution of the presence of a CPO based on the company size. Here, we see the presence of a CPO increase with the company size. We also observe that companies of all sizes have a sizable number of 'Unsure' responses.

We further asked the participants "When you have a question about compliance with regulations, what do you do?". The participants could select more than one option. Appendix E - Table 15 shows the distribution of the responses. About half of the respondents (50.1%) mention they ask lawyers or a CPO, while 23.1% look at the best practices and standards (such as NIST guidelines), and 18.5% use developers' forums (such as Stack Overflow). Among 'Other' sources, they mainly mention 'search Internet' or 'ask a colleague'.

We analyzed whether the existence of a CPO (i.e., access to a legal or privacy expert) could impact the creation of PIA (**H2a**), the familiarity with PETs (**H2b**), the number of privacy breaches (**H2c**), or is influenced by the company size (**H2d**). Appendix E and Table 14 show the list of the hypotheses and the results of the tests. With Bonferroni correction, our results show that the presence of a CPO correlates with the size of a company ($p-value < 0.00001$). This correlation indicates that larger companies are more likely to have a CPO or a legal/privacy expert to help mitigate privacy risks, which is aligned with findings in [7]. However, with the p-value adjustment, we do not find a correlation between the presence of a CPO (or a similar position) and the creation of a PIA ($p-value = 0.1005$) and the use of PETs ($p-value = 0.008$). This may not be surprising, especially since a majority of SD roles, who are the main users of PETs and are involved in the PIA creation, are unaware of a CPO role. Our analysis also did not reveal a significant correlation between the presence of a CPO and the number of privacy breaches experienced by the organization ($p-value = 0.359$). This suggests that the presence of a CPO, while important and necessary, may not be sufficient to help minimize privacy breaches.

Although a CPO could improve confidence in a company's privacy measures, it has limited effectiveness in enhancing privacy practices and reducing breaches.

## 6.3 Experience with Privacy

We ask members of software teams in various roles about their *experience* with creating privacy policies and/or PIA, as well as practices to ensure the protection of users' data to better understand their privacy challenges (i.e., RQ2&3).

### 6.3.1 Creation of a Privacy Impact Assessment

A Privacy Impact Assessment (PIA) is a critical tool for identifying and mitigating privacy risks at any stage of software development. Recently, PIAs and their variations, such as the Data Protection Impact Assessments (DPIAs), have become a requirement in GDPR [27] and CCPA [35]. This tool allows organizations to address privacy and security issues before they become problems. In our survey, we asked participants if a PIA was created at any time throughout development, and if the answer is yes: at what stage it was created, who was involved, and what the process for creation was.

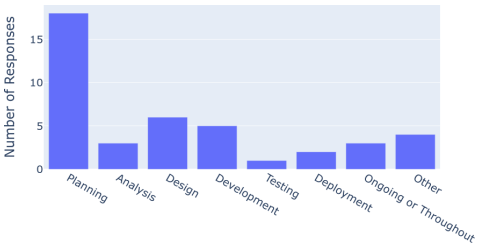


Figure 2: Stages of the SDLC When PIAs are Created.

We received 311 responses, where only 43 (14%) of them (where more than half were outside of EU+UK) reported that they created a PIA at any point in the SDLC, while a significant proportion (57.2%) reported that they did not (see Appendix F - Figure 5). This indicates a lack of awareness regarding the existence or the need for PIAs (i.e., the PIAs are non-existent or are conducted without their knowledge by the CPO or other teams). We also observed that ∼25% are unsure about whether a PIA was created, which may highlight a gap in communication within a company about its privacy practices. ∼4% chose 'Prefer not to say'.

Among the 43 who created a PIA, 3 (∼7%) did not answer the follow-up questions. Our results show that PIAs were created at various stages in the SDLC (see Figure 2), but ∼51.0% are at the planning and analysis stages. One participant who reported that a PIA was created during the planning stage said, *"At the start of development of idea because privacy is more important than all things"*. Some participants reported creating a PIA at the start of development, e.g., *"We created a [PIA] at the beginning of the software development process. This allowed us to identify potential privacy risks and develop strategies to mitigate them"*. Others mentioned during the design, or even towards the end of development.

The sizable number of participants (42%) involved in PIA during the later stages in SDLC may indicate that privacy requirements are not considered early on, and are only included as an afterthought – which is aligned with the findings in [40]. Furthermore, the variation in the timing of the PIA creation shows the need for a more standardized approach to incorporating privacy considerations into software development.

Figure 3 shows the distribution of the roles involved in PIA creation. More than one category was allowed. The responses are also diverse. Some participants reported that they created the PIA themselves or it was a team effort (i.e, SD & QA teams), while others reported that it was done by the CPO or external Legal team, ISec teams, upper management (i.e., CEO or CTO), or even the client (External). This shows that

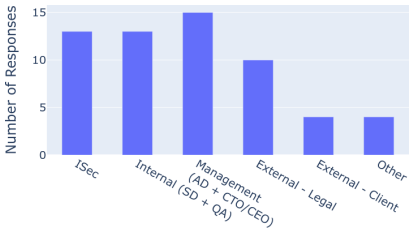


Figure 3: Distribution of Roles Involved in PIA Creation.

the responsibility for privacy can be distributed across various roles, which again highlights the need for clear communication, collaboration, and defined privacy practice processes.

The responses to the question regarding the process for creating the PIA also varied. Some initiate the process by downloading a template and collaborating with internal specialists, while others seek guidance from lawyers, executives, or third-party experts. A common approach involved consulting with professionals, with one participant mentioning that they *"outsourced [a] developer that specialises in data privacy and security"*. Several participants mentioned the involvement of specific roles, such as the Data Chief, IT teams, and privacy protection specialists. The process often involved cross-functional teams. In some cases, senior leadership, e.g., the CTO, CEO, or owner, played a pivotal role in the process.

Lastly, we evaluated the PIA correlation between the company size (**H3a**) and the participants' confidence in privacy and security measures (**H3b**) (see Appendix F). The results show a significant correlation $p-value < 0.0001$ for both tests - even after Bonferroni correction.

> Most members of software teams are not familiar with PIA or are unaware of its creation. However, those involved in PIA emphasize the need for its creation in the initial phases of the SDLC in a collaborative process with experts from various departments and consultants.

#### 6.3.2 Creation of Privacy Policies

Privacy policies describe how, why, and how long an application uses personal information. Regulations [27, 35] and the FTC [92] require companies to provide users with detailed privacy policies. Research shows that these policies may be inconsistent with apps [78, 99], since they are either created by outside legal experts (who may not fully comprehend the apps) or by using privacy policy generators [98].

We asked the AD team about their experience and challenges with privacy policies (as other roles are often only indirectly involved). Out of the 70 participants, 3 did not provide any answer. Of the rest, only 11 (17%) have been involved in the creation of a privacy policy, and they used 'legal experts' the most (64.0%), followed by 'templates' (45.5%), and 'privacy policy generators' (36.4%). More than one response could be selected. Two of them mentioned that they either 'search the Internet' or 'ask for team input', in addition to using privacy policy generators and templates. In 60% (out of 45.5%) of cases that used 'templates', and in 50% (out of 36.4%) of cases that used 'privacy policy generators', a 'legal expert' has also been selected. This result matches with prior research that legal experts in a company are mainly involved in the privacy policy creation, which may lead to inconsistencies between the app and the policy [78]. We also noticed that those who said 'Yes' are mostly from companies with less than 100 employees (∼64%) and with a CPO (∼55%).

Finally, we asked the 11 participants who responded 'Yes', "What challenges did you face when creating your privacy policy?". We received 10 responses. Six of them describe the challenges regarding compliance with regulations in multiple international jurisdictions, and understanding legal jargon, rules, and standards. One specifically had concerns regarding compliance, since they use privacy policy generators: *"...differences between different countries and their requirements since we are international."* Five respondents describe their main challenge as ensuring completeness (i.e., covering all personal information), soundness, and language of privacy policies. E.g.: *"Whether the wording I chose was going to cover all the bases I needed it to and whether it was clear and easy to understand."* or *"Which rules and text to inform users;..."* One of those five respondents was also concerned about which template to choose. Four others did not find the process challenging since they trusted the legal expert to help.

> Compliance with regulations, and ensuring completeness and correctness are among the most common challenges in creating a privacy policy. Software teams use several tools besides legal experts to help create privacy policies.

#### 6.3.3 Privacy Practices to Protect Users' Data

We tailored some of the privacy practice questions based on the role, specifically for ISec, SD, and QA teams. We asked ISec members: "How do you ensure that data collected from users is used only for intended purposes?". They discussed various approaches. ∼32% emphasized the importance of documentation to ensure transparency and accountability, with one noting *"the meticulous documentation of every step in the data usage process"*. Encryption emerged as a common theme, with participants mentioning sending encrypted documents and ensuring data is stored securely. A respondent states "*I would send documents encrypted and compressed into a zip file, and instruct them to delete the file once the information is accessed.*" Limiting access to data is another frequent approach, with 30.2% stressing the importance of restricting data access to only those who need it and maintaining logs to track any access. 16.98% stated the significance of transparency, ensuring they only collect necessary data, obtaining user consent, and regularly monitoring data usage. A few (9.43%) pointed out the importance of adhering to spe-

cific regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) [43] and the Family Educational Rights and Privacy Act (FERPA) [64]. 16.98% admitted to not having direct control over data but trusted their organization's protocols and training to handle data responsibly.

We also asked the same group "How do you manage access to sensitive user data in your organization?". Role-based access controls, multi-factor authentication, and encryption are common strategies employed to safeguard sensitive information. One respondent shared, "*We limit access to systems based on who really needs to access that data.*". Such measures ensure that only authorized personnel can access sensitive data, thereby minimizing potential breaches. Regarding data retention practices, only 47.17% of respondents state that they have been involved in removing user data either after its predetermined lifespan or upon user request.

We asked the SD members: "If you encounter a privacy concern at any point in the software development process, what steps would you take?". More than 95% of them take the concerns very seriously. For example, one participant mentions *"run a risk assessment"* and another mentions *"We take the app offline and start iteratively testing parts of the app to see where the privacy concern is."* About 36% deal with the concern internally to fix it and communicate it with the client and upper management. Another 35% directly escalate it to their supervisors, while 20% seek help from the ISec team or lawyers. A handful contact the client first.

We asked the QA team: "How do you verify that third-party systems used in your products are privacy compliant?". Similarly, we received diverse responses. Only 56.6% confirmed that their companies conduct security audits of these third-party systems. Some mentioned the significance of conducting vulnerability assessments and penetration testing to ensure third-party systems' compliance (23%). Some respondents discussed that they rely on reading privacy policies and contracts of third-party systems (28%), while others emphasized the importance of legal agreements and monitoring data transfers (15%). About 22% admitted to not being directly involved in this process, placing trust in their organization's legal and security teams, which is aligned with findings in [19].

Lastly, regarding the QA teams' practices for testing for privacy breaches and data leaks, they emphasized the importance of understanding the data they work with and always being vigilant about potential breaches. Regular manual or automated testing is a common theme. ∼27% of them mentioned the use of penetration testing, both internally and via third-party services. Others stressed the importance of using fake data during testing phases and ensuring that real user data is always encrypted and protected. ∼16% of respondents admitted to not being directly involved but trusted their organization's protocols and cybersecurity measures.

The most common privacy practices among SD, ISec, or QA teams are documentation, auditing, and security techniques (such as access control and encryption). QA teams rely heavily on legal and ISec teams to ensure data protection and are less involved themselves.

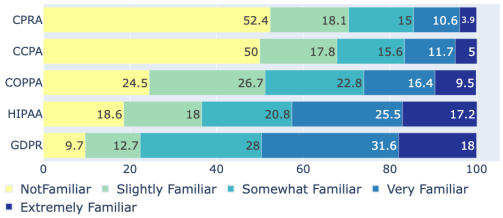


Figure 4: Familiarity with Different Regulations.

## 6.4 Privacy Awareness and Behaviors

We assess privacy *behaviors* based on familiarity with regulations, PbD, PETs, and such knowledge sources (i.e., RQ4).

### 6.4.1 Familiarity with Privacy Regulations

In recent years, several regulations have been introduced that developers need to comply with. Non-compliance with these regulations may lead to financial penalties, sometimes up to 4% of the annual turnover of the company [27]. However, these regulations include legal terminologies that may not be familiar to members of the software teams. To understand the degree of familiarity and awareness, we asked all 362 participants about their familiarity with GDPR, HIPAA, COPPA, CCPA, and the California Privacy Rights Act (CPRA). The answers are on a Likert Scale (see Figure 4).

We combined the results from 'somewhat familiar', 'very familiar', and 'extremely familiar' together and found that software teams' members, regardless of their region and roles, are more familiar with GDPR (77.35%) and HIPAA (63.26%). COPPA, CCPA, and CPRA are all 50% or below. ISec teams are the most familiar with all regulations among all roles, followed by the SD and AD teams. The QA teams are the least familiar with 7.5% familiarity with CCPA and CPRA, and 65.0%, and 57.5% with GDPR and HIPAA.

We asked participants "How did you learn about the previous regulations?". More than one option could be selected. As shown in Table 4, the majority are self-taught while university education ranks second. Among all roles, the ISec team has the highest percentage of learning about regulations through university education (33.3%), which is more likely through cybersecurity courses. We also asked the participants to describe the other sources they used to learn about privacy regulations. In most cases, they mentioned 'training at work' as the source; however, 2 participants mentioned 'social media' and 'YouTube' as their source.

Table 4: Distribution of Participants' Learning Experience

| Role | Self Taught | Lawyer | University Education | IAPP Cert. | Others |
|---|---|---|---|---|---|
| **AD** | 57.1% | 5.7% | 20.0% | 1.4% | 15.8% |
| **SD** | 56.6% | 3.5% | 21.7% | 2.5% | 15.7% |
| **QA** | 85.0% | 7.5% | 2.5% | 0.0% | 5.0% |
| **ISec** | 42.6% | 3.7% | 33.3% | 9.3% | 11.1% |
| **Total** | **57.7%** | **4.4%** | **21.0%** | **3.1%** | **13.8%** |

GDPR is the most familiar regulation among all participants due to its comprehensiveness. ISec teams are more likely to learn about regulations through university education; hence, are more familiar with them than other groups. QA teams are the least familiar.

#### 6.4.2 Familiarity with Privacy by Design (PbD)

Privacy by design (PbD) strategies introduced by Hoepman et al. [45] have gained interest in helping developers to be compliant with regulations. We asked the SD members (i.e., 198 participants) if they are aware of PbD, and if they answered yes, whether they used them (see Appendix G - Table 17) and to list the ones they used. ∼46% are familiar with PbD approaches while ∼25% are unsure, which indicates the potential knowledge gap and opportunity for educating developers. Out of those who answered 'Yes' to the awareness of the PbD question, only 57.1% had employed such strategies in their work. Of the remaining, 23.1% did not use them and 16.5% were unsure. This result suggests that even among developers who are familiar with such strategies, not everyone acts on this awareness – which may indicate the lack of usability and readiness of PbD for day-to-day developers' tasks [88] or other organizational factors, such as lack of resources [51].

Lastly, we evaluated the responses about the usage of specific PbD strategies (multiple answers were possible). Interestingly, our results are aligned with the findings of Tahaei et al. [87] (see Table 5). Our top categories are 'hide' (22), 'minimize' (21), 'inform' (17), and 'control' (12), while 'enforce' (1) and 'abstract' (2) are rarely discussed. One participant mentions *"Mostly minimise. Its the most straightforward."* This response reinforces our result in that 'minimize' is one of the easiest strategies to implement. We also received responses regarding Anne Cavoukian's PbD principles [17] such as 'privacy by default' (6 times) and 'proactive' (twice). The use of PIA was also mentioned 5 times as a strategy.

Our findings show that PbD approaches are not yet commonly used, and their lack of adoption underscores the gap in developers' knowledge regarding PbD and their usability in day-to-day developers' tasks.

Table 6: Usage of PETs in Software Development Process

| Privacy Enhancing Technology (PET) | Percentage |
|---|---|
| Encryption | 70.48% |
| Access Control/Identity Protection | 34.29% |
| Anonymity and Pseudonymity | 9.52% |
| Differential Privacy Approaches | 8.57% |
| Secure Communication/VPN | 8.57% |
| Privacy-Enhanced Anti Web Tracking | 0.0% |

#### 6.4.3 Use of Privacy-Enhancing Technologies (PETs)

Using PETs is another critical component of privacy protection, that allows better protection and maintenance of data privacy against outside threats. We asked the SD team, who are the main users of PETs, if they used any PETs, and if so to list them. Out of the 198 participants, 2 did not respond. 111 of them (56.63%) mentioned that they use some PETs while 36 (18.37%) do not. About 25% are unsure. These results are almost aligned with the degree of familiarity and usage of PbD. There was an increase (∼10%) in PETs familiarity and/or usage in comparison to PbD, which shows that these technologies are more common and tangible for developers, especially those related to encryption and access control. We grouped responses into 6 categories shown in Table 6 (definitions in Appendix K). Encryption and access control, which are primarily security-focused, were the most common, followed by anonymization methods and differential privacy.

Lastly, we investigated the correlations between the PETs' familiarity and the company size (**H4a**), confidence in security and privacy measures (**H4b**), and education level (**H4c**) (see Appendix H). With the adjusted p-value, we find no correlations ($p-value = 0.254$, 0.529, and 0.704, respectively).

PETs are slightly more commonly used than PbD strategies. However, there is still a gap in their familiarity, where more than 40% of developers do not use them or are unsure of their usage. The most commonly used PETs are more security-oriented concepts, than privacy.

#### 6.4.4 Developers' Sources for Privacy Information

As discussed in Section 2, developers sometimes seek privacy-related guidance on forums, such as Reddit or Stack Overflow (SO). We asked the SD teams how often they use various developers' forums for their privacy-related questions. Table 7 shows the distribution of the responses and their frequencies. ∼70% and ∼58% of the respondents use SO and GitHub *at least 1-3 times per month*, while for Reddit and Quora, this number is about 34.5% and 18.5%. About 57% of the respondents find these forums very or extremely useful, while less than 6% find them not useful at all. In cases where they do not find the answer on these forums, the SD team discusses their questions with the security or privacy experts, asks their

Table 5: Distribution of PbD Strategies Used by Developers

| Minimize | Hide | Separate | Abstract | Inform | Control | Enforce | Demonstrate |
|---|---|---|---|---|---|---|---|
| 21 | 22 | 7 | 2 | 17 | 12 | 1 | 4 |

Table 7: Frequency of Usage of the Developers' Forums

| Forums | Never | Rarely | 1-3/M | 1-3/W | Daily |
|---|---|---|---|---|---|
| **SO** | 13.1% | 17.1% | 26.1% | 24.1% | 19.6% |
| **GitHub** | 18.4% | 23.9% | 23.4% | 19.9% | 14.4% |
| **Reddit** | 30.5% | 35.0% | 20.0% | 10.5% | 4.0% |
| **Quora** | 54.5% | 27.0% | 12.5% | 5.5% | 0.5% |

teammates, or uses AI tools. In Appendix I, we provide a more detailed analysis regarding the usage of the forums.

Developers often seek privacy-related information from online forums, where more than 50% of participants use either Stack Overflow or GitHub at least 1-3 times per month and they find these forums useful.

## 7 Location Analysis

Our large-scale survey has responses from the US (189 responses) and non-US (173 responses from 22 countries: EU+UK, South Africa, Canada (CA), Mexico, and Chile), enabling us to examine differences in perceptions, experiences, and behaviors. We group the countries into three regions based on their similarities in privacy regulations: US+CA (192), EU+UK (132), and 'Other' countries (38). To evaluate the difference in *perception*, we examine whether participants' location correlates with their confidence in privacy and security measures (**H6a** in Appendix J) and the presence of a CPO (**H6b**). Both hypotheses do not hold ($p-values$ are 0.0567 and 0.6470). Table 8 shows the presence of a CPO across the three regions. The percentage of 'Yes' is almost equal between US+CA, EU+UK, and the 'Other' countries, while slightly more US+CA participants mentioned "no CPO" than elsewhere. This is not surprising since GDPR, the UK Data Protection Act of 2018, and the US HIPAA (Art.164.530) all require having a privacy officer or officials in a similar role.

We evaluated whether there is a significant difference between participants' *experience* in the three regions regarding the creation of PIA (**H6c**) and the number of privacy breaches (**H6d**). With $p-value$ 0.7724, we find no correlation for PIA.

Table 8: Distribution of Location-based CPO Presence

| Locations | Yes | No | Unsure | Others |
|---|---|---|---|---|
| **US+CA** | 43.7% | 41.5% | 14.1% | 0.7% |
| **EU+UK** | 41.7% | 36.1% | 20.3% | 1.9% |
| **Other Countries** | 43.5% | 30.4% | 26.1% | 0% |

Table 9: Distribution of Regulations Familiarity

| Location | GDPR | HIPAA | COPPA | CCPA | CPRA |
|---|---|---|---|---|---|
| **US+CA** | 71% | 84% | 53% | 48% | 44% |
| **EU+UK** | 89% | 37% | 38% | 11% | 9% |
| **Others** | 69% | 51% | 57% | 29% | 29% |

However, there is a correlation between the regions and privacy breaches ($p-value = 0.0010$). We also analyzed the *privacy behaviors* in the three regions concerning familiarity with PbD (**H6e**) and usage of PETs (**H6f**). With $p-values$ 0.3120 and 0.8588, we do not find any correlation that suggests that usage of PETs and PbD are equally (un)common in all regions. Since participants are from regions governed by different privacy laws, we investigated their familiarity with CCPA [35] (**H6g**) and GDPR [27] (**H6h**). As expected, we find a significant correlation between the participant's familiarity with the two regulations ($p-values$ are $< 0.0001$ and 0.0009 respectively). Due to the global reach of many apps, SDLC teams are responsible for complying with various regulations. We further evaluated the responses to the familiarity with each regulation in various regions. We combined the responses given for at least *somewhat familiarity* (i.e., somewhat, very, extremely familiar) and found that participants in the US+CA are most familiar with HIPAA while the rest are most familiar with GDPR. Those from 'Other' countries are also more familiar with the US regulations than those residing in the EU+UK. Table 9 shows the distribution.

## 8 Discussion

**Summary of Findings** Concerning *privacy perception*, our survey identifies that the majority of the participants define privacy in terms of control over personal information and disclose only when needed, or in terms of security. In other research [48, 85], data protection and security were the most common definitions. Having a CPO or a similar role positively impacts confidence in protecting users' data. However, we found out that a sizable portion of the participants are unaware of such a role in their company, which may lead to ineffectiveness in utilizing privacy tools or reducing privacy breaches. Lack of proper communication among various roles is a challenge that other research also identified [48, 85]. Our findings also align with [7] and [48], which observed a correlation between company size and having a CPO. However, we did not observe significant location-based differences in these perceptions. This is interesting but not surprising, since GDPR, HIPAA, the UK Data Protection Act, and Protection of

Personal Information Act (POPIA) all require a CPO or similar roles. Several of our US participants mentioned (in Q27) that they collect Protected Health Information (PHI), which falls under HIPAA; e.g., one participant says *"Health related data about people involved with our insurance companies"*. The extensive privacy requirements from these regulations likely explain why we observed no significant geographical differences in terms of participants' confidence, familiarity with PbD, and the usage of PETS.

In terms of *privacy experience*, most participants rely on legal experts to help create privacy policies; unlike [7] where creating a privacy policy was not the priority. Our study also shows that participants are primarily concerned about multi-jurisdictional compliance. Most of them are not involved in creating a PIA. The majority of those involved believe a PIA should be created during the planning or analysis phases; this is almost similar to findings in [40, 48]. Our participants emphasized the importance of detailed documentation regarding the data lifecycle, as well as using encryption and access control tools to protect the confidentiality and integrity of data. Interestingly, the QA teams rely more than others on security, privacy, and legal experts to implement and enforce privacy and security rules. Other studies did not examine the privacy practices of QA roles, separately.

Regarding *privacy behavior*, we identified that less than half of the participants are aware of PbD and an even smaller number use them. Similar to [87], 'hide', 'minimize', 'inform', and 'control' are more commonly used. The usage of PETs is slightly more prevalent than PbD, but the focus is more on security practices, such as encryption and access control; similar to other research that found security concepts are more tangible [7,40,48,85]. Anonymization techniques are not used frequently enough. We also find that although $\sim 53\%$ of our participants are from the US+CA, most are more familiar with GDPR than US-based regulations such as COPPA and CCPA. ISec experts are among the most knowledgeable about various regulations, while QA teams are the least familiar. Other works focus mainly on GDPR and CCPA and do not explore details regarding participants' familiarity [7, 40, 48, 85]. Most participants tend to seek answers to their privacy questions from developers' forums in addition to legal/policy experts; unlike [7] where they used 'friends' or 'social media'.

**Research Directions** Insights from the related work and our survey results highlight the need for approaches to operationalize PbD strategies and incorporate them into design and development. PbD patterns [93, 94] provide detailed information about their usage and high-level solutions, but still lack implementation. Approaches that detect privacy behaviors in code [53, 55] and further link them to patterns, or leverage automated code generation techniques to generate code from privacy patterns are yet to be explored.

Our survey highlights software teams' challenges in creating accurate PIAs and privacy policies. Research directions that focus on automated approaches to detect the information types, privacy practices, and purposes pre- [49] and post-development [53, 55], or to generate privacy statements from code [54] could alleviate the challenges regarding accuracy, consistency, and compliance.

Developers seek answers to their privacy-related questions from developers' forums, though increasingly use tools such as ChatGPT [15, 67]. However, these tools may not always provide accurate responses [24]. Developing methods to help translate developers' privacy-related questions into accurate privacy code snippets requires further attention [30].

Our survey indicates that software teams face challenges in understanding and adhering to privacy regulations; thus, there is a need for approaches to help better understand such regulations, and establish and maintain compliance. However, most research focuses on detailed requirements analysis, not suitable for agile app development. Future studies could focus not only on automated extraction of legal/privacy requirements but also on generating (privacy-related) user stories to be used in agile development. Research directions on automated approaches to monitor compliance and nudge developers towards compliant approaches are also worth addressing [18].

**Educational Takeaway** Similar to other work [7, 48, 85], our work shows the need for a more focused educational approach toward privacy in the SDLC. While currently, many courses emphasize security, it is important to tailor specific courses that include advanced privacy topics such as: regulations; the importance of PIA and other artifacts; challenges in privacy policy creation; and approaches such as PbD, differential privacy, and federated learning. This distinction between privacy from security is crucial since privacy encompasses a broad spectrum of concerns, including data handling, user consent, and transparency. Software teams should be equipped with educational modules and tools that foster and support life-long learning of dynamic privacy concepts. Nudging developers towards more privacy-preserving solutions through online support and tools is important. Balebako et al. [6] suggest that with the right guidance, developers can be encouraged to prioritize privacy in their design and development processes.

## 9 Conclusion

In this paper, we examined privacy perceptions, practices, and behaviors of SDLC team members during software development. Our findings suggest a need for standardized privacy practices, educational awareness and implementation of PbD, and a privacy expert to promote privacy awareness and compliance. We identified gaps in privacy practices among software teams. Finally, we provide research and educational directions to reduce the challenges in implementing these practices.

In the future, we will extend our research to conduct a comparative analysis within the US states. We will also evaluate whether developers over-claim their expertise in a new study. We will look into how privacy is taught at educational institutes, both in computer science and at Law schools.

## Acknowledgments

This research was funded by NSF Award # 2238047.

## References

[1] Mahmoud Alfadel, Diego Elias Costa, and Emad Shihab. Empirical analysis of security vulnerabilities in python packages. *Empirical Software Engineering*, 28(3):59, 2023.

[2] Atheer Aljeraisy, Masoud Barati, Omer Rana, and Charith Perera. Privacy laws and privacy by design schemes for the internet of things: A developer's perspective. *ACM Computing Surveys (CSUR)*, 54(5):1–38, 2021.

[3] Noura Alomar and Serge Egelman. Developers say the darnedest things: Privacy compliance processes followed by developers of child-directed apps. *Proc. on Privacy Enhancing Technologies*, 4(2022):24, 2022.

[4] Orlando Amaral, Sallam Abualhaija, Mehrdad Sabetzadeh, and Lionel Briand. A model-based conceptualization of requirements for compliance checking of data processing against gdpr. In *2021 IEEE 29th Int. Requirements Engineering Conf. Workshops (REW)*, pages 16–20, 2021.

[5] Renana Arizon-Peretz, Irit Hadar, Gil Luria, and Sofia Sherman. Understanding developers' privacy and security mindsets via climate theory. *Empirical Software Engineering*, 26:1–43, 2021.

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

[7] Rebecca Balebako, Abigail Marsh, Jialiu Lin, Jason I Hong, and Lorrie Faith Cranor. The privacy and security behaviors of smartphone app developers.(2014). *DOI: http://dx. doi. org/10.1184*, 1, 2014.

[8] Kenneth A Bamberger and Deirdre K Mulligan. Privacy on the books and on the ground. *Stanford Law Review*, pages 247–315, 2011.

[9] Kathrin Bednar, Sarah Spiekermann, and Marc Langheinrich. Engineering privacy by design: Are engineers ready to live up to the challenge? *The Information Society*, 35(3):122–142, 2019.

[10] Colin J Bennett and Charles D Raab. *The governance of privacy: Policy instruments in global perspective*. Routledge, 2017.

[11] J. Bhatia and T.D. et al. Breaux. Privacy risk in cybersecurity data sharing. In *Proc. of the ACM on Workshop on ISCS*, pages 57–64, 2016.

[12] Travis D. Breaux, Hanan Hibshi, and Ashwini Rao. Eddy, a formal language for specifying and analyzing data flow specifications for conflicting privacy requirements. *Requirements Engineering*, 19(3):281–307, 2014.

[13] Travis D. Breaux, Daniel Smullen, and Hanan Hibshi. Detecting repurposing and over-collection in multi-party privacy requirements specifications. In *Requirements Engineering Conference (RE), 2015 IEEE 23rd International*, pages 166–175. IEEE, 2015.

[14] Norman Breslow. A generalized kruskal-wallis test for comparing k samples subject to unequal patterns of censorship. *Biometrika*, 57(3):579–594, 1970.

[15] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[16] Fei Bu, Nengmin Wang, Bin Jiang, and Huigang Liang. "privacy by design" implementation: Information system engineers' perspective. *International Journal of Information Management*, 53:102124, 2020.

[17] Ann Cavoukian. Privacy by design - the 7 foundational principles implementation and mapping of fair information practices. *www.privacybydesign.ca*, 2009.

[18] Checks. Simplify compliance with google. https://checks.google.com/, 2024 (accessed Jun 6, 2024).

[19] Virginie Cobigo, Konrad Czechowski, Hajer Chalghoumi, Amelie Gauthier-Beaupre, Hala Assal, Jeffery Jutai, Karen Kobayashi, Amanda Grenier, and Fatoumata Bah. Protecting the privacy of technology users who have cognitive disabilities: Identifying areas for improvement and targets for change. *Journal of Rehabilitation and Assistive Technologies Engineering*, 7:2055668320950195, 2020.

[20] Michael Colesky, Jaap-Henk Hoepman, and Christiaan Hillen. A critical analysis of privacy design strategies. In *2016 IEEE security and privacy workshops (SPW)*, pages 33–40. IEEE, 2016.

[21] Asmita Dalela, Saverio Giallorenzo, Oksana Kulyk, Jacopo Mauro, and Elda Paja. A mixed-method study on security and privacy practices in danish companies. *arXiv preprint arXiv:2104.04030*, 2021.

[22] George Danezis, Josep Domingo-Ferrer, Marit Hansen, Jaap-Henk Hoepman, Daniel Le Metayer, Rodica Tirtea, and Stefan Schiffner. Privacy and data protection by design-from policy to engineering. *arXiv preprint arXiv:1501.03726*, 2015.

[23] Anastasia Danilova, Alena Naiakshina, Stefan Horstmann, and Matthew Smith. Do you really code? designing and evaluating screening questions for online surveys with programmers. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, pages 537–548. IEEE, 2021.

[24] Zack Delile, Sean Radel, Joe Godinez, Garrett Engstrom, Theo Brucker, Kenzie Young, and Sepideh Ghanavati. Evaluating privacy questions from stack overflow: Can chatgpt compete? In *2023 IEEE 31st International Requirements Engineering Conference Workshops (REW)*, pages 239–244. IEEE, 2023.

[25] Cynthia Dwork. Differential privacy. In *International colloquium on automata, languages, and programming*, pages 1–12. Springer, 2006.

[26] Anirudh Ekambaranathan, Jun Zhao, and Max Van Kleek. "money makes the world go around": Identifying barriers to better privacy in children's apps from developers' perspectives. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2021.

[27] European Union. The eu general data protection regulation (gdpr). http://www.eugdpr.org/, 2024 (accessed February 10, 2024).

[28] Saad Ezzini, Sallam Abualhaija, Chetan Arora, Mehrdad Sabetzadeh, and Lionel C. Briand. Using domain-specific corpora for improved handling of ambiguity in requirements. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, pages 1485–1497, 2021.

[29] Federal Trade Commission. Children's online privacy protection rule; final rule. http://tinyurl.com/5fh55th2, 2024 (accessed Feb 12, 2024).

[30] Zhangyin Feng, Daya Guo, Duyu Tang, Nan Duan, Xiaocheng Feng, Ming Gong, Linjun Shou, Bing Qin, Ting Liu, Daxin Jiang, and Ming Zhou. CodeBERT: A pre-trained model for programming and natural languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1536–1547. ACL, 2020.

[31] Sepideh Ghanavati, Daniel Amyot, and Liam Peyton. Towards a Framework for Tracking Legal Compliance in Healthcare. In John Krogstie, Andreas Opdahl, and Guttorm Sindre, editors, *Advanced Information Systems Engineering*, pages 218–232. Springer, 2007.

[32] Sepideh Ghanavati, Daniel Amyot, and Liam Peyton. Compliance analysis based on a goal-oriented requirement language evaluation methodology. In *2009 17th IEEE International Requirements Engineering Conference*, pages 133–142. IEEE, 2009.

[33] Sepideh Ghanavati, Daniel Amyot, and André Rifaut. Legal Goal-oriented Requirement Language (Legal GRL) for Modeling Regulations. In *Proc. of the 6th International Workshop on Modeling in Software Engineering*, pages 1–6, New York, NY, USA, 2014. ACM.

[34] Alessandra Gorla, Ilaria Tavecchia, Florian Gross, and Andreas Zeller. Checking app behavior against app descriptions. In *Proc. of the 36th Int. Conference on Software Engineering*, pages 1025–1035, 2014.

[35] Government of California. California consumer privacy act (ccpa). https://oag.ca.gov/privacy/ccpa, 2022 (accessed July 20, 2022).

[36] Matthew Green and Matthew Smith. Developers are not the enemy!: The need for usable security apis. *IEEE Security & Privacy*, 14:40–46, 2016.

[37] Priscilla E Greenwood and Michael S Nikulin. *A guide to chi-squared testing*, volume 280. John Wiley & Sons, 1996.

[38] Sara Gustavsson. An assessment of privacy by design as a stipulation in gdpr. 2020.

[39] Irit Hadar, Tomer Hasson, Oshrat Ayalon, Eran Toch, Michael Birnhack, Sofia Sherman, and Arod Balissa. Privacy by designers: Software developers' privacy mindset. *Journal of Empirical Software Engineering*, 23(1):259–289, February 2018.

[40] Irit Hadar, Tomer Hasson, Oshrat Ayalon, Eran Toch, Michael Birnhack, Sofia Sherman, and Arod Balissa. Privacy by designers: software developers' privacy mindset. *Empirical Software Engineering*, 23(1):259–289, 2018.

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

[42] Yongzhong He, Xuejun Yang, Binghui Hu, and Wei Wang. Dynamic privacy leakage analysis of android third-party libraries. *Journal of Information Security and Applications*, 46:259–270, 2019.

[43] US Department Health and Human Services. The Health Insurance Portability and Accountability Act (HIPAA). https://www.hhs.gov/hipaa/index.html, 2024 (accessed Feb 10, 2024).

[44] J. Hoepman. Privacy design strategies (extended abstract). 2014.

[45] J-H Hoepman. Making privacy by design concrete. 2018.

[46] Chris Jay Hoofnagle, Jennifer King, Su Li, and Joseph Turow. How different are young adults from older adults when it comes to information privacy attitudes and policies? *Available at SSRN 1589864*, 2010.

[47] Jaap-Henk Hopeman and Marc van Lieshout. Privacy: a fundamental right.

[48] Stefan Albert Horstmann, Samuel Domiks, Marco Gutfleisch, Mindy Tran, Yasemin Acar, Veelasha Moonsamy, and Alena Naiakshina. "those things are written by lawyers, and programmers are reading that." mapping the communication gap between software developers and privacy experts. *Proc. Priv. Enhancing Technol.*, 2024:151–170, 2024.

[49] Tianjian Huang, Vaishnavi Kaulagi, Mitra Bokaei Hosseini, and Travis Breaux. Mobile application privacy risk assessments from user-authored scenarios. In *Proceedings of the 31st IEEE International Requirements Engineering Conference*, pages 1–12. IEEE, 2023.

[50] International Association of Privacy Professionals. Taxonomy of privacy. https://iapp.org/resources/article/a-taxonomy-of-privacy/a, 2024 (accessed June 1, 2024).

[51] Leonardo Horn Iwaya, Muhammad Ali Babar, and Awais Rashid. Privacy engineering in the wild: Understanding the practitioners' mindset, organisational aspects, and current practices. *IEEE Transactions on Software Engineering*, 2023.

[52] Akshath Jain, David Rodriguez, Jose M del Alamo, and Norman Sadeh. Atlas: Automatically detecting discrepancies between privacy policies and privacy labels. *arXiv preprint arXiv:2306.09247*, 2023.

[53] Vijayanta Jain, Sepideh Ghanavati, Sai Teja Peddinti, and Collin McMillan. Towards fine-grained localization of privacy behaviors. In *IEEE 8th European Symposium on Security and Privacy*, pages 258–277, 2023.

[54] Vijayanta Jain, Sanonda Datta Gupta, Sepideh Ghanavati, and Sai Teja Peddinti. Prigen: Towards automated translation of android applications' code to privacy captions. In *Int. Conference on Research Challenges in Information Science*, pages 142–151. Springer, 2021.

[55] Vijayanta Jain, Sanonda Datta Gupta, Sepideh Ghanavati, Sai Teja Peddinti, and Collin McMillan. Pact: Detecting and classifying privacy behavior of android applications. In *Proc. of the 15th ACM Conf. on Security and Privacy in Wireless and Mobile Networks*, WiSec '22, page 104–118. ACM, 2022.

[56] Harjot Kaur, Sabrina Amft, Daniel Votipka, Yasemin Acar, and Sascha Fahl. Where to recruit for security development studies: Comparing six software developer samples. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 4041–4058, 2022.

[57] Rishabh Khandelwal, Asmit Nayak, Paul Chung, and Kassem Fawaz. Unpacking privacy labels: A measurement and developer perspective on google's data safety section. *arXiv preprint arXiv:2306.08111*, 2023.

[58] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

[59] Tianshi Li, Elizabeth Louie, Laura Dabbish, and Jason I Hong. How developers talk about personal data and what it means for user privacy: A case study of a developer forum on reddit. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):1–28, 2021.

[60] Xueqing Liu, Yue Leng, Wei Yang, Wenyu Wang, Chengxiang Zhai, and Tao Xie. A large-scale empirical study on android runtime-permission rationale messages. In *2018 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 137–146. IEEE, 2018.

[61] Laura MacLeod, Andreas Bergen, and Margaret-Anne Storey. Documenting and sharing software knowledge using screencasts. *Empirical Software Engineering*, 22:1478–1507, 2017.

[62] European Union Agency For Network and Information Security. Pets controls matrix a systematic approach for assessing online and mobile privacy tools. 2016.

[63] Serge Egelman Noura Alomar and and Jordan L. Fischer. Developers say the darnedest things: Privacy compliance processes followed by developers of child-directed apps. *Proceedings on Privacy Enhancing Technologies*, 2022(4), 2022.

[64] US Department of Education. The Family Educational Rights and Privacy Act (FERPA). https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html, 2024 (accessed Feb 10, 2024).

[65] Ehimare Okoyomon, Nikita Samarin, Primal Wijesekera, Amit Elazari Bar On, Narseo Vallina-Rodriguez, Irwin Reyes, Álvaro Feal, and Serge Egelman. On the

ridiculousness of notice and consent: Contradictions in app privacy policies. 2019.

[66] United Nations Conference on Trade and Development. Data protection and privacy legislation worldwide. https://tinyurl.com/puev83dt, 2021.

[67] R OpenAI. Gpt-4 technical report. *arXiv*, pages 2303–08774, 2023.

[68] Stefan Palan and Christian Schitter. Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, 2018.

[69] Rahul Pandita, Xusheng Xiao, Wei Yang, William Enck, and Tao Xie. {WHYPER}: Towards automating risk assessment of mobile applications. In *Presented as part of the 22nd {USENIX} Security Symposium ({USENIX} Security 13)*, pages 527–542, 2013.

[70] Jonathan Parsons, Michael Schrider, Oyebanjo Ogunlela, and Sepideh Ghanavati. Understanding developers privacy concerns through reddit thread analysis. *Joint Proc. of REFSQ-2023 Workshops, Doctoral Symposium, Posters & Tools Track and Journal Early Feedback co-located with the 28th Int. Conf. on Requirements Engineering: Foundation for Software Quality (REFSQ 2023), Barcelona, Catalunya*, 2023.

[71] Zhengyang Qu, Vaibhav Rastogi, Xinyi Zhang, Yan Chen, Tiantian Zhu, and Zhong Chen. Autocog: Measuring the description-to-permission fidelity in android applications. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 1354–1365, 2014.

[72] David Rodriguez, Akshath Jain, Jose M Del Alamo, and Norman Sadeh. Comparing privacy label disclosures of apps published in both the app store and google play stores. In *IEEE European Symp. on Security and Privacy Workshops*, pages 150–157, 2023.

[73] Nayan B. Ruparelia. Software development lifecycle models. *SIGSOFT Softw. Eng. Notes*, 35(3):8–13, 2010.

[74] Raphael Serafini, Marco Gutfleisch, Stefan Albert Horstmann, and Alena Naiakshina. On the recruitment of company developers for security studies: results from a qualitative interview study. In *19th Symposium on Usable Privacy and Security*, pages 321–340, 2023.

[75] J P Shaffer. Multiple hypothesis testing. *Annual Review of Psychology*, 46(1):561–584, 1995.

[76] Yun Shen and Siani Pearson. Privacy enhancing technologies: A review. *Hewlet Packard Development Company. Disponible en https://bit. ly/3cfpAKz*, 2011.

[77] Rocky Slavin, Xiaoyin Wang, Mitra Bokaei Hosseini, James Hester, Ram Krishnan, Jaspreet Bhatia, Travis D Breaux, and Jianwei Niu. Pvdetector: a detector of privacy-policy violations for android apps. In *2016 IEEE/ACM Int. Conf. on Mobile Software Engineering and Systems (MOBILESoft)*, pages 299–300, 2016.

[78] Rocky Slavin, Xiaoyin Wang, Mitra Bokaei Hosseini, James Hester, Ram Krishnan, Jaspreet Bhatia, Travis D Breaux, and Jianwei Niu. Toward a framework for detecting privacy policy violations in android application code. In *Proceedings of the 38th International Conference on Software Engineering*, pages 25–36, 2016.

[79] Daniel J Solove. A taxonomy of privacy. *University of Pennsylvania law review*, pages 477–564, 2006.

[80] Sarah Spiekermann and Lorrie Faith Cranor. Engineering privacy. *IEEE Transactions on Software Engineering*, 35(1):67–82, 2009.

[81] Sarah Spiekermann, Jana Korunovska, and Marc Langheinrich. Inside the organization: Why privacy and security engineering is a challenge for engineers. *Proceedings of the IEEE*, 107(3):600–615, 2018.

[82] Sarah Spiekermann-Hoff, Jana Korunovska, and Marc Langheinrich. Understanding engineers' drivers and impediments for ethical system development: The case of privacy and security engineering. 2018.

[83] Latanya Sweeney. k-anonymity: A model for protecting privacy. *Int. journal of uncertainty, fuzziness and knowledge-based systems*, 10(05):557–570, 2002.

[84] Mohammad Tahaei, Julia Bernd, and Awais Rashid. Privacy, permissions, and the health app ecosystem: A stack overflow exploration. In *Proc. of the 2022 European Symposium on Usable Security*, pages 117–130, 2022.

[85] Mohammad Tahaei, Alisa Frik, and Kami Vaniea. Privacy champions in software teams: Understanding their motivations, strategies, and challenges. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2021.

[86] Mohammad Tahaei, Adam Jenkins, Kami Vaniea, and Maria Wolters. "i don't know too much about it": On the security mindsets of computer science students. In Thomas Groß and Theo Tryfonas, editors, *Socio-Technical Aspects in Security and Trust*, pages 27–46, Cham, 2021. Springer International Publishing.

[87] Mohammad Tahaei, Tianshi Li, and Kami Vaniea. Understanding privacy-related advice on stack overflow. *Proceedings on Privacy Enhancing Technologies*, 2022(2):114–131, 2022.

[88] Mohammad Tahaei and Kami Vaniea. A survey on developer-centred security. In *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 129–138. IEEE, 2019.

[89] Mohammad Tahaei and Kami Vaniea. "developers are responsible": What ad networks tell developers about privacy. In *Extended Abstracts in CHI Conf. on Human Factors in Computing Systems*, pages 1–11, 2021.

[90] Mohammad Tahaei and Kami Vaniea. Recruiting participants with programming skills: A comparison of four crowdsourcing platforms and a cs student mailing list. In *CHI Conference on Human Factors in Computing Systems*, CHI '22. ACM, 2022.

[91] Mohammad Tahaei, Kami Vaniea, and Naomi Saphra. Understanding privacy-related questions on stack overflow. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–14, 2020.

[92] The Federal Trade Commission. Privacy and security enforcement. 2024 (accessed Feb 10, 2024).

[93] UC - Berkeley - School of Information. Privacy patterns - collaborative development of privacy software design patterns. https://github.com/privacypatterns, 2024 (accessed Feb. 10, 2024).

[94] UC Berkeley - School of Information. Privacy patterns org. https://privacypatterns.org/, 2024 (accessed February 10, 2024).

[95] Varonis. 84 must-know data breach statistics for 2023. https://www.varonis.com/blog/data-breach-statistics, (accessed Feb. 10, 2024).

[96] L. Yu and X. et al. Lou. Can we trust the privacy policies of android apps? In *46th Annual IEEE/IFIP Int. Conf. on (DSN)*, pages 538–549. IEEE, 2016.

[97] Xian Zhan, Lingling Fan, Sen Chen, Feng We, Tianming Liu, Xiapu Luo, and Yang Liu. Atvhunter: Reliable version detection of third-party libraries for vulnerability identification in android applications. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, pages 1695–1707, 2021.

[98] Sebastian Zimmeck, Rafael Goldstein, and David Baraka. Privacyflash pro: Automating privacy policy generation for mobile apps. 2021.

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

[100] Michael Zimmer. The gaze of the perfect search engine: Google as an infrastructure of dataveillance. In *Web search: Multidisciplinary perspectives*, pages 77–99. Springer, 2008.

## A Survey Questions

Survey questions can be found here: http://tinyurl.com/2p9n49e4

## B Participants' Demographic Information

Table 10 below shows the various demographics of our participants.

## C Details of Solove's Taxonomy

Solove's Taxonomy and the mapping of subcategories.

Table 11: Solove's Categories and Subcategories

| **Main Category** | **Solove's Subcategories** |
|---|---|
| Information Collection | Surveillance, Interrogation |
| Information Processing | Aggregation, Identification, Insecurity, Secondary Use, Exclusion |
| Information Dissemination | Breach of Confidentiality, Disclosure, Exposure, Increased Accessibility, Blackmail, Appropriation, Distortion |
| Invasion | Intrusion, Decisional Interference |

## D Confidence in Security & Privacy Measures

The hypotheses list for the correlation between confidence in security and privacy measures and various factors are:
– **H1a:** The size of the company correlates with confidence in privacy and security measures.
– **H1b:** The participants' role at the company correlates to confidence in privacy and security measures.
– **H1c:** The education level correlates to confidence in privacy and security measures.
– **H1d:** The presence of a CPO or similar position correlates to confidence in privacy and security measures.

The p-value results of the Chi-Square tests are as follows:

Table 12: P-Value for Hypothesis **H1a** to **H1d**

| | **H1a** | **H1b** | **H1c** | **H1d** |
|---|---|---|---|---|
| **P-Value** | 0.494 | 0.654 | 0.570 | 0.0007 |

## E Presence of a CPO or a Similar Role

The participant's knowledge about the presence of a CPO in their company is as follows:

Table 10: Demographic Information about the Participants

| | | | | | |
|---|---|---|---|---|---|
| **Gender** | Female (25.48%) | Male (73.41%) | Non-Binary (0.55%) | Other (0.55%) | PnS (0%) |
| **Age** | 18-25 (19.89%) | 26-35 (45.86%) | 36-45 (20.99%) | 46-55 (8.84%) | >55 (3.87%) |
| **Education** | High school (10.22%) | BSc. (61.05%) | MSc. (22.10%) | PhD (1.66%) | Other (3.87%) |
| **Degree** | CS/ECE/DS (34.8%) | IT (26.24%) | Business (11.05%) | Other (24.04%) | PnS (3.87%) |
| **Company Size** | 100+ emp. (50.00%) | 50-100 (13.54%) | 21-50 (12.43%) | 11-20 (7.46%) | 0-10 (16.57%) |

Table 13: Distribution of Knowledge about a CPO

| Yes | No | Unsure | Others |
|---|---|---|---|
| 42.6% | 38.4% | 17.9% | 1.1% |

The hypotheses list for the correlation between the presence of a CPO/a similar role and the PIA creation, familiarity with PETs, number of privacy breaches, and the company size are:
– **H2a:** The creation of a PIA correlates to the presence of a CPO or similar position at the company.
– **H2b:** Familiarity with PETs correlates to the presence of a CPO or similar position at the company.
– **H2c:** The higher number of privacy breaches correlates to the presence of a CPO or similar position at the company.
– **H2d:** The size of a company correlates to the presence of a CPO or similar position at the company.

The p-value results of the Chi-Square tests are as follows:

Table 14: P-Value for Hypothesis **H2a** to **H2d**

| | H2a | H2b | H2c | H2d |
|---|---|---|---|---|
| **P-Value** | 0.1005 | 0.008 | 0.359 | $< 0.00001$ |

The distribution of how participants address their compliance questions:

Table 15: Distribution of Sources for Compliance Questions

| Lawyer | CPO | Best Practices | Forums | Others |
|---|---|---|---|---|
| 24.2% | 25.9% | 23.1% | 18.5% | 8.3% |

## F The Creation of a PIA

The hypotheses list for the correlation between the creation of a PIA and the company size and confidence in privacy and security measures are:
– **H3a:** The size of the company correlates to the PIA creation.
– **H3b:** The participants' confidence in an organization's privacy and security measures correlates to the PIA creation.

The p-value results of the Chi-Square tests are as follows:

Table 16: P-Value for Hypothesis **H3a** to **H3b**

| | H2a | H2b |
|---|---|---|
| **P-Value** | $< 0.00001$ | $< 0.00001$ |

The distribution of responses to the creation of a PIA in their company is shown in Figure 5.

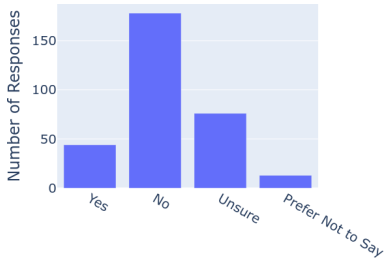


Figure 5: Distribution of Responses to the Creation of a PIA.

## G Privacy by Design Approaches

The distribution of participants who are familiar with PbD:

Table 17: Distribution of Familiarity with PbD Strategies

| Role | Yes | No | Unsure | PnS |
|---|---|---|---|---|
| **SD** | 91 (46%) | 54 (27.3%) | 49 (24.7%) | 4 (2%) |

## H Detailed Analysis of PETs' Familiarity

The list of the hypotheses for the correlation between the usage of PETs and the size of the company, participants' confidence, and the presence of the CPO is as follows:
– **H4a:** The size of the company correlates to the use of PETs.
– **H4b:** The participant's confidence in an organization's privacy and security measures correlates to the use of PETs.
– **H4c:** The participant's education level correlates to the use of PETs.

Table 18 shows the results of the hypotheses analysis.

Table 18: P-Value for Hypothesis **H4a** to **H4c**

| | H4a | H4b | H4c |
|---|---|---|---|
| **P-Value** | 0.254 | 0.704 | 0.529 |

Table 19: P-Value and H Value for Hypothesis **H5a** to **H5c**

| | H5a | H5b | H5c |
|---|---|---|---|
| **P-Value** | 0.04 | 0.17 | 0.08 |
| **H Value** | 4.03 | 7.80 | 9.83 |

Table 20: Categories of PETs

| Categories of PETs | Definition in Literature [62,76] | Example from Our Survey |
|---|---|---|
| **Encryption** | A system of communication where the only people who can read the messages are the people communicating. | We use encryption and a number of security features offered by the platform we implement. It is primarily the responsibility of the back-end programmers. |
| **Access Control/ Identity protection** | Deals with identifying individuals and controlling access to resources in a system. | We implement role-based access for the various features of our product as well as internally |
| **Anonymity and Pseudonymity** | Involves removing personally identifiable information (PII) to prevent individual users from being identified. Pseudonymity involves replacing identifiers with pseudonyms [83]. | Data anonymization, our managers would be the primary users for that subject |
| **Differential Privacy** | Involves adding noise to the data to protect individual user information while still providing useful insights. It is particularly useful in data analysis and machine learning applications. [25] | We use encryption and al little bit of **differential privacy** where it is applicable and it varies from project to project with who is tasked with implementing these features. |
| **Secure Communication/ VPN** | Involves encrypting all communications within the software using standard protocols like HTTPS and SSL/TLS. | All of our internal communication is done over an internal VPN, and all web access is done with https. |
| **Privacy-Enhance Anti Web Tracking** | Involves blocking attempts of different types of trackers to monitor users' online activity and personal data. | - |

# I Factors Influencing Usage of Forums

To further evaluate the impact of the size of the company, familiarity with PETs, and the presence of a CPO on the usage of developer forums, we employed the Kruskal-Wallis test which is a non-parametric test that is used to compare two or more independent samples for statistically significant differences between groups [58]. Below is the list of hypotheses for the frequency of the usage of the developers' forums:

– **H5a:** The size of the company correlates to the use of developer forums to ask privacy-related questions.

– **H5b:** The presence of a Chief Privacy Officer or similar position at a participant's organization correlates to the use of developer forums to ask privacy-related questions.

– **H5c:** Familiarity with PETs correlates to the use of developer forums to ask privacy-related questions.

As shown in Table 19 when comparing forum usage with the size of the company, a statistically significant difference was found between the groups ($H-Value = 4.03, p-value = 0.04$). However, no significant difference was noted when comparing forum usage with the presence of a Chief Privacy Officer (CPO) ($H-value = 9.83, p-value = 0.08$) or with the usage of Privacy Enhancing Technologies (PETs) ($H-value = 3.92, p-value = 0.56$). These findings suggest that only the size of the company is more likely to influence the frequency with which developers consult forums for privacy-related inquiries.

# J Details for the Location Analysis

Below is the list of hypotheses for location analysis.

– **H6a:** The participants' confidence in their organization's privacy and security measures correlates to their region of origin.

– **H6b:** The presence of a CPO or similar position at a participant's organization correlates to their region of origin.

– **H6c:** The participants' creation of a PIA correlates to their region of origin.

– **H6d:** The participants' organization being a victim of a breach of privacy correlates to their region of origin.

– **H6e:** The participants' familiarity with PbD strategies correlates to their region of origin.

– **H6f:** The participants' use of PETs correlates to their region of origin.

– **H6g:** The participants' familiarity with the CCPA correlates to their region of origin.

– **H6h:** The participants' familiarity with the GDPR correlates to their region of origin.

# K Qualitative Analysis Guidelines

Table 20 shows the different categories of PETs and Table 21 describes the privacy taxonomy, both of which were considered as guidelines for our qualitative analysis.

Table 21: Taxonomy of Privacy

| Taxonomy of Privacy | Solove's Definition [79] | Example from IAPP [50] | Example from Our Survey |
|---|---|---|---|
| **Surveillance** | Watching, listening to, or recording of an individual's activities | A website monitoring the cursor movements of a visitor while visiting the website. | Privacy is the ability to keep information or activities out of public knowledge |
| **Interrogation** | Questioning or probing for personal information | An interviewer asking an inappropriate question, such as marital status, during a employment interview. | As far as I'm the internet, not asking for private information from our customers such as addresses or any sensitive information. |
| **Aggregation** | Combining of various pieces of personal information | A credit bureau combining an individual's payment history from multiple creditors. | Keeping unnecessary information from being exchanged at the minimum amount possible. |
| **Insecurity** | Carelessness in protecting information from leaks or improper access | An e-commerce website allowing others to view an individual's purchase history by changing the URL (e.g. enterprivacy.com?id=123) | Having confidential and private information secured and stored away safely from malicious users. |
| **Identification** | Linking of information to a particular Individual. | A researcher linking medical files to the Governor of a state using only date of birth, zip code and gender. | I think it can be defined as a set of personal information of each individual that should not be accessible to other people |
| **Secondary Use** | Using personal information for a purpose other than the purpose or which is was collected | The U.S. Government uses census data collected for the purpose of apportioning Congressional districts to identify and intern those of Japanese descent in WWII. | Ensuring the minimum amount of data is available only to those that genuinely need it for business purposes, and that it's only available for the specified amount of time that the data is needed. |
| **Exclusion** | Failing to let an individual know about the information that others have about them and participate in its handling or use | A company using customer call history, without the customer's knowledge, to shift their order in a queue (i.e. "Your call will be answer in the order [NOT] received") | to have the authority of controlling information about yourself who can or can not see. to be from from any interference, and to be able to interact with anyone I want. |
| **Breach of Confidentiality** | Breaking a promise to keep a person's information confidential | A doctor revealing patient information to friends on a social media website. | Having confidential and private information secured and stored away safely from malicious users. |
| **Disclosure** | Revealing truthful personal information about a person that impacts the ways others judge their character or their security | A government agency revealing an individual's address to a stalker, resulting in the individual's murder. | Data must be kept safe, and users need that information to be seen only by those they authorize. |
| **Exposure** | Revealing an individual's nudity, grief, or bodily functions | A store forcing a customer to remove clothing revealing a colostomy bag. | Freedom of your own information. |
| **Increased Accessibility** | Amplifying the accessibility of personal information | A court making proceeding searchable on the Internet without redacting personal information. | A state where one can be sure no one else knows what they are doing |
| **Blackmail** | Threatening to disclose personal information | A dating service for adulters charging customers to delete their accounts. | - |
| **Appropriation** | Using an individual's identity to serve the aims and interests of another | A social media site using customer's images in advertising | Being able to be secure in your information so that none of it gets accessed or leaked by outside sources |
| **Distortion** | Disseminating false or misleading information about an individual | A creditor reporting a paid bill as unpaid to a credit bureau. | Privacy refers to an individual's right to control [..]on. This includes protecting sensitive data from [..], and providing individuals with the ability to access, correct, or delete their PI. |
| **Intrusion** | Disturbing an individual's tranquility or solitude | An augmented reality game directing players onto private residential property. | The right to be let alone,or freedom from interference or intrusion. |
| **Decisional Inference** | Intruding into an individual's decision regarding their private affairs | A payment processor declining transactions for contraceptives | The right to be let alone,or freedom from interference or intrusion. |