Journal of Internet Services and Applications, 2025, 16:1, doi: 10.5753/jisa.2025.5302

Data Privacy in Software Practice: Brazilian Developers' Perspectives

Aryely Matos [**Federal University of Ceará** | ***aryelymatos@alu.ufc.br***]
Mario Patrício [**Federal University of Ceará** | ***jose.mariopatricio@alu.ufc.br***]
Maria Isabel Nicolau [**Pontifical Catholic University of Rio de Janeiro** | ***mariaisabel@aise.inf.puc-rio.br***]
Edna Dias Canedo [**University of Brasília** | ***ednacanedo@unb.br***]
Juliana Alves Pereira [**Pontifical Catholic University of Rio de Janeiro** | ***jpereira@inf.puc-rio.br***]
Anderson Uchôa [**Federal University of Ceará** | ***andersonuchoa@ufc.br***]

Campus of Itapajé – Jardins de Anita, Federal University of Ceará, Rua Francisco José de Oliveira, s/n, Centro, Itapajé, CE, 62600-000, Brazil

Received: 10 December 2024 • **Accepted:** 12 May 2025 • **Published:** 12 June 2025

Abstract Data privacy is an essential principle of information security, aimed at protecting sensitive data from unauthorized access and information leaks. As software systems advance, the volume of personal information also grows exponentially. Therefore, incorporating privacy engineering practices during development is vital to ensure data integrity, confidentiality, and compliance with legal regulations, such as the General Data Protection Regulation (GDPR). However, there is a gap in understanding developers' awareness of data privacy, their perceptions of the implementation of privacy strategies, and the influence of organizational factors on this adoption. Thus, this paper aims to explore the level of awareness among Brazilian developers regarding data privacy and their perceptions of the implementation strategies adopted to ensure data privacy. Additionally, we seek to understand how organizational factors influence the adoption of data privacy practices. To this end, we surveyed 88 Brazilian developers with privacy-related work experience. We got 21 statements grouped into three topics to measure the Brazilian developers' awareness of data privacy in software. Our statistical analysis reveals substantial gaps between groups, e.g., developers have Direct v.s. Indirect data privacy-related work experience. We also reveal some data privacy strategies, e.g., Encryption, are both widely used and perceived as highly important, others, such as Turning off data collection, highlight strategies where ease of use does not necessarily lead to widespread adoption. Finally, we identified that the absence of dedicated privacy teams correlates with a lower perceived priority and less investment in tools. Even in organizations that recognize the importance of privacy. Our findings offer insights into how Brazilian developers perceive and implement data privacy practices, emphasizing the critical role organizational culture plays in decision-making regarding privacy. We hope that our findings will contribute to improving privacy practices within the software development community, particularly in contexts similar to Brazil.

Keywords: Data Privacy; Software Development; Data Privacy Strategies; Organizational Factors; Awareness

1 Introduction

The increasing complexity of software products and services has introduced significant challenges to software engineering, particularly in achieving regulatory compliance [Kempe and Massey, 2021]. Regulations targeting technologies such as artificial intelligence or governing processes like personal data processing are becoming increasingly prevalent and stringent, forcing organizations to adapt quickly. Furthermore, advancements in the use of generative AI have sparked a growing demand for new regulatory frameworks specifically tailored to generative AI applications. As a result, the proliferation of laws and regulations is expected to intensify, addressing diverse concerns ranging from user data privacy to ethical issues associated with the use of emerging technologies [Kshetri, 2024].

For software developers, these evolving regulatory landscapes necessitate aligning their practices with structured principles and demonstrating compliance to avoid penalties such as substantial fines or even the removal of non-compliant products from the market [Peixoto *et al.*, 2023]. Data privacy, in particular, has emerged as a critical concern for regulatory compliance, especially in light of global data protection laws such as the General Data Protection Regulation (GDPR) and Brazil's General Data Protection Law (LGPD). Software engineers must navigate these requirements to ensure that personal data is handled responsibly throughout the Software Development Life Cycle (SDLC) [Canedo *et al.*, 2023]. However, this task is often complicated by overlapping regulatory demands and varying interpretations of privacy and security requirements. At the core of regulatory compliance is software requirements engineering, which translates legal and regulatory demands into actionable, testable requirements. Despite its importance, existing literature reveals a lack of systematic understanding of the challenges and practices related to integrating data privacy into software development. Furthermore, the regulations are designed to safeguard individuals' rights by governing how personal data is collected, processed, and stored, imposing strict requirements on organizations to uphold privacy standards [Canedo *et al*., 2023; Peixoto *et al*., 2023].

The importance of addressing the challenges associated with ensuring user data privacy extends far beyond avoiding penalties. Compliance with data privacy regulations serves as a foundation for building trust with end users and stakeholders, particularly as concerns over data violations and the misuse of personal information continue to grow. Developers must navigate multiple, often overlapping regulatory frameworks originating from different jurisdictions, each with its unique scope and requirements [Ferrão *et al*., 2024]. This complexity highlights the critical need for structured approaches to managing compliance and embedding privacy considerations into software engineering practices. By doing so, organizations can not only meet legal requirements but also foster trust and demonstrate their commitment to ethical data management in an increasingly privacy-conscious world [Sangaroonsilp *et al*., 2023].

By prioritizing privacy and regulatory compliance, organizations not only mitigate risks but also position themselves competitively in the market by demonstrating their commitment to ethical and responsible data management practices. For developers, mastering these challenges is essential for delivering software products and services that are not only legally compliant but also aligned with societal expectations and ethical standards. This makes regulatory compliance a cornerstone of sustainable and responsible software engineering in today's digital economy [Landis and Kroll, 2024].

This study aims to bridge this gap by conducting a survey with 88 Brazilian developers with privacy-related work experience. Based on the Knowledge-Attitude-Behaviour (KAB) model [Schrader and Lawless, 2004], we explore key aspects of developers' awareness, including their understanding of data privacy principles, laws, and best practices; their attitudes toward data privacy; and their actual behaviors and actions related to privacy. Additionally, we analyze their perceptions of 13 implementation strategies and the influence of organizational factors on data privacy practices. By focusing on the Brazilian context, this research provides valuable insights into the challenges developers face in aligning their practices with data privacy regulations.

Our contributions include: (i) insights into the most frequent terms Brazilian developers associate with data privacy; (ii) an analysis of empirical evidence highlighting both strengths and gaps in their awareness of data privacy; (iii) findings on Brazilian developers' perceptions of various data privacy strategies, focusing on their frequency of use, perceived importance, and ease of implementation; and (iv) a discussion on how organizational structures, priorities, and resources impact the adoption and enforcement of data privacy practices within development teams.

2 Background

This section overviews the theoretical foundation for understanding the key concepts of the study as follows.

Privacy vs Software Security. The term privacy is often confused with security, which is understandable since privacy is a critical component within the broader domain of Information Security. Tahaei *et al*. [2021] conducted interviews with privacy experts, asking them to define data privacy. Most practitioners responded that privacy is the protection of personal data against unauthorized access. On the other hand, security, as described in the study by Lester and Jamerson [2009], is defined as the ability of software to resist, tolerate, and recover from events that intentionally threaten its reliability while preserving the availability, integrity, and confidentiality of information. Although the concepts of data privacy and software security are distinct, they are closely related, as privacy is a subset of security. In our study, we explore the immediate associations and key concepts that Brazilian developers link to data privacy.

Privacy within the Software Development Process. Privacy within the software development process is often associated with *Privacy by Design* (PbD) [Hadar *et al*., 2018], a framework that guides developers in applying solutions and strategies aimed at ensuring privacy protection. In other words, PbD ensures that data privacy is integrated within the software development process, from the initial conception to the system's maintenance. Additionally, the growing need for data protection became even more evident with the introduction of regulations such as the General Data Protection Regulation (GDPR) in the European Union [União Europeia, 2016] and, later, the General Data Protection Law (LGPD) in Brazil [Brasil, 2018]. In this context, PbD has become a key solution for embedding privacy directly into system design, making it an inherent feature of systems to prevent unauthorized exposure or violations of personal data. Our study explores Brazilian developers' awareness of data privacy, focusing on: (i) their knowledge of privacy principles, laws, and best practices; (ii) their attitudes toward privacy, including beliefs, values, and emotional responses; and (iii) their actions concerning data privacy.

Practices of Privacy Engineering. Privacy engineering is a multidisciplinary field that integrates privacy principles into the design, development, and maintenance of systems [Stallings, 2019]. In other words, privacy engineering includes both technical capabilities and management processes. Among these principles, we can highlight the use of implementation strategies to ensure data privacy within the software development process. An example of a data privacy strategy is *Encryption*, which protects data by converting it into an unreadable format, ensuring confidentiality and integrity. In fact, as stated by a previous study [Iwaya *et al*., 2023], the use of different data privacy strategies might be necessary depending on the context and specific requirements of the system being developed, as each strategy offers unique advantages in addressing various privacy concerns. Table 1 overview 13 different data privacy strategies that are commonly used by developers.

Each of these strategies plays a distinct role in ensuring the protection of personal data, depending on the needs of the system and the regulatory requirements it must comply with [Iwaya *et al*., 2023; Stallings, 2019]. In our study, we explore the perceptions of Brazilian developers regarding the frequency, importance, and ease of use of each of the aforementioned data privacy strategies. By examining these perceptions, we aim to gain insights into the specific preferences, challenges, and priorities developers face when implement-

Table 1. Data Privacy Strategies and Their Descriptions

Strategy	Description
1. Encryption	A technique used to protect information by making it inaccessible to unauthorized individuals. It converts readable data (plaintext) into an unreadable format (ciphertext).
2. Minimization of personal data collection	Refers to the practice of limiting the amount of personal data collected by the organization.
3. Decentralization	Involves the distribution of data collection, storage, and processing across different locations or systems.
4. Data sovereignty	Based on respecting data protection laws and regulations in different jurisdictions.
5. Temporal data	Refers to using and storing data only for the time necessary.
6. User control	Involves giving users control over their own data, allowing them to access, modify, and delete their personal information as needed.
7. Turning off data collection	Consists of providing options for users to opt out of data collection, respecting their privacy preferences.
8. Anonymization	Involves removing or altering personally identifiable information so that the data cannot be associated with specific individuals.
9. Data classification tools	These are systems or software used to classify and catalog organizational data, facilitating the management and identification of sensitive or personal information.
10. Design and code reviews	Practices that ensure the development process complies with techniques that guarantee security throughout the development process.
11. Risk management	Involves assessing and planning for risks associated with data processing.
12. Data flow modeling	A technique used to visually represent the movement of data within an information system.
13. Proxy	A proxy is used for all requests to third-party services to hide users' identities and prevent third parties from obtaining information about them.

ing data privacy strategies, ultimately helping to inform best practices and decision-making in privacy engineering.

3 Study Settings

To define the goal, and research questions (RQs), we rely on the Goal-Question-Metric (GQM) template [Caldiera and Rombach, 1994]. Our goal is: **analyze** the awareness and implementation strategies related to data privacy among software developers; **for the purpose of** understanding their awareness and practices related to data privacy in software development; **with respect to** key aspects, such as awareness, perception about data privacy strategies, and the influence of organizational factors; **from the viewpoint of** researchers; **in the context of** Brazilian software developers. We detail each RQ as follows:

RQ$_1$: ***To what extent are developers aware of data privacy in software development?*** – RQ$_1$ aims to explore the multifaceted nature of developers' awareness of data privacy. To capture this awareness, we use 21 carefully crafted statements divided into three key topics: (i) *Knowledge* that assesses the developers' understanding and knowledge of data privacy principles, laws, and best practices; (ii) *Attitudes and Feelings* which delves into the developers' attitudes towards data privacy, including their beliefs, values, and emotional responses; and (iii) *Behaviors and Actions* which focuses on the actual behaviors and actions developers take concerning data privacy. By answering RQ$_1$, we can identify both strengths and gaps in developers' awareness of data privacy. This understanding can contribute to building more effective training and educational programs about data privacy within the software industry.

RQ$_2$: ***How do developers' perceptions of frequency, importance, and ease of use differ for various data privacy strategies?*** – RQ$_2$ aims to understand the developers' perceptions of different data privacy strategies in terms of how often they are used, how important they are considered, and how easy they are to use. Thus, in this RQ we will compare these perceptions across a range of strategies (e.g., encryption, anonymization, access control) to identify which strategies are most and least favored by developers. By answering RQ$_2$, we will help to determine whether certain strategies are underutilized due to perceived challenges or overemphasized due to their perceived importance.

RQ$_3$: ***How do organizational factors influence adherence to data privacy practices within development teams?*** – RQ$_3$ aims to explore four organizational factors: (i) the presence or not of dedicated privacy teams; (ii) the work dynamic for handling privacy and data protection within organizations; (iii) the use of not of specific tools for managing private data; and (iv) the participants' perceptions of their organization's priority regarding privacy and data protection. By answering RQ$_3$, we can shed light on how organizational structures, priorities, and resources influence the implementation and adherence to data privacy practices within development teams.

3.1 Survey Design

To answer our RQs, we choose to use the survey method. We consider this the most suitable research method to investigate our RQs through a relatively large sample of software developers. For this purpose, we use the Linåker *et al.* [2015] guidelines for setting up and conducting our survey. We describe each step in detail as follows.

Step 1: Defining objectives for information collection. In this step, we conducted a brainstorming session to define the goals and scope of the survey. Our primary goal is to assess developers' awareness of data privacy in software development starting from its conception. Additionally, we aim to examine the practices related to implementing data privacy strategies and how organizational factors influence adherence to these practices within development teams. These goals guided the design and focus of our survey questions, ensuring the collection of relevant and comprehensive data.

Step 2: Identifying the target population and sampling. Our target population consists of individuals who are over 18 years old and work in the fields of software development or information security in Brazil. To ensure the accuracy of our target audience, we included a control question to confirm their familiarity with data privacy in the software development process. We employed snowballing sampling procedures [Kitchenham and Pfleeger, 2002], in which initial participants help to identify and recruit additional participants, leading to the iterative expansion of our participant network.

Step 3: Designing survey instrument. All authors collaboratively designed our survey, which was administered via Google Forms. The survey consists of 35 questions with different types, i.e., open-ended, closed-ended, multiple-choice, and Likert scale. The survey was structured into five sections as illustrated in Table 2. We intentionally organized and grouped the questions in sections to avoid biasing participants' responses. Each section was presented to the participant only after answering the previous one.

Table 2. Survey Questions

ID	Question	Type
General Participant Characterization		
Q1	What is your age?	C
Q2	Which state do you currently reside in?	C
Q3	What is your gender?	C
Q4	What is your highest completed level of formal education?	C
Q5	What is your work model?	C
Q6	What role best describes your current activities in software development projects?	C
Q7	Please indicate the seniority level of your position.	C
Q8	How many years of experience do you have in roles related to software development or IT?	C
Q9	What types of systems do you currently work on?	M
Q10	What domains do the systems you currently work on belong to?	M
Q11	What sector is the organization you currently work for in?	C
Q12	What is the size of the company you currently work for?	C
Characterization of Experience with Data Privacy		
Q13	Do you have (or have you had) any work experience related to data privacy in the software development process?	C
Q14	Please briefly describe your experience with data privacy in software. If none, state that you have no experience.	O
Q15	What are the main sources or methods you use to learn about data privacy issues in software?	M
Q16	What types of personal data do you handle in your work?	M
Awareness of Data Privacy		
Q17	[T1: Knowledge] Provide at least five words in the order they come to mind when you think of privacy.	O
Q18	[T1: Knowledge] Based on your experience, rate and classify each of the following statements on a five-point scale.	L
Q19	[T1: Knowledge] For statements you strongly disagreed with, please describe the reason for your rating. (Optional)	O
Q20	[T2: Attitudes and Feelings] Based on your experience, rate and classify each of the following statements on a five-point scale.	L
Q21	[T2: Attitudes and Feelings] For statements you strongly disagreed with, please describe the reason for your rating. (Optional)	O
Q22	[T3: Behaviors and Actions] Based on your experience, rate and classify each of the following statements on a five-point scale.	L
Q23	[T3: Behaviors and Actions] For statements you strongly disagreed with, please describe the reason for your rating. (Optional)	O
Data Privacy Implementation Strategies		
Q24	How frequently do you use privacy techniques and strategies to ensure the protection of personal data in the following development phases?	L
Q25	How frequently do you use or have you used the following privacy strategies?	L
Q26	Do you use any other implementation strategies to ensure privacy that were not mentioned in the previous question? If so, which ones? (Optional)	O
Q27	In your opinion, what is the level of importance of the following privacy implementation strategies?	L
Q28	For the strategies you considered important, please describe the reason for your rating. (Optional)	O
Q29	Regarding these strategies, how would you characterize the ease of using them?	L
Organizational Factors		
Q30	In your organization, is there a dedicated team solely for handling privacy?	C
Q31	In your organization, what is the work dynamic for handling privacy and data protection? Please describe briefly.	O
Q32	In your organization, are one or more tools used for handling private data? If so, which ones?	M
Q33	What is your perception of your organization's priority regarding privacy and data protection?	L
Q34	In your view, what will be the biggest challenges for organizations in the coming years regarding practices and regulations to better protect users' privacy rights?	O
Q35	Please enter the current year to confirm that you are not a bot.	O

C: Closed; M: Multiple; O: Open; L: Likert

At the beginning of the survey, we provided study information, including the objective, methodology, data handling procedures, and researchers' contact details. Additionally, we presented an informed consent statement that outlined the conditions and stipulations governing participation.[1] Participants were informed that their participation was entirely voluntary, and that they had the freedom to decline participation or withdraw their consent at any time without facing any penalties. The survey was conducted anonymously, with no requests for respondents' contact information.

The first section comprises general questions designed to gather information about the developers' backgrounds and skills, defined by a set of 12 questions. The second section consists of four questions aimed at collecting data on the developer's experience working with data privacy in software. The third section includes seven questions aimed at assessing the developers' awareness of data privacy throughout the software development process.

To build this section, we rely on the Knowledge-Attitude-Behaviour (KAB) model from social psychology, which provides a framework for interpreting developers' personal aspects. The KAB model has been widely adopted to study information security awareness (e.g., [Thomson and von Solms, 1998; Kruger and Kearney, 2006; Parsons *et al.*, 2014]). Briefly, knowledge refers to all information a person possesses or accumulates in a specific domain [Schrader and Lawless, 2004]. Attitude comprises three components [Schrader and Lawless, 2004]: a cognitive component, such as a belief or idea associated with a psychological object; an affective component of the individual's evaluation and emotion associated with a psychological object; and a conative component, which represents predispositions or actions related to the object. In this study, the psychological object is the participants' conceptualization of data privacy within the software development process. Lastly, behavior refers to observable actions [Schrader and Lawless, 2004].

With the KAB model in mind, and based on the findings provided by prior studies [Iwaya *et al.*, 2023; Tahaei *et al.*, 2021; Hadar *et al.*, 2018], we create 21 statements categorized into three topics (T) to capture distinct dimensions of awareness: T1 - *Knowledge* (with seven statements), T2 - *Attitudes and Feelings* (with eight statements), and T3 - *Behaviors and Actions* (with six statements). Participants were asked to score the statements based on their work experience and opinions, using a 5-point Likert scale [Likert, 1932], ranging from *Strongly Disagree*, to *Strongly Agree*. Apart from scoring the statements, the participants were encouraged to provide a rationale for any statement that they strongly disagreed with on each topic.

The fourth section includes six questions designed to characterize the use of 13 data privacy strategies (e.g., encryption, minimizing collection of personal data, and proxy), focusing on the frequency of use, level of importance, and ease of use. Finally, the fifth section comprises five questions aimed at identifying organizational factors that may influence adherence to data privacy practices. We also made an additional question to verify if the participant is a bot or not.

Step 4: Validating survey instrument. Before conducting the survey, we performed a pilot test [Litwin and Fink, 1995] with four PhD students in Computer Science not involved in our study. Their feedback helped us refine the survey, offering suggestions on question phrasing and adjustments such as adding options and reordering alternatives for closed questions. We followed their advice and improved the survey questions. Note that the collected responses from the pilot survey are excluded from the presented results in this paper. Regarding the time to complete the survey, the pilot respondents, took an average of 15 minutes. We informed respondents of this average time when the survey questions were made publicly available.

Step 5: Administering the survey. We hosted and ad-

[1] The conditions adhered to ethical privacy standards as per the Brazilian General Data Protection Law (Law No. 13,709/2018).

ministered our survey at the Google Forms platform. We promoted the survey through social media posts and direct messages to reach our target audience. We shared posts on LinkedIn and sent direct messages to profiles on this platform, as well as via WhatsApp and email. The survey was available for 132 days, from June 25 to November 5, 2024.

3.2 Data Analysis Procedures

In total, we obtained, 122 responses, and we excluded 34 respondents who had reported no direct or indirect data privacy-related work experience. Finally, we considered 88 responses for analysis. In our data sample, about 63.3% of the participants have direct data privacy-related work experience, while 38.6% have indirect ones. Our analysis approach varied based on the question types. For closed and multiple-choice questions, we report the percentage of each option selected, supplemented by data visualizations. For the Likert scale questions, we report correlation analysis. Finally, we employed Grounded Theory (GT) procedures for the open-ended questions by performing open and axial coding [Corbin and Strauss, 2008]. GT refers to a method of inductively generating theory from data. Studies often include unstructured text, for instance, interview transcripts and field notes [Glaser and Strauss, 2017]. We present the analysis performed to answer our RQs as follows.

3.3 Analysis to Answer RQ1

To address RQ_1, we conducted both quantitative and qualitative analyses. We describe each analysis as follows.

The most frequent words that developers associate with data privacy. To capture the words developers associate with data privacy, participants were asked to list at least five words in the order they came to mind when thinking about privacy (Q17). In this context, each word was recorded alongside its position in the sequence, reflecting the order of mention (from the 1st to the 5th word). The analysis focused on identifying the most commonly mentioned words overall and examining how the order of mention influences their perceived importance. In this context, the Words mentioned in the first position often reflect immediate, core associations, while those listed later may indicate complementary or more reflective thoughts. The analysis provides insights into developers' mental models of data privacy, highlighting the key concepts and priorities they associate with this topic. The frequency distribution of the words is shown in Figure 5.

Correlations analysis on the awareness of data privacy between different groups. To gain deeper insights into participants' awareness of data privacy, we divided all survey respondents into different demographic groups. After, we compared their Likert scale responses reported in each statement of Q18, Q20, and Q23. We defined the following groups: (a) Respondents with direct data privacy work experience (54 respondents); (b) Respondents with indirect data privacy work experience, i.e., their groups have done data privacy-relevant work, but the respondents have not been directly involved, (34 respondents); (c) Respondents who are working in relatively larger organizations (56 responses); and (d) Respondents who are working in relatively smaller organizations (25 responses); (e) Respondents who have a privacy-related role (14 responses); and (f) Respondents who do not have a privacy-related role (74 responses).

To categorize participants based on their direct versus indirect experience with data privacy, we utilize responses to Q13, where developers indicate one of the following options: *Direct experience*: "I have worked (or currently work) with data privacy in the software development process (e.g., design, development, and testing)."; *Indirect experience*: "My development group/team works (or has worked) with privacy, but I am not directly involved in any tasks."; and *No experience*: "I have never had direct or indirect experience with data privacy in software.". Regarding the organization size, we refer to responses from Q12: "What is the size of the company that you currently work for?" Based on the responses, organizations are categorized as follows: *Small organizations*: Up to 9 employees, 10–49 employees, and 50–99 employees; and *Big organizations*: 100–499 employees, 500–999 employees, and more than 1,000 employees. Finally, we consider privacy-related roles (Q6), when the participants have selected one of these roles in the software development projects: Cryptographer, Privacy engineer, Penetration tester, and Security engineer.

The statements and results of the comparison are summarized in Table 3. To conduct this analysis, we used the Wilcoxon Rank Sum Test [Whitley and Ball, 2002] and the Cliff's Delta (d) measure [Grissom and Kim, 2005] to determine the level of agreement with each statement between the previously defined groups. The Cliff's Delta (d) measure [Grissom and Kim, 2005] quantifies the strength of the difference between groups. For instance, how strong is the difference between developers with direct experience in data privacy and those with indirect experience for the statement **[S1]**. We used p-values to test whether the observed differences between the two groups were statistically significant at a 95% confidence level (p-value < 0.05). As a result, we found that 12 statements presented statistically significant differences between the groups. To interpret the Cliff's Delta (d) effect size, we employ a well-known classification [Romano *et al.*, 2006], that defines four categories of magnitude, which are represented in Table 3: negligible (without symbol), small (*), medium (**), and large (***). The positive d magnitudes are represented by the (+) symbol and the negative ones are represented by the (−) symbol.

To illustrate the interpretation of (d) measure, Table 3 provides the following example: for statement **[S18]**, developers with direct experience in data privacy scored a mean of 4.19, while those with indirect experience scored a mean of 3.76. The associated p-value is 0.05, indicating statistical significance. The computed Cliff's Delta (d) is 0.2015, positive, and classified as small according to the magnitude categories [Romano *et al.*, 2006]. This result indicates that developers with direct experience agree more strongly with statement **[S18]** compared to those with indirect experience.

Qualitative analysis. We applied GT procedures to analyze responses to the open-ended questions (Q19, Q21, Q23), where participants optionally described their disagreement, with any statements under analysis.

3.4 Analysis to Answer RQ2

Similar to the previous question, to answer RQ$_2$, we conducted both quantitative and qualitative analyses, which we describe in detail as follows.

Frequency of use of data privacy strategies within development phases. Participants were asked to indicate how often they applied privacy strategies across specific phases of the software development lifecycle (Q24). The response options ranged from "Never" to "Always", capturing variations in frequency. The development phases considered included Requirements Analysis, Design, Coding, Feasibility Study, Installation, Deployment, Testing, and Maintenance. The analysis involved the following steps: (1) Responses were categorized by frequency level ("Never", "Rarely", "Sometimes", "Often", "Always") for each development phase. This step enabled a clear understanding of how privacy strategies were distributed across the lifecycle, and (2) We calculated the proportion of participants selecting each frequency level for each development phase. This allowed us to identify phases where privacy strategies were most and least utilized.

Correlations analysis on the perceptions of frequency, perceived importance, and ease of use of data privacy strategies. To evaluate developers' perceptions of data privacy strategies, we analyzed data on the frequency of use (Q25), perceived importance (Q27), and ease of use (Q29) for 13 strategies, based on responses from 88 developers. In this context, to consider the varying levels of expertise among the developers, we adapted a methodology inspired by a previous study [Dias-Neto *et al.*, 2017]. This approach involved three key steps, detailed as follows:

Step 1 - Weight Assignment: Each participant was assigned a weight based on their years of experience, level of seniority, academic degree, and the relation between direct experience and expertise in data privacy (Eq.1). The assigned weight reflects the participant's overall expertise, considering both formal qualifications and practical experience.

$$W(i) = E(i) + S(i) + \left(\frac{Y(i)}{\text{Median}(Y)}\right) + R(i), \text{where:} \quad (1)$$

- $W(i)$: is the total weight assigned to the participant i;
- $E(i)$: is the highest academic degree level of the participant i, such as (1) High School, (2) Bachelor's degree, (3) Specialization, (4) Master Student, (5) Master's degree, (6) PhD Student, and (7) Doctoral degree;
- $S(i)$: is the seniority level reported by the participant i in they current position, such as (1) Intern/Trainee, (2) Junior (up to 5 years), (3) Mid-Level (6 to 9 years), and (4) Senior (10+ years);
- $Y(i)$: is the number of years of experience reported by the participant i in software development, such as (1) Less than 1 year, (2) Between 1 and 3 years, (3) Between 4 and 6 years, (4) Between 7 and 14 years, and (5) More than 15 years;
- Median(Y): is the median of years of development experience, considering the answers from all participants; and
- $R(i)$: represents the relationship value between direct experience and expertise in data privacy.

We explain the $R(i)$ as follows.

$$R(i) = \frac{D(i) + X(i)}{2}, \text{where:}$$

- $R(i)$: is the value of the relationship between direct experience and expertise in data privacy;
- $D(i)$: indicates direct experience in data privacy, with a value of 1 if the developer has direct experience in data privacy, otherwise 0;
- $X(i)$: denotes the developer's expertise in data privacy, assigned a value of 1 if the developer is an expert in data privacy, i.e., if their current position is a Cryptographer, Privacy engineer, Penetration tester or Security engineer, otherwise 0.

In summary, $R(i)$ computes the average of two binary attributes, $D(i)$ and $X(i)$, to reflect the participant's overall relationship between direct experience and professional expertise in data privacy. The interpretation for each case is as follows:

- $D = 0, X = 0$: The participant has neither direct experience nor professional expertise in data privacy. In this case, $R = 0$, indicating no relation;
- $D = 1, X = 0$: The participant has direct experience but is not considered an expert. Here, $R = 0.5$, represents a moderate involvement.
- $D = 0, X = 1$: The participant is considered an expert but does not have direct experience. Similarly, $R = 0.5$.
- $D = 1, X = 1$: The participant possesses both direct experience and professional expertise in data privacy. In this case, $R = 1$, indicating the highest level of relevance.

Step 2 - Weighted Answers: After assigning weights, each participant's response to the frequency of use, perceived importance, and ease of use of the 13 strategies was multiplied by their corresponding weight. This results in a weighted score for each strategy. The total value for each strategy is then computed by summing these weighted responses across all participants, as defined in Eq.2.

$$T(j) = \sum_{i=1}^{n} (\text{Answer(i, j)} \times \text{W(i)}), \text{where:} \quad (2)$$

- $T(j)$: is the total value of frequency of use, perceived importance, and ease of use for the data privacy strategy j;
- $Answer(i, j)$: is the answer value ranging from 1 to 5, relating to the frequency of use, perceived importance, and ease of use of participant i in for the data privacy strategy j;
- $W(i)$: is the weight for the participant i.

Step 3 - Normalization: Finally, a normalized value was calculated for the levels of use, perceived importance, and ease of use, ranging from 0 to 100%, by normalizing the value obtained in Step 2 for each data privacy strategy, i.e., divide the value achieved in the previous step by the maximum possible value (as defined in Eq.3).

$$N(j) = \frac{T(j)}{\sum_{i=1}^{n}(\text{W(i)}) \times MaxScorePerParticipant}, \text{ where:} \tag{3}$$

- $N(j)$: is the normalized value for the frequency of use or perceived importance or ease of use of a data privacy strategy j;
- $T(j)$: is the total value calculated for strategy j in Step 2;
- $W(i)$: is the weight value for the participant i;
- $MaxScorePerParticipant$: is the maximum score a participant can provide (e.g., on a 5-point Likert scale).

The normalized value (Eq.3) is used in Table 5 to compare the frequency of use, perceived importance, and ease of use for each data privacy strategy.

To demonstrate the application of the formula, we use the perceived importance answers for the Encryption strategy (Table 5) as an example as follows:

Calculate the Total Perceived Importance Value: The total perceived importance (T) for this strategy is computed by multiplying each developer's weight ($W(i)$) by their corresponding response ($Answer(i, Encryption)$) and summing the results across all participants (n). In this case, (T(Encryption)) $\approx$ 1136.98:

$$T(\text{Encryption}) = \sum_{i=1}^{n}(\text{Answer(i, Encryption)} \times \text{W(i)})$$

Normalize to Determine the Perceived Importance Level: The level of perceived importance is determined as a percentage, ranging from 0% to 100%, by normalizing the total importance value. This is achieved by dividing $T(Encryption)$ by the sum of the possible weights for all participants $\sum_{i=1}^{n}(\text{W(i)})$, multiplied by 5. The normalization ensures the results are comparable across all strategies, regardless of the number of participants or their weights.

$$N(j) = \frac{1136.98}{241.85 \times 5} \approx 0.940$$

The result, $N(j) \approx 0.940$, indicates that the perceived importance of the Encryption strategy is approximately 94.02%.

Qualitative analysis. We also applied GT procedures to analyze responses to the open-ended questions (Q26 and Q28), both optional questions. For Q26, participants were asked whether they used any additional implementation strategies to ensure privacy beyond those mentioned in the previous question (Q25), and if so, to specify them. This question aimed to uncover unique or less conventional strategies that could complement or expand the predefined set of strategies. For Q28, participants provided explanations for rating certain strategies as important. These responses added depth and context, shedding light on participants' decision-making processes, priorities, and the specific challenges they faced in implementing privacy strategies.

3.5 Analysis to Answer RQ3

To address RQ$_3$, we conducted both qualitative and quantitative analyses to uncover how organizations manage privacy and data protection and the influence of organizational factors on adopting privacy strategies.

Qualitative analysis. We analyzed responses to two open-ended questions (Q31 and Q34) where 88 participants described their organization's practices for managing privacy and data protection, as well as their perception of future challenges for organizations in protecting user privacy rights. Through this analysis, we identified nine distinct categories of practices related to privacy management and eight categories of challenges. We also compare our findings with prior studies. Additionally, we compared our findings with prior studies to contextualize and validate the results. **Quantitative analysis.** We examined how organizational structures influence the adoption of privacy strategies. Our analysis focused on three key aspects: (1) the presence of dedicated privacy teams (Q30); (2) tool adoption and usage patterns (Q32); and (3) perceived organizational priority (Q33).

4 Results and Discussion

As mentioned in Section 3.2, our data sample is the 88 responses for analysis. We considered only participants who reported a direct or indirect data privacy-related work experience. For brevity, we use the notation (n/88) along the text to denote a number n out of the 88 of valid participants.

General characterization. The participants are distributed across 11 Brazilian states (Q2), with a significant predominance in Ceará (28 participants, representing 35% of the total), followed by the Federal District (21 participants, 26%) and São Paulo (16 participants, 20%). Other states with lower representation include Alagoas (5 participants), Rio de Janeiro (3 participants), Minas Gerais (3 participants), Paraíba (2 participants), Paraná (2 participants), Pernambuco, Santa Catarina, and Tocantins (1 participant each). This distribution reflects a regional concentration in the Northeast and Central-West, with a notable representation from the Southeast.

The participants are predominantly male (Q3), representing 85.2% (75/88) of our sample, with only 14.8% (13/88) of women, a trend commonly observed in tech-related fields [Ashcraft *et al.*, 2016]. In terms of age (Q1), the majority of participants are in the 25-34 year range, comprising 44.3% (39/88), followed by 35-44 years at 25% (22/88), 18-24 years at 20.5% (18/88), and one participant (1.1%) in the 55-64 year age group. Additionally, one participant (1.1%) preferred not to say their age. Regarding their academic background (Q4), most participants have at least a bachelor's degree, with 29.5% (26/88) holding a bachelor's degree. Furthermore, 19.3% (17/88), have completed a specialization, while 13.6% (12/88) have only completed high school. Additionally, around 20% of participants are pursuing or have obtained higher education degrees: 11.4% (10/88) are master's students, 10.2% (9/88) hold a doctoral degree, and another 10.2% (9/88) have completed a master's degree. Finally, 5.7% (5/88) are PhD students. This indicates that our sample has a highly academic background.

We asked developers about their current work model type (Q5), with the majority, 62.5% (55/88) working remotely, reflecting a common trend in the tech industry, especially post-pandemic [Ralph *et al.*, 2020]. While Hybrid 18.2% (16/88) and in-person models with 19.3% (17/88) are less frequent, they still account for a substantial portion of the sample (around 37%). Additionally, most participants carry out traditional positions within software development projects (Q6). Especially, 50.6% (44/88) are developers (including backend, frontend, and fullstack), and 10.3% (9/88) work as security engineers. Other positions include project leads and data scientists each with (5.7%, 5/88). Also, appears test managers or testers, solution architects, and privacy engineers each representing 3.4% (3/88), followed by product owners and penetration testers with each representing 2.3% (2/88). Other roles, including data engineers, UX/UI designers, database administrators, and IT auditors, also appear, each representing 1.1% (1/88). This suggests that most of the participants work directly in development roles.

The majority of participants are seniors (Q7), with 39.8% (35/88) having at least 10 years of experience or more in their fields, and 33% (29/88) holding a Mid-Level, i.e., 6 to 9 years of experience. In contrast, 21.6% (19/88) are Junior (with up to 5 years of experience), and 5.7% (5/88) hold Intern/Trainee positions. This distribution indicates a well-experienced sample, with nearly 40% of participants in senior roles, reflecting the depth of expertise among the sample. Regarding years of experience in software development (Q8), the majority of participants are highly experienced, with 34.1% (30/88) having between 7 and 14 years of experience, and 19.3% (17/88) having over 15 years. Additionally, 26.1% (23/88) have 4 to 6 years, 13.6% (12/88) have 1 to 3 years, and 6.8% (6/88), have less than 1 year of experience, providing some diversity in terms of work history.

We also asked the participants about the type of systems they worked with (Q9) and system domain (Q10). These were multiple-choice questions so that participants could select as many options as they wanted. Figure 1 shows that the most common type of system was web applications at (78), followed by mobile applications at (41), and desktop applications at (27). Other system types included developer tools (22), operating systems (16), middleware (12), and embedded applications (10).

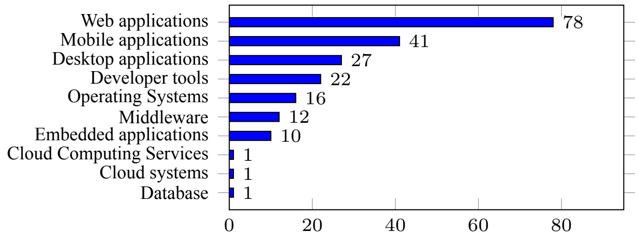

Figure 1. System type (top 10)

Figure 2 shows that the most common system domain was baking/financial at (21), followed by education (18), sales/e-commerce (15), defense & security (13), telecommunication (12), and many others.

Characterization of the organizations. We asked participants on organization size (Q12) and sector (Q11) where they work. Most participants come from large organizations with more than 1000 employees (43.1%, 38/88), followed by organizations with 100 to 499 employees (13.6%, 12/88), 10 to 49 employees (12.5%, 12/88), up to 9 employees (11.3%, 10/88), 500 to 999 employees (6.8%, 6/88), and 50 to 99 employees (4.5%, 4/88). Only seven of the participants indicated that they did not know the size of their organization. Most participants work in private organizations (48.8%, 43/88), followed by those in federal public administration (22.7%, 20/88).

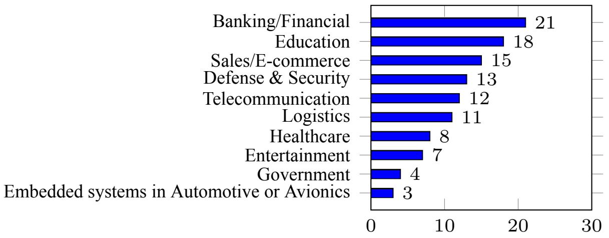

Figure 2. System domain (top 10)

Specific experience with data privacy within the development process. We also asked the participants if they had any work experience related to data privacy (Q13). A total of 61.4% (54/88)) of participants responded that have worked (or currently work) with data privacy in the software development process (e.g., design, development, and testing), while 38.6% (34/88) indicated having an indirect experience, stating that their team works (or has worked) on data privacy, but they were not directly involved in any tasks. The participants have reported their experience with data privacy in the Q14 (open-ended question), for instance:

> *#P79 said: "My experience with data privacy is related to the criticality of financial information, both for transactions and personal data of customers. We must handle the use and sharing of this data with extreme care to ensure compliance with the LGPD and other applicable laws/regulations. This involves data processing, storage issues, non-storage, and redirection."*

We also asked about the main sources or methods that participants usually use to learn about issues related to data privacy in software (Q15), and about what types of personal data they handle in their work (Q16). These were multiple-choice questions so that participants could select as many options as they wanted. The results, summarized in Figure 3, show that the most common sources include documentation and guidelines (73), followed by training programs (51), and tools and libraries (41). Other sources were lectures (29), workshops (20), seminars (17), mentorship (6), and hackathons (3).

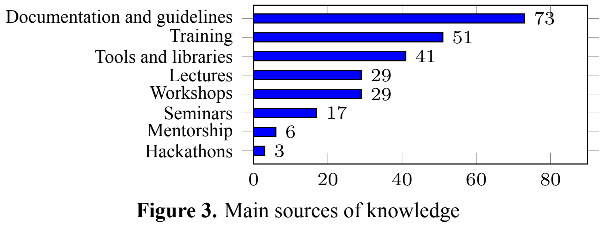

Figure 3. Main sources of knowledge

Figure 4 shows that the most frequently handled data type is basic contact identifiers, with 83 responses, which include data such as names, emails, phone numbers, and personal user IDs. This is followed by location information (42 responses), e.g., geolocation data, and online activity data (37 responses), which includes Website logs link IP address, user agents, and device IDs. Other types of personal data include financial information and legal or judicial information, both with 35 responses, indicating the handling of sensitive data related to finances or legal processes.

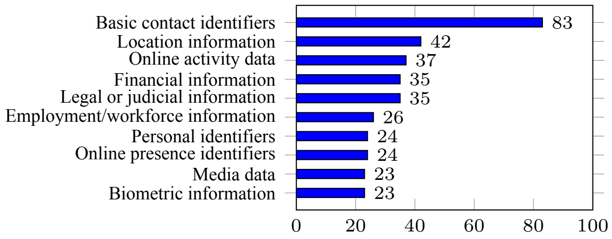

Figure 4. The top 10 types of personal data handled by participants

4.1 Developers' Awareness of Data Privacy (RQ1)

To answer **RQ$_1$**, we analyzed: (i) the most frequent words that immediately come to developers' minds when they think about data privacy (Q17); and (ii) the developers' agreement with 21 statements, considering different groups, across three dimensions: *Knowledge*, which assesses the developers' understanding of data privacy principles, laws, and best practices (Q18); *Attitudes and Feelings* which explores their beliefs, values, and emotional responses towards data privacy (Q20); and *Behaviors and Actions* which focuses on the actual behaviors and actions developers take concerning data privacy (Q22). We explain the data procedures used to answer RQ_1 in Section 3.3.

The most frequent words that developers associate with data privacy. We asked participants to *Provide at least five words in the order they come to mind when you think about privacy* (Q17). This question aimed to capture the immediate associations and key concepts developers link to data privacy. Figure 5 shows the frequency distribution of words most associated with data privacy by participants. The x-axis represents the position of the words in the order provided by the participants (from 1 to 5). These numbers correspond to the sequence in which participants listed the words.

Figure 5 reveals that the word *Security* appears most frequently at $ranking = 1,\ 2\ and\ 3$ with 33, 11, and 8 mentions, respectively. This suggests that security is the central and immediate idea linked to data privacy in the respondents' minds. At $ranking = 2$, the distribution of words become more varied, with *Protection* (7 mentions), *Confidentiality* (6 mentions), *Anonymization* (6 mentions) and *General Data Protection Law* (5 mentions). These associations indicate that participants link privacy not only to protective measures but also to legal frameworks, such as those outlined in the *General Data Protection Law*.

At $ranking = 3$, the frequency of words declines, with *Encryption* appearing for the second time, with four mentions. This indicates that privacy concerns extend to more specific issues, such as data privacy strategies such as encryption. At $ranking = 4$, *Protection* appears for the third time with six mentions, followed by *Trust* and *Control* with five mentions, and *Confidentiality* and *People* with four mentions. Finally, at $ranking = 5$ *Consent* appears as the most mentioned with four mentions, followed by *Protection*, *Trust*, *Access Control*, and *Integrity*, each with three mentions. This progression reflects a shift from more general concepts of security and protection to a mix of technical, legal, and interpersonal factors that play a role in data privacy.

> **Finding 1:** Developers frequently associate the word *security* with data privacy, as it ranked highest across the first three positions in the sequence provided by participants (33 mentions at ranking 1, 11 at ranking 2, and 8 at ranking 3). This highlights security as the primary and immediate concept linked to privacy.

Developers' awareness, perspectives, and practices. Table 3 overviews the results on awareness of data privacy statements. We used color coding to indicate which group was more likely to agree with each statement: a Dark Grey color indicates the former group had a higher likelihood of agreement, while a Light Grey color indicates the latter group was more likely to agree. Additionally, we annotated each statement to indicate statistical significance between groups as follows: (1) 🔧 indicates a significant difference between participants with **Direct** data privacy work experience and those with **Indirect** data privacy work experience; (2) 📱 denotes a significant difference between groups **Specialist** v.s. **Non-Specialist** roles; and (3) 👥 indicates a significant difference between participants working in **Large** v.s. **Small** organizations. For instance, **[S1]** shows there are no significant differences between any participant groups regarding their perception of the statement. On the other hand, **[👥 📱 S2]** shows that there is a significant difference between participant groups with **Big** and **Small** organizations, as well as between groups of **Specialist** and **Non-Specialist**, regarding their perception of S2.

🔧 Direct vs. Indirect Data Privacy Work Experience. Direct Privacy Work Experience showed a higher impact on knowledge of privacy laws **[S1]**, initiative to learn **[S4]**, and addressing privacy issues **[S16]**. Participants with direct privacy work experience reported slightly higher awareness of relevant privacy laws and standards, with a near-significant difference (p=0.07). Direct-experience participants showed greater initiative in independently learning about privacy, a trend supported by a marginally significant result (p=0.08). Identifying privacy issues during development was more common among participants with direct privacy roles, as evidenced by the significant difference (p=0.10). These findings suggest that direct involvement in privacy-related work positively impacts self-reported awareness and proactive learning behaviors.

👥 Big vs. Small Organizations. Participants from larger organizations were significantly more likely to have participated in privacy training (p=0.01) **[👥 S3]**, indicating structured opportunities in larger workplaces. Moreover, they scored higher on understanding relevant laws, such as ISO/IEC 29100 and Privacy by Design **[👥 S1]**, with a statistically significant difference (p=0.02). Additionally, they

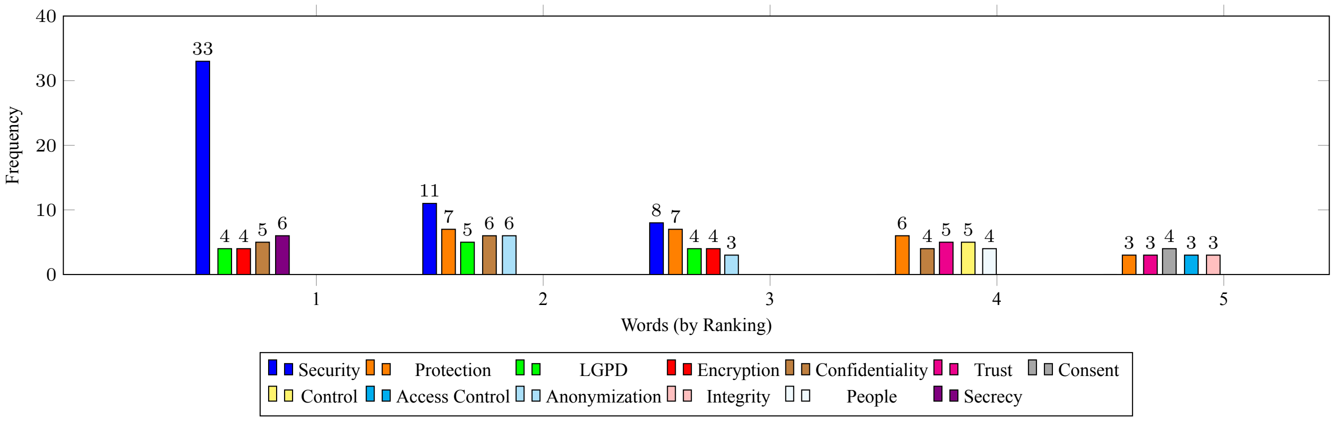

Figure 5. Frequency of the top-5 most mentioned words by ranking.

Table 3. Survey results on awareness of data privacy statements

Statement	ID	Likert	Direct v.s Indirect				Big v.s. Small				Specialist v.s. Non-Specialist			
		Score	Score	Score	P-value	*d*	Score	Score	P-value	*d*	Score	Score	P-value	*d*
Topic 1: Knowledge about Data Privacy in Software														
I have knowledge of relevant privacy laws for my field, such as ISO/IEC 29100 and Privacy by Design.	S1	3.43	3.61	3.15	0.07	(+)*	3.63	3.08	**0.02**	(+)*	4.29	3.17	**0.00**	(+)***
I have a solid understanding of the specific privacy laws and regulations that apply to my work and how they influence software development.	S2	3.65	3.67	3.62	0.44	(+)	3.88	3.32	**0.04**	(+)*	4.43	3.38	**0.00**	(+)***
I have participated in internal training or industry workshops on security and privacy, including internal privacy policies and industry-specific regulations.	S3	3.56	3.59	3.50	0.35	(+)	3.95	3.16	**0.01**	(+)*	3.93	3.40	0.08	(+)*
I strive to learn about privacy on my own initiative, including reading privacy laws and regulations and consulting experts.	S4	3.82	3.94	3.62	0.08	(+)*	3.88	3.80	0.45	(+)	4.43	3.66	**0.00**	(+)**
I recognize the risks and concerns related to data privacy in software systems, such as personal data leakage, unauthorized access, and lack of control over user data.	S5	4.57	4.57	4.56	0.51	(-)	4.61	4.56	0.76	(-)	4.79	4.52	0.12	(+)*
I am aware of best practices and techniques for protecting user privacy in software systems, including data anonymization and access control practices.	S6	3.90	3.98	3.76	0.22	(+)	4.00	3.76	0.11	(+)*	4.64	3.69	**0.00**	(+)***
I understand the difference between security and privacy, recognizing that privacy goes beyond data protection and includes aspects such as control over data and informed consent.	S7	4.13	4.19	4.03	0.25	(+)	4.16	4.08	0.44	(+)	4.29	4.05	**0.04**	(+)*
Topic 2: Attitudes and Feelings about Data Privacy in Software														
I am concerned about the possibility of being monitored or manipulated when using apps, social media, or browsing the internet.	S8	4.30	4.30	4.29	0.65	(-)	4.32	4.20	0.40	(+)	4.71	4.25	**0.02**	(+)*
I believe that personal privacy is a fundamental right and that we should be vigilant about privacy violations.	S9	4.60	4.61	4.59	0.36	(+)	4.55	4.68	0.89	(-)	4.50	4.63	0.66	(-)
I feel a personal responsibility to protect user privacy in my work as a software developer.	S10	4.42	4.44	4.38	0.46	(+)	4.38	4.44	0.61	(-)	4.64	4.42	0.20	(+)
I perceive that many people do not care about privacy, but I believe it is still important to educate and raise awareness about privacy rights.	S11	4.34	4.33	4.35	0.43	(+)	4.18	4.56	0.98	(-)*	4.64	4.29	0.10	(+)*
I feel frustrated with the idea that privacy is unattainable in today's digital society.	S12	3.58	3.48	3.74	0.86	(-)	3.50	3.56	0.58	(-)	3.79	3.62	0.30	(+)
I value informed consent from users before collecting data and believe it is essential for a trusting relationship between the user and the developer.	S13	4.28	4.43	4.06	0.06	(+)*	4.14	4.48	0.95	(-)*	4.07	4.32	0.66	(-)
I recognize that the retention of personal data should be limited and depends on the type of data and the purpose of the system.	S14	4.58	4.54	4.65	0.68	(-)	4.55	4.68	0.89	(-)	4.71	4.57	0.27	(+)
I consider privacy to be a shared responsibility of the entire development team.	S15	4.51	4.61	4.35	0.06	(+)*	4.45	4.60	0.89	(-)*	4.71	4.51	0.13	(+)*
Topic 3: Behaviors and Actions regarding Data Privacy in Software														
Identifying privacy issues during development activities is an important part of my professional routine.	S16	3.75	3.87	3.56	0.10	(+)*	3.66	3.84	0.79	(-)	4.29	3.58	**0.00**	(+)**
When I encounter privacy issues, I direct them to project leaders, more experienced colleagues, or alert security operations teams.	S17	4.28	4.31	4.24	0.41	(+)	4.20	4.36	0.91	(-)*	4.43	4.22	0.18	(+)
I propose solutions to privacy issues encountered during software development.	S18	4.02	4.19	3.76	**0.05**	(+)*	3.95	4.08	0.83	(-)	4.43	3.91	**0.03**	(+)*
I have played the role of a privacy advocate (”privacy champion”) in my team or company.	S19	2.77	2.81	2.71	0.35	(+)	2.82	2.76	0.40	(+)	3.57	2.54	**0.00**	(+)**
When dealing with privacy conflicts with clients, I seek to negotiate with the client to implement privacy controls.	S20	3.43	3.43	3.44	0.50	(+)	3.39	3.48	0.57	(-)	4.00	3.25	**0.01**	(+)**
I have faced suspicious requests for excessive data collection from clients.	S21	2.76	2.57	3.06	0.94	(-)*	3.05	2.20	**0.01**	(+)**	3.79	2.48	**0.00**	(+)***

often have applied this knowledge in a daily basis work (p=0.04) **[👥 S2]**. This trend reflects a greater emphasis on formal privacy education and resources in larger companies compared to smaller ones.

📱 **Specialist vs. Non-Specialist Roles.** Specialists consistently reported significantly higher knowledge of privacy laws and their application (p<0.00 for **[📱 S1]**, p=0.04 for **[📱 S2]**). Moreover, this knowledge is obtained by their initiative, including reading privacy laws and regulations and consulting experts (p<0.00 for **[📱 S4]**). Moreover, they demonstrated greater awareness of risks like unauthorized access and data leakage **[S5]**, though this difference was not statistically significant (p=0.12). However, a statistically significant (p<0.00) was observed concerning their awareness of best practices and techniques for protecting user privacy in software systems, including data anonymization and access control practices **[📱 S6]**. A similar observation applies to the **[📱 S7]** with p=0.04, regarding the knowledge that privacy goes beyond data protection and includes aspects such as control over data and informed consent.

Additionally, the specialists are more likely to the concerned about the possibility of being monitored or manipulated when using apps, social media, or browsing the internet (p<0.02 for **[▯ S8]**). Also, they were far more likely to identify privacy issues during development activities (p<0.00 for **[▯ S16]**), to propose solutions to privacy issues encountered during software development (p=0.03 for **[▯ S18]**), and serve as "privacy champions" in their organizations (p<0.00 for **[▯ S19]**). Similarly, specialists were more active in negotiating privacy-related conflicts with clients (p=0.01 for **[▯ S20]**), and have faced suspicious requests for excessive data collection from clients (p<0.00 for **[▯ S21]**). These results emphasize the crucial role of specialists in advancing privacy awareness and embedding privacy practices within development processes.

Cross-Group Comparisons. We observed significant differences when multiple factors were examined together: (1) Participants with direct experience and those specializing in privacy were more likely to propose solutions to privacy issues (p=0.05, and p=0.03 for **[🔧 ▯ S18]**); (2) Similarly, the groups that work in big organizations and the specialists have faced suspicious requests for excessive data collection from clients (p=0.01, and p<0.00 for **[👥 ▯ S21]**). These observations show the synergistic effects of experience, specialization, and organization size in enhancing privacy practices.

In general, participants demonstrated a moderate to high level of agreement on statements related to awareness of data privacy, suggesting that awareness of privacy-related concepts is widespread but varies significantly across certain groups and contexts. Thus, our findings highlight the need for tailored strategies to improve awareness and integration of privacy practices in diverse software development contexts. Complementary, we asked participants to describe the reasons for their classification regarding statements they strongly disagreed with (Q19). The majority of participants who answered this question stated that it was due to a lack of knowledge about privacy laws or a lack of training.

> *#P16 said: "I haven't read any specific regulations regarding privacy, nor have I received any training on this in my organization. There are people on the team with more experience who handle this activity[...]"*

For the statements with which the participant strongly disagreed in (Q20), we asked them to describe the reasons for their classification (Q21). The participants who answered this question expressed disagreement, citing the lack of standardization in defining what constitutes privacy, insufficient awareness of privacy practices within organizations, and the absence of specialized teams focused on privacy and data protection in organizations.

> *#P101 said: "I believe software development teams are already overwhelmed with other technical and managerial tasks. It is necessary for organizations to have a dedicated team for data privacy and personal data handling to oversee and support the development team during their activities."*

Participants strongly disagreed (Q23) because they believe there is a lack of awareness among people about the importance of data privacy.

> *#P116 said: "It is still very difficult to convince people of the importance of privacy. While the resistance was worse in the past, it still persists. I believe people are still coming to understand the benefits of privacy."*

Finding 2: Awareness of data privacy statements is nuanced, with variations driven by 🔧 direct privacy experience, 👥 organizational size, and ▯ professional specialization. These differences underscore the importance of targeted interventions, such as privacy training in smaller organizations and empowering non-specialists to engage with privacy concepts.

4.2 Developers' Perception of Data Privacy Strategies (RQ2)

We address **RQ**$_2$ by analyzing the developers' perception of 13 data privacy strategies based on three factors: their frequency of use (Q25), perceived importance (Q27), and ease of use (Q29). Additionally, we examined responses from Q24 regarding the frequency of data privacy strategy usage across development phases, as well as insights from the open-ended questions in Q26 and Q28. The data procedures employed to answer RQ_2 are detailed in Section 3.4.

Frequency of use of data privacy strategies within development phases. Table 4 overviews the response for the Q24, about how frequently developers apply data privacy strategies across different software development phases.

Table 4. Frequency of privacy strategies in development phases

Development Phase	Frequency					
	Never	Rarely	Sometimes	Often	Always	Total
Requirements analysis	3 (3.41%)	9 (10.23%)	19 (21.59%)	38 (43.18%)	19 (21.59%)	88
Design	11 (12.5%)	15 (17.05%)	27 (30.68%)	20 (22.73%)	15 (17.05%)	88
Coding	4 (4.55%)	8 (9.09%)	17 (19.32%)	21 (23.86%)	38 (43.18%)	88
Viability study	10 (11.36%)	14 (15.91%)	23 (26.14%)	23 (26.14%)	18 (20.45%)	88
Instalation	15 (17.05%)	14 (15.91%)	27 (30.68%)	17 (19.32%)	15 (17.05%)	88
Deploy	14 (15.91%)	15 (17.05%)	19 (21.59%)	24 (27.27%)	16 (18.18%)	88
Testing	8 (9.09%)	8 (9.09%)	24 (27.27%)	21 (23.86%)	27 (30.68%)	88
Maintenance	5 (5.68%)	7 (7.95%)	22 (25%)	29 (32.95%)	25 (28.41%)	88

We can observe that *Coding* and *Maintenance* phases show relatively higher adoption of data privacy strategies with significant counts of Always and Often responses (38 and 25 Always, respectively, and 21 and 29 Often). Moreover, the *Requirements Analysis* phase demonstrates a mix of frequent usage (Always = 19, Often = 38) and occasional usage (Sometimes = 19), suggesting it is a critical but variably addressed phase. Phases such as *Design and Deploy* show relatively fewer Always responses (15 and 16, respectively) and more moderate usage (Sometimes and Rarely). This indicates a potential gap in integrating privacy strategies during these phases. Finally, the *Viability Study* and *Installation* phases have fewer developers applying privacy strategies consistently (Always = 18 and 15), and a notable number of Never responses (10 and 15), indicating limited prioritization in these areas.

Finding 3: Data privacy strategies are consistently used in the *Coding*, *Testing*, and *Maintenance* phases. In contrast, the *Design* and *Viability Study* phases showed a less consistent adoption, highlighting areas for improvement in integrating privacy practices across the entire development lifecycle.

Analysis of the frequency of use, perceived importance, and ease of use of data privacy strategies. Table 5 presents a comparative analysis of various data privacy strategies based on three metrics: frequency of use, perceived importance, and ease of use. The data are presented in a heatmap format, where color intensities reflect the magnitude of the corresponding metric values for each strategy. In this context, the greener color highlights strategies that are widely used, highly important, and considered easy to use. Conversely, the red color denotes strategies that are less frequently adopted, perceived as less important, or more difficult to use. In other words, the greener a strategy appears, the more it is recognized for its utility, importance, or practicality. Conversely, redder tones suggest lower prioritization or adoption in practice.

Table 5. Comparison on the frequency of use, perceived importance, and ease of use of data privacy strategies

Data Privacy Strategy	Frequency of Use	Perceived Importance	Ease of Use
Encryption	76.46%	94.02%	68.51%
Minimization of personal data collection	71.40%	85.76%	71.82%
Decentralization	58.58%	75.08%	57.07%
Data sovereignty	67.79%	82.21%	60.68%
Temporal data	64.53%	78.10%	65.66%
User control	77.20%	86.89%	69.41%
Turning off data collection	56.20%	73.65%	72.68%
Anonymization	69.07%	84.74%	62.04%
Data classification tools	60.45%	77.13%	61.38%
Design and code reviews	72.62%	83.78%	66.44%
Risk management	70.28%	89.86%	61.15%
Data flow modeling	69.12%	83.25%	63.48%
Proxy	57.61%	72.22%	59.76%

By analyzing the color patterns, we can observe that the three most frequently used strategies were: *User control* (77.20%), *Encryption* (76.46%), and *Design and code review* (72.20%). Regarding the privacy strategies with the highest perceived importance, we can observe that most of the strategies are perceived as important, in which the *Encryption* (94.02%), *Risk management* (89.86%), and *User control* (86.89%) strategies appear in the top 3. Conversely, most of the strategies appear as not easy to use, except for *Turning off data collection* (72.68%) and *Minimization of personal data collection* (71.82%) which present moderate values. Additionally, we also can observe that *Turning off data collection* has the lowest frequency of use (56.20%) despite its ease of use and moderate importance. In summary, this analysis helps prioritize strategies based on their frequency of use, perceived importance, and ease of use.

Finding 4: While certain strategies like *Encryption* are both widely used and perceived as highly important, others, such as *Turning off data collection*, highlight strategies where ease of use does not necessarily lead to widespread adoption.

Analysis of correlation between frequency of use, perceived importance, and ease of use. To better understand the relation between the data privacy strategies, we assessed the correlation between the frequency of use, perceived importance, and ease of use reported by participants (N = 88).

We hypothesize that $\mathbf{HA_1}$**:** There is a strong correlation between frequency of use and perceived importance in the $[\mathit{privacy\ strategy}]_n$. The null hypothesis is $\mathbf{HA_0}$**:** There is no strong correlation between frequency of use and perceived importance in the $[\mathit{privacy\ strategy}]_n$. Additionally, we hypothesize that $\mathbf{HB_1}$**:** There is a strong correlation between perceived importance and ease of use in the $[\mathit{privacy\ strategy}]_n$. The null hypothesis is $\mathbf{HB_0}$**:** There is no strong correlation between perceived importance and ease of use in the $[\mathit{privacy\ strategy}]_n$. Similarly, we hypothesize that $\mathbf{HC_1}$**:** There is a strong correlation between ease of use and frequency of use in the $[\mathit{privacy\ strategy}]_n$. The null hypothesis is $\mathbf{HC_0}$**:** There is no strong correlation between ease of use and frequency of use in the $[\mathit{privacy\ strategy}]_n$.

We applied the Shapiro-Wilk test [Shapiro and Wilk, 1965] to assess our data distributions. We confirmed that our data is not normally distributed. Thus we decided to compute the Spearman's rank correlation coefficient [Wohlin *et al.*, 2012]. We considered a confidence interval of 95% (p-value < 0.05). As a result, we obtained a p-value < 0.01 for all data privacy strategies across all correlations. Thus, the computed correlations have statistical significance for all data privacy strategies.

Table 6 presents the correlation results. The first column lists each data privacy strategy The second column presents the correlation between perceived importance and frequency of use. The third column presents the correlation between perceived importance and ease of use. Finally, the last column presents the correlation between ease of use and frequency of use. We have categorized the correlation values according to a previous work [Salkind, 2012]: Very strong relationship (0.8 to 1.0 or -0.8 to -1.0); Strong relationship (0.6 to 0.8 or -0.6 to -0.8); Moderate relationship (0.4 to 0.6 or -0.4 to -0.6); Weak relationship (0.2 to 0.4 or -0.2 to -0.4); and Weak or no relationship (0.0 to 0.2 or 0.0 to -0.2).

By analyzing Table 6, we can observe that there is a high correlation between **perceived importance and frequency of use** across data privacy strategies. More specifically, privacy strategies such as *User control* (ρ=0.785), *Encryption* (ρ=0.770), and *Risk management* (ρ=0.756) exhibit strong correlations, indicating that users tend to frequently adopt these strategies when they perceive them as highly important. However, the *Turning off data collection* (ρ=0.55) shows a moderate correlation, suggesting that their perceived importance might not be always correlated to the frequency of use. Thus, our results confirm $\mathrm{HA_1}$ for all data privacy strategies, except for the *Turning off data collection* strategy. For this strategy, we reject $\mathrm{HA_1}$ (and confirm $\mathrm{HA_0}$).

The correlation between **perceived importance and ease of use** showed that strategies such as *User control* (ρ=0.696), *Minimization of personal data collection* (ρ=0.691), and *Encryption* (ρ=0.676) present strong correlations, indicating that ease of use significantly influences how important users perceive these strategies. Conversely, strategies such as *Temporal data* (ρ=0.514) and *Anonymization* (ρ=0.506) have moderate correlations, suggesting usability challenges may decrease their perceived importance. Thus, we confirm $\mathrm{HB_1}$

Table 6. Spearman's correlation between frequency of use, perceived importance, and ease of use of data privacy strategies

Data Privacy Strategy	Correlation(Importance, Use)		Correlation(Importance, Ease)		Correlation(Ease, Use)	
	Stat (rho)	Category	Stat (rho)	Category	Stat (rho)	Category
Encryption	0.770000	Strong	0.676000	Strong	0.536000	Moderate
Minimization of personal data collection	0.719000	Strong	0.691000	Strong	0.557000	Moderate
Decentralization	0.661000	Strong	0.633000	Strong	0.662000	Strong
Data sovereignty	0.736000	Strong	0.621000	Strong	0.567000	Moderate
Temporal data	0.643000	Strong	0.514000	Moderate	0.438000	Moderate
User control	0.785000	Strong	0.696000	Strong	0.601000	Strong
Turning off data collection	0.550000	Moderate	0.525000	Moderate	0.388000	Weak
Anonymization	0.658000	Strong	0.506000	Moderate	0.530000	Moderate
Data classification tools	0.635000	Strong	0.574000	Moderate	0.443000	Moderate
Design and code reviews	0.647000	Strong	0.609000	Strong	0.533000	Moderate
Risk management	0.756000	Strong	0.568000	Moderate	0.427000	Moderate
Data flow modeling	0.673000	Strong	0.510000	Moderate	0.465000	Moderate
Proxy	0.647000	Strong	0.594000	Moderate	0.421000	Moderate

for the following strategies: *User control*, *Data sovereignty*, *Decentralization*, *Minimization of personal data collection*, *Design and code reviews* and *Encryption*. For the other strategies, we reject HB_1 (and confirm HB_0).

Regarding the correlation between **ease of use and frequency of use**, most strategies exhibit moderate correlations in this category, such as *Proxy* (ρ=0.421), *Risk management* (ρ=0.427) and *Data flow modeling* (ρ=0.465) indicating that usability somewhat affects adoption but may not be the sole factor. Conversely, the *Decentralization* (ρ=0.662) and *User control* (ρ=0.601) strategies was the only two that demonstrated a strong correlation, indicating that the ease of use has a positive impact on the adaptation of these strategies. Additionally, *Turning off data collection* (ρ=0.388) strategy was the only one that demonstrated a weak correlation, suggesting that ease of use has a limited impact on the frequency of adoption. Thus, we confirm HC_1 only for the *Decentralization* and *User control* strategies. For the other strategies we reject HC_1 (and confirm HC_0).

Finding 5: The majority of data privacy strategies 92% (12/13) presented a strong correlation between perceived importance and frequency of use. Additionally, 38% (5/13) showed a strong correlation between perceived importance and ease of use, while 61% (8/13) exhibited a moderate correlation. Finally, the ease of use and frequency of use were moderately correlated in 84% (11/13) of the strategies, with one strategy presenting a weak correlation.

Specific data privacy strategies across correlations. When we look for patterns of specific strategies, we observed that *Encryption*, *Minimization of personal data collection*, *Data sovereignty*, and *User control* presented a high correlation across all three dimensions (Strong for both Importance–Use and Importance–Ease; and Moderate for Ease–Use). This suggests that these privacy strategies are both valued and frequently used, despite moderate usability challenges. On the other hand, *Decentralization* was the only strategy that demonstrated a strong correlation across dimensions, suggesting a consistent alignment between perceived importance, ease of use, and frequency of use. Additionally, *Turning off data collection* was the only strategy that demonstrated the weakest correlations in all dimensions, particularly in Ease-Use (ρ=0.367). This may reflect usability barriers or limited perceived relevance in specific contexts. Conversely, the *User control* strategy has the highest correlations in Importance-Use (ρ=0.777), emphasizing the need for tools that empower users to manage their data.

We also asked participants whether they use any additional implementation strategies to ensure data privacy (Q26) that were not mentioned in the previous question (Q25), and if so, what those strategies were. Only seven participants responded to this question and mentioned the use of tools such as *Static Application Security Testing (SAST)*, *Dynamic Application Security Testing (DAST)*, *Data Loss Prevention (DLP)*, and *Virtual Private Network (VPN)*.

Finding 6: Privacy strategies such as *Encryption*, *Decentralization*, and *User control* emerge as effective and widely adopted, reflecting their strong alignment between perceived importance, usability, and frequency of use. In contrast, strategies such as *Turning off Data Collection* and *Temporal Data* highlight the need for targeted improvements to enhance their usability and perceived relevance, encouraging broader adoption.

Analysis of the intersections between the frequency of use, perceived importance, and ease of use. We analyzed the occurrence of each privacy strategy across four categories: (1) *More use, most important, and easiest to use*; (2) *Least used, least important, and hardest to use*; (3) *Most used, least important, and hardest to use*; and (4) *Least used, most important, and easiest to use*. This analysis was based on the Likert scale, where 5 indicates *more*, *most*, and *easiest*, while 1 indicates *least* and *hardest*. Figure 6 shows the frequency of privacy strategies within these four categories.

By analyzing Figure 6, we can observe that each data privacy strategy appears at least once in the category (1): *More use, most important, and easiest to use*. More specifically, the *Design and Code Reviews* appears five times, followed by the *Temporal Data*, *Data Sovereignty*, *Proxy* and *Minimization of Personal Data Collection* strategies, each with four occurrences, respectively. Furthermore, we also can observe that in the category (4): Least used, most important,

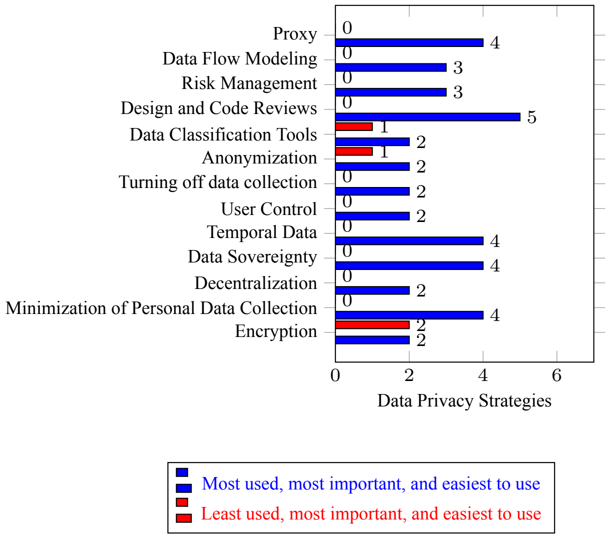

Figure 6. The distribution of data privacy strategies across categories

and easiest to use the *Encryption* strategy appears two times, followed by *Anonymization* and *Data Classification Tools*, with 1 occurrence, respectively. Finally, no data privacy strategy appears simultaneously in the category *Least used, least important, and hardest to use*. This observation also applies to the third category, i.e., *Most used, least important, and hardest to use*. Regarding the strategies that participants considered important, we asked them to explain the reasons behind their ratings (Q28). The respondents stated that all the strategies presented in the survey are important.

> #P30 said: *"All the options provide an additional level of privacy for user data. The use of each strategy depends on the scenario in which it will be applied. For example, encryption can be used for user passwords, while data control is more related to both permissions and system privacy, ensuring users prevent unauthorized access. Code review is essential to ensure that nothing that could affect data privacy is introduced into the system's production."*

Finding 7: No data privacy strategy is simultaneously classified as *Least used, least important, and hardest to use*. Similarly, none appear in the category *Most used, least important, and hardest to use*. On the other hand, all data privacy strategies appear at least two times in the category *Most used, most important, and easiest to use*.

4.3 Adherence to Data Privacy Practices within Software Organizations (RQ3)

To address RQ$_3$, we analyzed organizational factors that influence the adoption of data privacy practices, using responses from Q30, Q32, and Q33. Additionally, we analyzed the responses from two open-ended questions: (i) Q31, which explores work dynamics and organizational practices for handling privacy and data protection, and (ii) Q34, which captures Brazilian developers' perceptions of organizational challenges in protecting users' privacy rights. The analysis procedures are detailed in Section 3.5.

Influence of organizational factors on the adoption of privacy strategies. To understand how organizational factors influence the adoption of privacy strategies within development teams, we examined three key aspects (see Table 7): *(i)* the presence of dedicated privacy teams (Q30), *(ii)* the use of specific tools for managing private data (Q32), and *(iii)* the perceived priority given by organizations to privacy and data protection (Q33). Our analysis, based on 88 responses, provides insights into organizational structures and their impact on privacy practices.

Organizational structures and privacy management. The presence of dedicated privacy teams emerged as a significant factor in the adoption of privacy strategies. Among the participants surveyed, 43 reported having an internal team solely focused on privacy, while 39 addressed privacy concerns on an ad hoc basis. Additionally, 4 participants employed outsourced privacy teams, while others mentioned hybrid approaches or were unsure about their organization's structure.

Tool adoption and usage patterns. Organizations with dedicated privacy teams were more likely to use specific tools for managing privacy, with OneTrust being the most frequently mentioned (11.63%), followed by solutions like BigID and custom-built systems (each at 2.33%). Conversely, 89.74% of organizations without dedicated teams reported not using any specific privacy tools. This shows a clear correlation between structured organizational support and the adoption of technological solutions for privacy.

The number of tools employed by organizations further illustrates the influence of structural factors. Most respondents (78) reported using a single tool, while a smaller number used two (6 responses), three (4 responses), or as many as five tools (1 response). This limited adoption of multiple tools suggests that budget constraints, implementation complexity, or a lack of prioritization may hinder technological integration, even in organizations that value privacy.

Perceived organizational priority. The perceived priority assigned to privacy also varied significantly based on organizational structure: (i) *Organizations with dedicated teams* - Respondents predominantly viewed privacy as either essential (41.86%) or of high priority (41.86%), with only 16.28% considering it a medium priority; (ii) *Ad hoc privacy management* - The perception of priority was lower, with only 17.95% rating privacy as essential, while 35.90% classified it as medium priority and 20.51% as low priority; (iii) *Outsourced privacy teams* - Half of the respondents rated privacy as essential, and the other half considered it a high priority; and (iv) *Uncertain structures* - Respondents who were unaware of their organization's privacy management structure considered privacy a medium priority.

Finding 8: The absence of dedicated privacy teams correlates with a lower perceived priority and less investment in tools. Even in organizations that recognize the importance of privacy, the limited adoption of specific technologies reflects persistent challenges, such as budget limitations, implementation difficulties, or insufficient prioritization.

These insights emphasize the need for a structured ap-

Table 7. Organizational Structure and Tools Used

Dedicated Team (Q30)	**#responses**	**Most Used Tools (Q32)**	**Organizational Priority (Q33)**
Yes, dedicated internal team	43	None (58.14%), OneTrust (11.63%), Others (30.23%)	Essential (41.86%), High (41.86%), Medium (16.28%)
No, we handle it ad-hoc	39	None (89.74%), OneTrust, Ketch, Vanta (2.58%), BigID, Ketch, Vanta (2.56%), Securiti, Vanta (2.56%), OneTrust, Securiti, Ketch (2.56%)	Essential (17.95%), Medium (35.90%), Low (20.51%), High (15.38%), Not a priority (10.26%)
Outsourced team	4	None (75%), OneTrust (25%)	Essential (50%), High (50%)
Uncertain	2	Be Compliance (100%)	Medium (100%)

proach to privacy management, which requires specialized teams, tools, and organizational commitment to prioritize privacy in software development. Addressing these gaps will enhance privacy strategies and better protect user data in the evolving digital landscape.

Indeed, from finding 8, the company culture plays a crucial role in shaping how privacy decisions are made. Several participants mentioned that their ability to implement privacy measures was influenced, either supported or limited, by management priorities, available resources, and the overall organizational stance on privacy. In practice, this means developers can be either empowered or constrained by decisions made at higher levels. For instance, if privacy isn't treated as a priority by leadership, it becomes much harder for developers to justify the time and effort needed to build in strong protections. On the other hand, when privacy is supported from the top down, with clear policies, training, and dedicated resources, developers are better positioned to make informed decisions and integrate privacy effectively into their work. Although our study did not focus specifically on organizational culture, these aspects emerged in qualitative responses, suggesting that developers are often constrained by decisions made at higher levels.

The work dynamics and organization practices for handling privacy and data protection. To better understand how organizations address data privacy issues, we asked participants to describe the specific work dynamics and practices their organization employs to manage privacy and data protection (Q31 from Table 2). A total of 88 participants responded to this question, and we identified nine categories. The most commonly used work dynamic by organizations to handle privacy and data protection is the use of `laws`, mentioned by 24 participants (Table 8). Fifteen participants mentioned that there is a `compliance team` in the organization, and 14 of them stated that the `development team` is responsible for ensuring the privacy and protection of user data. The use of `techniques` was mentioned by eight participants, with anonymization and security being the most commonly used. Additionally, six participants stated that organizations conduct `training` to educate and raise awareness among employees about handling user privacy and data protection. Table 8 presents all identified categories and subcategories from the coding of Q31 responses.

Our findings are similar to the work conducted by Canedo *et al.* [2021], which conducted a survey with 82 practitioners from software development companies across various organizations, as well as a focus group with 11 participants, to understand how LGPD principles are implemented by software development teams. Most participants stated that data privacy should be enforced by development teams using techniques such as data encryption and anonymization.

Franke *et al.* [2024] conducted a study with 56 open-source software (OSS) developers to understand the impact of General Data Protection Regulation (GDPR) implementations on development activities. Most participants acknowledged that the implementation of GDPR concepts affects development processes, highlighting the influence of privacy/compliance requirements on open-source software development. Fifteen OSS developers reported consulting legal teams to ensure GDPR compliance, and seven participants with experience consulting compliance teams observed a positive impact on their software development activities. The participants emphasized the benefits of seeking legal expertise, affirming the importance of consulting compliance teams to clarify privacy requirements and avoid misinterpretations of the legislation. This consultation facilitated easier implementation of GDPR compliance. Our findings also reinforce those of Franke *et al.* [2024], as most participants in our study reported that their work dynamics and practices for managing privacy and data protection within their organizations include consulting compliance teams.

Finding 9: Key practices for managing privacy and data protection include adhering to laws such as LGPD, engaging compliance and development teams, and implementing techniques like anonymization and security measures. Additionally, training programs are essential for enhancing employee awareness and promoting effective privacy management.

These findings reinforce the importance of integrating legal expertise and compliance consultation into software development to address privacy challenges effectively.

Organizational challenges in protecting users' privacy rights. We also investigated participants' perceptions of the biggest challenges organizations will face in the coming years regarding practices and regulations to better protect user data privacy rights (Q34 from Table 2). A total of 87 participants responded to this question, and eight categories were identified. The most frequently mentioned categories were `Privacy Laws` and `Generative AI`. Fifty-three participants stated that the biggest challenge will be the implementation of privacy laws, while 16 mentioned dealing with generative AIs. Table 9 presents the eight categories and their respective subcategories.

Table 8. Work dynamics for handling privacy and data protection

Category	Subcategory	#
1. Laws	Analysis of the collected data	14
	Legislation, consent, risk management, and anonymization	1
	Data governance system under rules and guidelines	1
	Lack of practical experience in privacy	1
	Compliance analysis with LGPD in the development process	1
	Data minimization and user identity protection	1
	Use of a unified database	1
	Data access limitation	1
	Apply LGPD guidelines and encryption	1
	Application of Privacy by Design	1
	Use of privacy policies	1
2. Compliance Team	There is a compliance team	15
3. Development Team	Discussion and collective decision	5
	Application of the LGPD from the requirements elicitation phase	4
	Integration of privacy into the requirements definition process	1
	Developer's responsibility for the data lifecycle	1
	During software development	1
	Guidance from senior members	1
	Peer review	1
6. Techniques	Use of anonymization techniques	2
	Implementation of security requirements	1
	Management of sensitive data and anonymization	1
	Phishing tests	1
	Prevention techniques	1
	Uses encryption	1
	Access control, user profiles, and data management	1
5. Training	Training and awareness	2
	Customer awareness	1
	Conducting seminars and training sessions	1
	Conducting courses	1
	Awareness programs	1
6. Work Dynamics	Lack of work dynamics	1
7. Techniques and Training	Encryption; Lack of training	1
8. Compliance Team and Training	There is a compliance team; Lack of training	1
9. Legislation; Training; Awareness; and Techniques	Governance and Privacy Policies; Employee Training and Awareness; Access Controls and Technological Security; Continuous Auditing and Monitoring; Incident Management and Rapid Response; Preventive Impact Assessment	1

Table 9. Future challenges for organizations in protecting user privacy rights

Category	Subcategory	#
1. Privacy laws	Implementing effective privacy policies	41
	Country-specific laws	4
	Adaptation of legacy systems	2
	Implementing secure auditing and data collection	1
	Lack of knowledge	1
	Lack of tools to support implementation	1
	Cybersecurity and transparency	1
	Security	1
	Lack of tools to optimize processes	1
2. Generative AI	Dealing with generative AIs	15
	Bias in training data	1
3. Privacy laws; Generative AI	Implementing effective privacy policies. Dealing with generative AIs	4
4. Privacy laws; Generative AI; Awareness	Implementing effective privacy policies. Dealing with generative AIs; Lack of user awareness	1
5. Awareness	Lack of user awareness	2
	Awareness of users	1
6. Engagement	Engage people	3
7. Training	Lack of specialized privacy teams	3
8. Cyberattacks	Cyberattacks	1

Rocha *et al*. [2023] also identified, through a survey with several practitioners, that all participants reported difficulties in implementing the principles of the LGPD due to a lack of knowledge about implementation techniques. They emphasized that software development practitioners need to enhance their understanding of techniques that ensure privacy and compliance with the LGPD principles in order to effectively implement privacy policies. Jesus *et al*. [2024] and Peixoto *et al*. [2023] also identify that practitioners face difficulties in effectively implementing data privacy law principles, either due to a lack of knowledge or the absence of automated tools to support the implementation of privacy requirements during the software development process. Our findings are also similar to those of [Golda *et al*., 2024; Cheung and Liu, 2023; Ferrara, 2023], who identified challenges related to ensuring user data privacy and biases in training data when using generative AI.

Finding 10: Organizations face significant challenges in protecting user privacy, implementing effective policies, adapting legacy systems, and complying with diverse laws. Emerging issues include managing generative AI and addressing biases, raising user awareness, engaging stakeholders, and building specialized privacy teams.

Based on these findings, we observed that addressing these challenges requires a multifaceted approach from organizations. Several concrete steps can help bridge the gap in privacy awareness. One key step is offering regular training that's tailored to specific roles and real-world situations so developers can really understand how privacy laws apply to their daily work. It is also important to incorporate privacy into development processes and tools, rather than treating it as an afterthought. Beyond that, building a company culture where privacy is taken seriously, starting with leadership, can make a big difference. Things like clear internal policies, working closely with legal and compliance teams, and making sure privacy is part of project planning from the start all help developers make better, more informed choices.

Additionally, the Brazilian organizations can adopt some helpful resources that could support Brazilian developers in strengthening their privacy efforts. For instance, adopting Privacy by Design principles is a great starting point. There are also well-established standards like ISO/IEC 27701, which focuses on privacy information management, and practical guidance from Brazil's own National Data Protection Authority (ANPD). On top of that, the NIST Privacy Framework can offer clear, actionable steps to help teams incorporate privacy into their development workflows.

5 Threats to Validity

Our study involved a survey with 88 developers, including only those participants who reported direct or indirect work experience related to data privacy. To ensure the validity of our findings, we discuss threats to the study validity [Wohlin *et al.*, 2012] as follows. First, regarding **construct validity**, there was a risk that the survey questions might not fully align with the intended constructs, such as developers' understanding and experience with data privacy. To mitigate this, we designed the survey questions based on established literature and validated them through a pilot study with four experts in the field. This process helped ensure that the questions were relevant and effectively captured the constructs of interest. The 21 statements created to answer RQ1 using the Knowledge-Attitude-Behavior (KAB) model were carefully crafted by two paper authors. In addition, the statements were based on findings from previous studies [Iwaya *et al.*, 2023; Tahaei *et al.*, 2021; Hadar *et al.*, 2018].

Second, for **internal validity**, participants' interpretations of the survey questions could have been influenced by their roles, organizational contexts, or prior experiences, potentially introducing bias. To address this, we provided clear definitions and illustrative examples of key terms within the survey instrument to minimize ambiguity and reduce the likelihood of misinterpretation. We followed strict procedures for creating the different groups used in RQ1. We collected their background and divided all survey respondents into different demographic groups. Third, **external validity** posed a challenge, as the generalizability of the findings might be limited due to the self-selected nature of the participants and the focus on individuals with data privacy experience. To enhance representativeness, we targeted developers from diverse sectors and professional roles, within the Brazilian software industry, ensuring a broad range of perspectives on data privacy practices. In addition, we used a snowball sampling strategy to increase participant reach. While effective for accessing participants with relevant experience, this approach may introduce sampling bias, as individuals tend to refer peers with similar backgrounds or perspectives. To help mitigate this, we included control questions related to the backgrounds of the participants (e.g., level of experience, company size, and region), which allowed us to examine possible biases in the sample. Nonetheless, we believe the adopted sampling strategy is appropriate to the Brazilian context, where the registries of developers with privacy experience are limited.

Finally, regarding **conclusion validity**, inconsistencies in participant responses due to misunderstanding or unclear survey questions could have impacted the reliability of our findings. To mitigate this threat, we pre-tested the survey and implemented iterative refinements to improve clarity and consistency throughout the instrument. For the descriptive and statistical analysis, two authors collaborated on analyzing all research questions (RQs). Specifically, for RQ1, we evaluated the data distribution before applying correlation tests to mitigate potential biases in the statistical analysis. We employed the Wilcoxon Rank Sum Test [Whitley and Ball, 2002] and Cliff's Delta (d) [Grissom and Kim, 2005] to assess the level of agreement with each statement between the predefined groups. Additionally, we used the conventional p-value threshold of 0.05 to establish statistical significance.

The data analysis process described in Section 3.4 examines participants' perceptions of the frequency, perceived importance, and ease of use of data privacy strategies, considering their individual characteristics. This analysis was based on a previous survey [Dias-Neto *et al.*, 2017] and employs a weighted average approach, where the weighting is informed by qualitative attributes of each participant. While this method provides a nuanced view, it may result in the loss of extreme data points. Regarding the qualitative data analysis, we have partially relied on Grounded Theory (GT) [Corbin and Strauss, 2008] to define our data analysis protocol. We aimed to reduce the inherent subjectivity of coding the responses to the open-ended questions. We analyzed all data in a pair to minimize biases and reach a consensus about the identified categories and subcategories.

6 Related Work

Data Privacy, whose importance has been growing in the context of software development, has been the focus of recent studies investigating the challenges related to organizational factors and developers' perceptions in implementing data privacy strategies. Hadar *et al.* [2018]; Tahaei *et al.* [2021]; Iwaya *et al.* [2023] conducted interviews with developers to understand the challenges related to ensuring data privacy. These studies provide detailed insights into privacy strategies and the challenges developers face. However, gaps remain in the quantitative measurement of these organizational factors and developers' perceptions. To address these gaps, this study combines qualitative and quantitative data collected through a structured survey, aiming for a broader understanding of the organizational factors that may influence development teams' adherence to privacy practices.

Hadar *et al.* [2018] investigated software developers' perceptions and mindsets regarding privacy by design. The authors conducted interviews with software developers to identify and analyze their perceptions and practices related to privacy. Our study differs from this work as we conducted a survey with both open and closed-ended questions and performed a quantitative and qualitative analysis, allowing us to reach a larger number of respondents. This approach actively contributes to understanding how developers' mindsets and perceptions influence the adoption of data privacy strategies in the systems they develop. The study by Hadar *et al.* [2018] is relevant to our work as it explores developers' perceptions of data privacy, providing a starting point for analyzing how organizational factors and technical knowledge can influence the application of privacy strategies in software systems.

Canedo *et al.* [2023] conducted a systematic literature review to identify the methodologies, techniques, and tools used for eliciting privacy requirements. Additionally, they conducted a survey with ICT practitioners to investigate their perceptions regarding the use of these techniques and tools in the software development process. The study revealed that the existing methodologies, techniques, and tools for privacy requirements elicitation are not commonly used by practitioners in the industry. However, practitioners acknowledged that their use could play an important role in ensuring user data privacy. This study differs from ours, as the authors focused on investigating whether the techniques identified in the literature were being used in the industry, rather than examining how organizational factors might influence adherence to these techniques and methods.

Peixoto *et al.* [2023] investigated the level of knowledge and understanding software developers have about privacy, exploring personal, behavioral, and external environmental factors that influence their decision-making regarding privacy requirements. They conducted semi-structured interviews with thirteen professionals from six different companies. The study identified nine personal factors, five behavioral factors, and seven external environmental factors that positively or negatively impact developers' decision-making concerning privacy. This work differs from our research, as it focuses solely on factors related to the Requirements Engineering phase, whereas our study examines factors spanning the entire software development lifecycle.

Tahaei *et al.* [2021] analyzed the role of privacy champions within software development teams. These practitioners are responsible for advocating for user data privacy within organizations by implementing privacy strategies and addressing challenges associated with the lack of data privacy. The authors conducted interviews with privacy champions to investigate their motivations and the strategies they employ in the context of sensitive data privacy in organizations. The findings highlight that privacy champions play a important role in raising awareness and promoting the adoption of effective privacy strategies, despite facing barriers such as organizational cultural resistance and resource constraints. The authors emphasized that having dedicated privacy practitioners can strengthen best practices in software development by aligning them with the principles of Privacy by Design. This study is directly related to our research, as we aim to identify the challenges and opportunities in adopting privacy strategies among developers.

Some studies have also examined how organizational aspects influence the adoption of privacy strategies in software development. Iwaya *et al.* [2023] investigated software developers' perceptions of privacy, focusing on organizational aspects. The authors conducted interviews with developers and identified three key dimensions: developers' knowledge, attitudes, and behaviors. Additionally, the study explored how these factors influence the adoption of privacy practices, as well as the impact of team leadership and organizational culture on their implementation. One of the main findings was that negative organizational cultures—characterized by a low prioritization of privacy and a lack of incentives for training—directly hinder the adoption of privacy strategies. Despite the growing awareness among team members about data privacy in recent years, the topic remains under-discussed, as there is still a lack of knowledge about privacy engineering strategies and standards. The authors also highlighted the importance of organizational culture and the role of management in promoting the use of privacy strategies.

The findings of Iwaya *et al.* [2023] correlate with those of Hadar *et al.* [2018], which also emphasized the importance of aligning organizational culture with the practices adopted by developers to ensure data privacy. Both studies highlight the need for more robust organizational initiatives to enhance the privacy culture, aligning with the goal of our study to investigate how organizational factors influence the adoption of privacy strategies. Our study measures this through the frequency, importance, and use of privacy strategies by software development teams.

7 Final Remarks

This paper presents a study investigating Brazilian software developers' awareness of data privacy. To this end, we conducted a survey with 88 developers to explore: (i) their awareness of data privacy using the Knowledge-Attitude-Behaviour (KAB) model; (ii) their perception of frequency, importance, and ease of use of 13 data privacy strategies; and (iii) organizational factors that might influence the adoption and adherence of data privacy practices within development teams. Our findings provide empirical evidence from Brazilian developers highlighting their awareness of data privacy, perceptions of data privacy strategies, and organizational factors. These findings can guide researchers and developers aiming to understand the challenges and benefits of incorporating data privacy into software projects in practice.

Our survey revealed that the majority of the Brazilian software developers surveyed exhibited a moderate level of knowledge and awareness regarding data privacy. While many developers were aware of the general importance of data privacy, their specific knowledge about legal regulations, such as the *General Data Protection Regulation* (GDPR), was limited. Moreover, developers generally rated data privacy strategies like *Encryption*, and *User control* as both important and easy to use. However, strategies such as *Anonymization* and *Turning off data collection* were perceived as less frequently used and somewhat more complex to implement, reflecting a gap between their recognized importance and practical adoption.

Furthermore, we identified that organizational factors such as the company's size, industry, and the presence of a dedicated privacy officer or team tend to influence the adoption of data privacy practices. Larger organizations and those in highly regulated sectors demonstrated a higher level of adherence to data privacy strategies. In contrast, smaller organizations often lack the resources or expertise to implement comprehensive privacy practices fully.

In conclusion, the study highlights the need for enhanced training and awareness among software developers, particularly in terms of legal requirements and more advanced privacy measures. Additionally, organizations must invest in clear privacy policies and allocate sufficient resources to ensure data privacy is effectively integrated into the software development lifecycle. As future work, we plan to conduct a longitudinal study to assess changes in privacy awareness and practices, especially in response to new regulations, and effectiveness of different approaches in bridging the privacy awareness gap. Additionally, we plan to explore how regional differences within Brazil and organizational characteristics, such as culture, company size, may influence how privacy is prioritized and supported within organizations.

Declarations

Funding

This work is partially supported by FUNCAP (BP6-00241-00276.01.00/25), PET-UFC, CNPq (404406/2023-8), CAPES (88881.879016/2023-01) and FAPESP (2023/00811-0).

Authors' Contributions

Aryely Matos: Conceptualization, Data curation, Investigation, Methodology, Validation, Writing (original draft, review, and editing). Mario Patrício: Investigation, Data curation, Visualization, Writing (original draft, review and editing). Maria Isabel Nicolau: Investigation, Data curation, Visualization, Writing (review and editing). Edna Dias Canedo: Investigation, Methodology, Validation, Visualization, Writing (original draft, review, and editing). Juliana Alves Pereira: Conceptualization, Methodology, Co-supervision, Validation, Writing (review and editing). Anderson Uchôa: Conceptualization, Data curation, Investigation, Writing (original draft, review, and editing), Supervision, Project administration, Funding acquisition. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The main artifacts involved in our study are publicly available at `https://doi.org/10.5281/zenodo.14345392`, including the survey script, statistical analysis, and the complete codebook for the open-ended questions.

References

Ashcraft, C., McLain, B., and Eger, E. (2016). *Women in tech: The facts*. National Center for Women & Technology (NCWIT) Colorado, CO, USA. Available at:`https://www.ceoplaybook.co/wp-content/uploads/2019/11/womenintech_facts_fullreport_05132016-1.pdf`.

Brasil (2018). Lei n.º 13.709, de 14 de agosto de 2018. Available at:`https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm`,.

Caldiera, V. R. B.-G. and Rombach, H. D. (1994). Goal question metric paradigm. *Encyclopedia of software engineering*, 1(528-532):6. Available at:`https://www.cs.umd.edu/~mvz/handouts/gqm.pdf`.

Canedo, E. D., Bandeira, I. N., Calazans, A. T. S., Costa, P. H. T., Cançado, E. C. R., and Bonifácio, R. (2023). Privacy requirements elicitation: a systematic literature review and perception analysis of IT practitioners. *Requir. Eng.*, 28(2):177–194. DOI: 10.1007/S00766-022-00382-8.

Canedo, E. D., Calazans, A. T. S., Cerqueira, A. J., Costa, P. H. T., and Masson, E. T. S. (2021). Agile teams' perception in privacy requirements elicitation: Lgpd's compliance in brazil. In *29th IEEE International Requirements Engineering Conference, RE 2021, Notre Dame, IN, USA, September 20-24, 2021*, pages 58–69. IEEE. DOI: 10.1109/RE51729.2021.00013.

Cheung, M. Y. M. and Liu, H. (2023). Information privacy concerns in generative AI. In *Australasian Conference on Information Systems, ACIS 2023, Wellington, New Zealand, December 5-8, 2023*. Available at:`https://aisel.aisnet.org/acis2023/24`,.

Corbin, J. and Strauss, A. (2008). Basics of qualitative research: Techniques and procedures for developing grounded theory. *Thousand Oaks*, 3:1–400. Available at:`https://archive.org/details/basicsofqualitat0000stra/page/n3/mode/2up`,.

Dias-Neto, A. C., Matalonga, S., Solari, M., Robiolo, G., and Travassos, G. H. (2017). Toward the characterization of software testing practices in south america: looking at brazil and uruguay. *Software Quality Journal*, 25:1145–1183. DOI: 10.1007/s11219-016-9329-3.

Ferrão, S. É. R., Silva, G. R. S., Canedo, E. D., and Mendes, F. F. (2024). Towards a taxonomy of privacy requirements based on the LGPD and ISO/IEC 29100. *Inf. Softw. Technol.*, 168:107396. DOI: 10.1016/J.INFSOF.2024.107396.

Ferrara, E. (2023). Should chatgpt be biased? challenges and risks of bias in large language models. *First Monday*, 28(11):13346/11369. DOI: 10.5210/FM.V28I11.13346.

Franke, L., Liang, H., Farzanehpour, S., Brantly, A., Davis, J. C., and Brown, C. (2024). An exploratory mixed-methods study on general data protection regulation (GDPR) compliance in open-source software. In *Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM 2024, Barcelona, Spain, October 24-25, 2024*, pages 325–336. ACM. DOI: 10.1145/3674805.3686692.

Glaser, B. and Strauss, A. (2017). Discovery of grounded theory: Strategies for qualitative research. DOI: 10.4324/9780203793206.

Golda, A., Mekonen, K., Pandey, A., Singh, A., Hassija, V., Chamola, V., and Sikdar, B. (2024). Privacy and security concerns in generative AI: A comprehensive survey. *IEEE Access*, 12:48126–48144. DOI: 10.1109/ACCESS.2024.3381611.

Grissom, R. J. and Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Lawrence Erlbaum Associates Publishers. DOI: 10.4324/9781410612915.

Hadar, I., Hasson, T., Ayalon, O., Toch, E., Birnhack, M., Sherman, S., and Balissa, A. (2018). Privacy by designers: software developers' privacy mindset. *Empirical Software Engineering*, 23:259–289. DOI: 10.1007/s10664-017-9517-1.

Iwaya, L. H., Babar, M. A., and Rashid, A. (2023). Privacy engineering in the wild: Understanding the practitioners' mindset, organizational aspects, and current practices. *IEEE Transactions on Software Engineering*, 49(9):4324–4348. DOI: 10.1109/TSE.2023.3290237.

Jesus, E. D. B. D., Vilela, J., and Silva, C. (2024). Requisitos de segurança e privacidade em startups: Um estudo empírico em uma aplicação de governança de dados. In *Anais do WER24 - Workshop em Engenharia de Requisitos, Buenos Aires, Argentina, August 7-9, 2024*. Even3, Brasil. DOI: 10.29327/1407529.27-13.

Kempe, E. and Massey, A. (2021). Regulatory and security standard compliance throughout the software development lifecycle. In *54th Hawaii International Conference on System Sciences, HICSS 2021, Kauai, Hawaii, USA, January 5, 2021*, pages 1–10. ScholarSpace. Available at:`https://scholarspace.manoa.hawaii.edu/items/5fd99dfc-2570-4605-97ff-09bd61a99fcf`.

Kitchenham, B. and Pfleeger, S. L. (2002). Principles of survey research: part 5: populations and samples. *ACM SIGSOFT Software Engineering Notes*, 27(5):17–20. DOI: 10.1145/571681.571686.

Kruger, H. A. and Kearney, W. D. (2006). A prototype for assessing information security awareness. *Computers & security*, 25(4):289–296. DOI: 10.1016/j.cose.2006.02.008.

Kshetri, N. (2024). Navigating EU regulations: Challenges for U.S. technology firms and the rise of europe's generative AI ecosystem. *Computer*, 57(10):112–117. DOI: 10.1109/MC.2024.3433088.

Landis, C. B. and Kroll, J. A. (2024). Mitigating inference risks with the NIST privacy framework. *Proc. Priv. Enhancing Technol.*, 2024(1):217–231. DOI: 10.56553/POPETS-2024-0013.

Lester, C. Y. and Jamerson, F. (2009). Incorporating software security into an undergraduate software engineering course. In *The Third International Conference on Emerging Security Information, Systems and Technologies, SECURWARE 2009, 18-23 June 2009, Athens/Glyfada, Greece*, pages 161–166. IEEE Computer Society. DOI: 10.1109/SECURWARE.2009.32.

Likert, R. (1932). *A Technique for the Measurement of Attitudes*. Number Nº 136-165 in A Technique for the Measurement of Attitudes. Archives of Psychology. Available at:`https://archive.org/details/likert-1932`.

Linåker, J., Sulaman, S. M., de Mello, R. M., and Höst, M. (2015). Guidelines for conducting surveys in software engineering. *Technical report*. Available at:`https://lup.lub.lu.se/search/files/6062997/5463412.pdf`.

Litwin, M. S. and Fink, A. (1995). *How to measure survey reliability and validity*, volume 7. Sage, https://methods.sagepub.com/book/how-to-measure-survey-reliability-and-validity. DOI: 10.4135/9781483348957.

Parsons, K., McCormac, A., Butavicius, M. A., Pattinson, M. R., and Jerram, C. (2014). Determining employee awareness using the human aspects of information security questionnaire (HAIS-Q). *Comput. Secur.*, 42:165–176. DOI: 10.1016/J.COSE.2013.12.003.

Peixoto, M. M., Ferreira, D., Cavalcanti, M., Silva, C., Vilela, J., Araújo, J., and Gorschek, T. (2023). The perspective of brazilian software developers on data privacy. *J. Syst. Softw.*, 195:111523. DOI: 10.1016/J.JSS.2022.111523.

Ralph, P., Baltes, S., Adisaputri, G., Torkar, R., Kovalenko, V., Kalinowski, M., Novielli, N., Yoo, S., Devroey, X., Tan, X., *et al.* (2020). Pandemic programming: How covid-19 affects software developers and how their organizations can help. *Empirical software engineering*, 25:4927–4961. DOI: 10.1007/s10664-020-09875-y.

Rocha, L. D., Silva, G. R. S., and Canedo, E. D. (2023). Privacy compliance in software development: A guide to implementing the LGPD principles. In *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, SAC 2023, Tallinn, Estonia, March 27-31, 2023*, pages 1352–1361. ACM. DOI: 10.1145/3555776.3577615.

Romano, J., Kromrey, J. D., Coraggio, J., Skowronek, J., and Devine, L. (2006). Exploring methods for evaluating group differences on the nsse and other surveys: Are the t-test and cohen'sd indices the most appropriate choices. In *annual meeting of the Southern Association for Institutional Research*, pages 1–51. Citeseer.

Salkind, N. (2012). *Exploring Research*. Pearson Education. Available at:`https://archive.org/details/exploringresearc0000salk_z7v0`.

Sangaroonsilp, P., Dam, H. K., Choetkiertikul, M., Ragkhitwetsagul, C., and Ghose, A. (2023). A taxonomy for mining and classifying privacy requirements in issue reports. *Inf. Softw. Technol.*, 157:107162. DOI: 10.1016/J.INFSOF.2023.107162.

Schrader, P. G. and Lawless, K. A. (2004). The knowledge, attitudes, & behaviors approach how to evaluate performance and learning in complex environments. *Performance Improvement*, 43(9):8–15. DOI: 10.1002/pfi.4140430905.

Shapiro, S. and Wilk, M. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611. DOI: 10.2307/2333709.

Stallings, W. (2019). *Engenharia de privacidade de informações e privacidade por design: Compreendendo ameaças à privacidade, tecnologia e regulamentações com base em padrões e melhores práticas*. Addison-Wesley Professional. Book.

Tahaei, M., Frik, A., and Vaniea, K. (2021). Privacy champions in software teams: Understanding their motivations, strategies, and challenges. In *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, pages 693:1–693:15. ACM. DOI: 10.1145/3411764.3445768.

Thomson, M. E. and von Solms, R. (1998). Information security awareness: educating your users effectively. *Inf. Manag. Comput. Secur.*, 6(4):167–173. DOI: 10.1108/09685229810227649.

União Europeia (2016). Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016 relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados). Jornal Oficial da União Europeia, L 119, 1-88.

Whitley, E. and Ball, J. (2002). Statistics review 6: Nonparametric methods. *Critical care*, 6:1–5. DOI: 10.1186/cc1820.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in Software Engineering*. Springer. DOI: 10.1007/978-3-642-29044-2.