Understanding Legal Professionals' Practices and Expectations in Data Breach Incident Reporting

Ece Gumusel*
Indiana University Bloomington
Bloomington, Indiana, USA
egumusel@iu.edu

Yue Xiao*
Indiana University Bloomington
& IBM Research
Yorktown Heights, New York, USA
xiaoyue@ibm.com

Yue Qin
Indiana University Bloomington
Bloomington, Indiana, USA
qinyue@iu.edu

Jiaxin Qin
China University of Political Science and Law
Beijing, Beijing, CHINA
220301276@cupl.edu.cn

Xiaojing Liao
Indiana University Bloomington
Bloomington, Indiana, USA
xliao@indiana.edu

ABSTRACT

Legal professionals are essential in analyzing data breach incident reports and guiding the response to comply with data privacy laws and regulations. Their expertise helps mitigate privacy and security risks and prevents failures in privacy compliance. However, little research has been done to understand how legal professionals perceive, react to, and face challenges within the data breach incident reporting procedure. In this study, we conducted a simulated incident report assessment experiment and semi-structured interviews with 33 legal professionals who varied in age, gender, and legal background. We reported the criteria used by legal professionals to identify privacy-related items and also uncovered that the agreement among legal professionals on the concepts of privacy-related items is low. Furthermore, we presented findings regarding the perceptions and strategies of legal professionals concerning legal and regulatory compliance, as well as the key features of incident responses that facilitate efficient analysis of data privacy and security law compliance. After taking into account the challenges and suggestions provided by legal professionals, we concluded this study with recommendations for enhancing the effectiveness of legal compliance analysis for incident responses.

CCS CONCEPTS

• **Security and privacy → Human and societal aspects of security and privacy**; **Social aspects of security and privacy**.

KEYWORDS

GDPR Compliance; Data Breach Incident Reporting; Legal Professionals' Practices; Security and Privacy in Legal Contexts;

ACM Reference Format:
Ece Gumusel*, Yue Xiao*, Yue Qin, Jiaxin Qin, and Xiaojing Liao. 2024. Understanding Legal Professionals' Practices and Expectations in Data Breach Incident Reporting . In *Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security (CCS '24), October 14–18, 2024, Salt Lake City, UT, USA*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3658644.3690357

1 INTRODUCTION

In recent years, new developments in data privacy laws and regulations have been rising, most notably the European Union's (EU) General Data Protection Regulation (GDPR) [37], which has set a global benchmark for data protection. Meanwhile, these laws and regulations introduced a new urgency to the demonstration of privacy compliance. According to the GDPR Enforcement Tracker [5], the total amount of GDPR non-compliance fines was €4,500,688,064 across 2,060 cases from July 2018 to April 2024. To avoid costly fines, organizations must ensure that they are taking all necessary measures to comply with the GDPR. This involves conducting appropriate privacy impact assessments and legal compliance analysis in incident responses.

Privacy compliance in incident responses. In data breach incident response procedures, organizations must ensure that they comply with specific privacy laws and regulations, which requires the implementation of a standardized incident response procedure [10]. One of the key considerations in incident response is the involvement of legal professionals, who play a crucial role in analyzing the incident and guiding the response in accordance with relevant data privacy laws and regulations [9, 32, 34, 38, 47, 54, 60, 61]. Specifically, legal professionals perform legal compliance checks and assess privacy risks from a legal standpoint, including analyses of any relevant laws, regulations, and contractual obligations that may have been breached. While the research community has already pointed out that the importance of technical-legal interactions in privacy compliance should not be underestimated [22, 25, 32, 39, 43, 56, 66],

*The first two authors are ordered alphabetically. This work was completed while Yue Xiao was a student at Indiana University Bloomington.

CCS '24, October 14–18, 2024, Salt Lake City, UT, USA

ACM ISBN 979-8-4007-0636-3/24/10
https://doi.org/10.1145/3658644.3690357

there has been little research on the strategies and barriers encountered by legal professionals in privacy risk assessment and legal compliance analysis for incident responses.

Research Questions. In this study, we investigate the perspectives, practices, and barriers faced by legal professionals specifically regarding GDPR compliance analysis in data breach incident reporting. We examined the following unaddressed research questions in this area:

RQ1 *What are the criteria by which legal professionals can identify privacy-related items (e.g., personal information, Apple advertising ID) under data protection laws and regulations from incident responses?*

RQ2 *How do legal professionals strategize to map privacy risks in incident responses to ensure compliance with the specific sections of relevant laws and regulations?*

RQ3 *How to improve the writing quality of incident responses for better understanding by legal professionals?*

Study and findings. In this paper, we conducted a simulated incident report assessment experiment with 33 legal professionals, along with semi-structured interviews. Our study revealed three main criteria that participants use to identify privacy-related items, including (1) personal information defined in law and regulations, (2) data associated with existing data breaches, and (3) elements of privacy-by-design/practice principles (**RQ1**). Notably, some participants (n=8) only replied on the literal definition of personally identifiable information (PII) as the only criteria to determine privacy-related items. They believed that device information (e.g., device serial numbers) should not be considered privacy-related items and therefore should not pose any privacy risks.

For **RQ2**, in our study, some of the participants (n=15) are comfortable conducting legal compliance analyses for well-documented incident responses with clearly defined privacy impacts. Conversely, our study found that legal professionals often struggle to agree on their analyses when the incident responses are less structured for privacy assessment, laden with technical jargon, and lack descriptions of law violations. Additionally, 27 participants emphasized the need for legal professionals to access additional resources or training to understand the technical aspects of these responses, identify privacy risks, and ensure compliance with relevant laws and regulations, especially with GDPR to mitigate significant impacts.

In terms of improving the quality of incident response writing (**RQ3**), 12 participants highlighted the importance of including discussions on privacy impacts in incident responses for more effective legal compliance analysis. Some participants (n=11) also expected incident responses to have well-organized structures. For instance, they emphasized the importance of background information that helps them understand the attack scenarios and cross-reference specific terminologies for clarity. Interestingly, the study participants showed varying preferences for incident responses. Some (n=19) preferred these responses to be generic, explaining regulatory compliance risks and summarizing the impact of the incident. However, others (n=23) expressed a need for additional technical background or details to fully grasp the implications of these responses. Notably, 9 participants emphasized the necessity for both generic explanations and additional technical details in the responses.

Contributions. Contributions of the paper are as follows:

- This study is the first to investigate legal professionals' perspectives, practices, and barriers in privacy risk assessment and GDPR compliance analysis during data breach incident response. We identify strategies used by legal professionals to identify privacy-related items in incident responses with related laws and regulations.
- This study reveals a low level of consensus among legal professionals regarding the concepts of privacy-related items.
- This study identifies several challenges in GDPR-compliant incident response, including discrepancies in data definitions across various sources, lack of technical and legal context, and varying levels of expertise.
- This study provides recommendations for enhancing the quality of incident responses to facilitate more effective legal compliance analysis.

2 BACKGROUND

2.1 Data Privacy Laws and Regulations

In recent years, numerous data privacy laws and regulations have been established worldwide to safeguard individuals from privacy violations and protect their personal information. In the EU, the GDPR is considered the gold standard for data protection laws. It sets strict standards for the collection and use of personal data and provides individuals with greater control over their personal information. In the United States (US), the data protection landscape is more fragmented, lacking standard federal data protection laws and regulations. Instead, regulations tend to be sectoral, focusing on specific areas such as health data. For instance, The Health Insurance Portability and Accountability Act of 1996 (HIPAA) governs the protection of health information, while The Family Educational Rights and Privacy Act (FERPA) safeguards student education records. Additionally, at the state level, there are various data protection laws in place, such as the California Consumer Protection Act (CCPA) further contributing to the complexity of the regulatory environment. The CCPA has served as a model for other states, including Colorado, Utah, Nebraska, and Virginia, to enact their own privacy laws and regulations. However, the absence of a standard federal data protection law in the US remains a challenge.

These laws and regulations bring in new challenges for regulatory and national standard compliance. For example, the EU's GDPR requires *appropriate technical measurements*[1] to prevent potential privacy and security challenges for organizations. This can raise various privacy and legal compliance challenges for both legal professionals and engineers [39, 53]. It can even cause insufficient compliance practices. The analysis of 261 publicly available GDPR enforcement orders issued by Data Protection Authorities during the first 24 months of the GDPR implementation shows that the largest volume of fines were issued mostly for *non-compliance with general data processing principles* and *insufficient legal basis for data processing* [59]. Indeed, GDPR is not the only measure for privacy compliance, yet it is a great example that shows how the companies (e.g., Equifax [17], Capital One [1], Meta [16], Google [51]) have been struggling with non-compliance with these regulations [5, 20, 50].

[1] [GDPR Article 25.2] The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed.

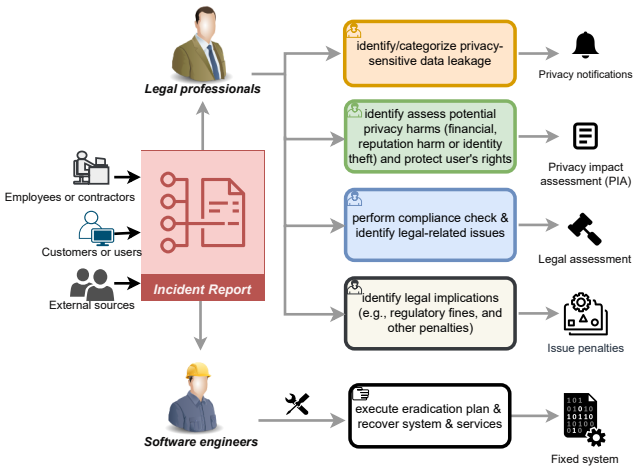

Figure 1: Legal professionals' duties in incident response

This study therefore primarily focuses on GDPR compliance, highlighting the critical intersection of legal and technical perspectives for effective implementation [39].

2.2 Privacy-related Items

Defining privacy-related items is a complex task that is challenged by multiple factors. One of the most prominent challenges is the obscure nature of data protection regulations [26]. These laws and regulations often have general and broad definitions that are difficult to interpret, making it challenging for organizations to determine which data should be considered private. The leading data protection law, GDPR Article 4 defines *personal data*, also referred to as PII or personal information, as "any information relating to an identified or identifiable natural person". According to the definition, any information that is related to or linked with an individual is considered as personal or private information.

However, there is still no universally correct answer to what information or item is identifiable of private, as this can be context-dependent based on each legal professional's interpretation of legal requirements. We therefore argue that the diversity of legal documents can pose a challenge, as sensitive data can be considered differently in different legal documents. For instance, data considered sensitive in privacy policies may differ from data defined by Terms of Service (ToS). Privacy policies typically focus on personal information, such as name, email address, and mailing address, guided by laws. ToS, on the other hand, focuses on security-critical data such as passwords and tokens, as well as Software Development Kit (SDK)-specific data such as Application Programming Interface (API) keys and access credentials, to prevent the abuse of sensitive data [57]. These factors present a significant challenge for organizations in defining personal data, which is crucial for developing effective privacy policies and data protection measures.

In our study, we conducted both quantitative and qualitative analyses to gain insight into the criteria used by legal professionals to identify personal data or privacy-related items within the GDPR's definition and to reveal the challenges and difficulties they encounter in real-world scenarios.

2.3 Legal Professionals in Incident Responses

To effectively manage and minimize the impact of privacy and security incidents, organizations usually follow a standardized incident response procedure. As shown in Figure 1, *legal professionals* play a primary role in ensuring that incident response is compliant with data protection laws and regulations and meets with the national incident response standards [39]. In literature, [60] highlighted the United States National Institute of Standards and Technology (NIST) 800-61 which regulates incident response best practices, emphasizing the involvement and role of legal professionals. According to this standard, legal professionals should review incident response plans, policies, and procedures to ensure their compliance with the law and federal guidance, including the protection of privacy rights. Particularly, they are responsible for ensuring that organizations adhere to the laws and regulations and for protecting the personal information of individuals. Some of the key responsibilities of legal professionals in incident reporting and compliance risks include; (1) identifying and categorizing personal data leakage and (2) performing compliance analysis, identifying legal-related issues, and determining the legal consequences of an incident (see Figure 1). However, effective communication between legal and technical professionals remains a challenge in this area [60].

It is worth noting that an incident response report is typically written by cybersecurity risk analysts or collected from external sources and includes information about the affected system, steps to reproduce the attack, privacy and/or security implications such as data breaches, and recommendations and an action plan. The personal data in incident responses is often described using technical and domain-specific language, which is different from the high-level privacy concepts defined in regulations such as FIPPs [35], DPD [48], and especially GDPR [37]. The responses are typically handled by Information Technology (IT) experts and legal professionals within organizations.

In our study, we investigated the procedures and strategies used by legal professionals in the data breach incident response process to assess how effectively they identify and understand privacy-related items within incident responses within addressing GDPR-compliance requirements.

2.4 Study Scope

We aim to explore the practices and expectations of legal professionals in the context of data breach incident reporting, with a primary focus on compliance with GDPR. This focus is driven by the importance of GDPR in shaping data protection strategies (see § 2.1), which legal professionals worldwide practice in ensuring organizational adherence to its requirements. Additionally, our study is specifically concentrated on the privacy impact assessment aspects of incident response under GDPR. However, it is important to acknowledge that other compliance concerns, such as those related to health data regulations or sector-specific requirements, fall outside the scope of this analysis.

3 METHODOLOGY

Our study aims to explore legal professionals' perspectives and practices of incident response assessment through three research

questions outlined in §1. These questions focus on legal professionals' criteria for identifying privacy-related items under data privacy laws and regulations (**RQ1**), understanding regulatory compliance challenges in incident responses (**RQ2**), and determining key features of high-quality incident responses that facilitate efficient legal compliance analysis and reduce privacy risks (**RQ3**).

3.1 Study Procedure

This section outlines the methodology of our simulated incident response investigation experiment and the semi-structured interview process. The entire study procedure for each participant lasted about 90 minutes, including a 5-min informed consent process, a 50-min experiment phase, and a 35-min post-experiment survey and interview phase. All procedures were approved by our institute's Institutional Review Board (IRB). Below, we elaborate on the study.

Informed consent process. Participants were informed about the survey's purpose, structure, length, and raffle. We obtained both verbal and written consent from them before the study (see Appendix A.4).

Experiment phase. In the experiment phase, each participant was randomly assigned 3 real-world incident responses and articles on Qualtrics, which the selection criteria were discussed later in this subsection. Participants were instructed to complete the following two tasks for each response or article:

• ***Incident response inspection task***, where participants conducted a simulated investigation on incident responses for a European Company, with a focus on privacy risk assessment and GDPR-compliance analysis. During the experiment, the participant should (1) highlight privacy-related items in the incident response report (**RQ1**); (2) match eight GDPR sections with incident responses they reviewed. These sections were associated with the enforcement sections against violations of users' privacy rights (e.g., GDPR Article 4 (1) [37]), and requirements ensuring appropriate level security measurements for data controllers (e.g., GDPR Recital 78 [37]) (**RQ2**). Participants had the opportunity to refer back to the response report during the experiment and were allowed to modify or refine their answers during the survey.

• ***Comparison task***, where participants were instructed to highlight privacy-related items and match eight GDPR sections in a new article related to data breaches.

Post-experiment survey and interview phase. Following up with the experiment, we asked study participants to evaluate whether the responses adequately help legal professionals recognize privacy and security laws and regulations (**RQ3**).

• ***Survey***. We developed three questions to gain a deeper understanding of the participants' assessment of the incident responses in our study. More specifically, the survey included questions about the participants' ability to understand the concept and evaluation of incident response qualities. We asked them to rate the quality of each response on 5-Likert Scale (where 1=Strongly Disagree, 5=Strongly Agree) in terms of its ability to assist legal professionals in recognizing privacy-related items and privacy risks, and performing compliance analysis. We also asked them to specify the factors they used in assessing the response's quality. The main survey is provided in [15] and the post-interview questions are listed in Appendix A.7.

• ***Semi-structured interview***. In the follow-up interview, we reiterated the experiment to gain insights into their reactions during the incident response report inspection. The interview was conducted immediately after the experiment via Zoom. We prepared a semi-structured questionnaire to standardize the questioning for all participants. The interviewers used this questionnaire to gather information about the participants' experiences and perspectives. The interview was recorded for subsequent analysis, and each interview lasted approximately 35 minutes. In our study, we designed the following interview questions to provide context, add nuance, and corroborate information for the survey:

• ***Strategy reasoning***. We designed eight interview questions with the objective of gaining insights into: (1) how participants identify privacy-related items; (2) how they align specific sections of laws and regulations with privacy risks in the incident responses; (3) how they use legal knowledge to interpret the responsible subjects, obligations, and terms in incident responses; and (4) if applicable, the reasons for not considering certain privacy-related items as "sensitive" under GDPR.

• ***Challenges, barriers and suggestions***. We developed 12 interview questions to gain an understanding of the following: (1) why participants experience difficulty in recognizing privacy-related items in incident responses, (2) why participants feel challenged while conducting privacy risk assessments, and (3) why the quality of the responses falls short of the participants' expectations, as well as their recommendations for improving incident responses.

The interview script for this study can be found in [8].

Report and article selection. In our study, we carefully collected 9 real-world data breach incident responses from 9 different collaborative tech organizations. Each response had its own structure and technical terminology. The average length of the incident responses in our study is 673 words, which is comparable to that of Cybersecurity and Infrastructure Security Agency (CISA) incident responses, i.e., 642 words [2]. These incident responses were technically detailed and utilize specific technical jargon, mirroring the type of incident responses that legal professionals typically handle. During our study, we collected participants' feedback on their experience with investigating the reports and the quality of the reports we provided. None of the participants indicated any discrepancies between these reports and the type of incident responses they typically review (see §4). Also, authors with legal backgrounds developed this study based on their professional experience to ensure the study design aligns with the legal professionals' work on incident report investigation.

Additionally, for the comparison task, we included 4 relevant news articles pertaining to data breaches This dual approach helped us gain insights into the criteria used by legal professionals and the challenges they face in privacy compliance analysis. The response reports and news articles can be found in [15]. The selection of the number of responses in the survey was based on the number of participants recruited (33), ensuring that each response was investigated by 9 participants on average to provide a diverse sample for our research.

3.2 Recruitment and Demographics

Participants were recruited through social media as well as alumni email lists from law schools and universities' online interdisciplinary programs in cyber risk and law (see the recruitment ads in Appendix A.1). Prior to the experiment, each candidate completed a screening survey (See Appendix A.2), indicating prior experience with privacy and security laws. Additionally, we conducted a brief screening interview to understand the participants' experience and jurisdictional knowledge in practicing data protection laws, particularly GDPR and its compliance standards. This involved asking questions to describe their GDPR compliance standards and how they interpret these standards with their typical job routines. (see Appendix A.3 for further details). It took approximately 90 minutes to finish the study, based on the hourly paycheck of legal professionals, participants will be compensated with $50 USD.

In total, we recruited 33 participants diverse in age, gender (female=17, male=16), and region (EU, US, and China). These individuals have varied professional experiences across law firms, government agencies, and corporate legal departments practicing areas that involve providing services to EU residents and require compliance with GDPR. 29 of them are licensed legal professionals, holding certifications such as Certified Information Privacy Manager, Certified Privacy Technologist, Certified Compliance & Ethics Professional, or law degrees, including Juris Doctor (JD), Master of Laws (LLM), and Bachelors of Law (LLB). 4 of them have been pursuing their related degrees. We demonstrate participants' demographics in Appendix A.5 Table **??**.

3.3 Data Analysis

3.3.1 ***Quantitative analysis.*** We measured the agreement among participants regarding the recognition of privacy-related items, where GDPR's definition of personal data were interpreted for privacy-related items and privacy compliance analysis (see §4) from three aspects:

• **For *item-level* measurement of privacy-related item recognition,** we considered each candidate privacy-related item as a subject to be identified, and the annotation is regarding whether a subject is recognized as a privacy-related item by each annotator. Based on Fleiss' Kappa [33], a measure to determine the level of agreement between two or more annotators, we calculated the inter-annotator agreement rate P_i for the subject i as:

$$P_i = \frac{1}{N_i(N_i-1)} \sum_{j=1}^{C_i} n_{ij}(n_{ij}-1), \tag{1}$$

where n_{ij} represents the number of annotators who assign the j-th annotation to the i-th subject, $i \in \{1, 2, \ldots, S\}$ is the index of subjects, $j \in \{1, 2, \ldots, C_i\}$ is the index of annotation categories, and N_i is the total number of annotators for the subject i. We let $Ci = 2$ where $j = 1$ indicates the item is identified as privacy-related and $j = 2$ indicates the item is not identified as privacy-related. To collect the candidate privacy-related items, we included those identified as privacy-related items by at least one participant. As a result, 108 candidate privacy-related items were recognized, and 1,724 annotations were made by the participants, which is calculated as

$$N_A = \sum_{i=1}^{S} N_i C_i. \tag{2}$$

• **For *report-level* measurement of privacy-related item recognition,** we measure the observed proportion of agreement regarding a report d (i.e., incident response or news article) , using the following formula

$$\bar{P}_d = \frac{1}{S_d} \sum_{i=1}^{S_d} P_i, \tag{3}$$

where P_i is the agreement on the i-th candidate privacy-related item in report d and S_d is the number of candidate items in report d.

• **For *privacy compliance* analysis,** we measure the inner-annotator agreement among participants regarding whether a report (i.e., incident response or news article) involves privacy-related items governed by a certain data protection provision. We calculate the Fleiss' kappa [33] as:

$$\kappa = \frac{\bar{P}_d - \bar{P}_e}{1 - \bar{P}_e} \tag{4}$$

$$\bar{P}_e = \sum_{j=1}^{C} \left(\frac{1}{N_d S_d} \sum_{i=1}^{N_d} n_{ij} \right)^2, \tag{5}$$

where $\bar{P}_d$ is computed using Equation 3 with considering each provision as an identified subject, N_d and S_d are the number of annotators and privacy-related items of report d, and $j \in \{1, 2\}$ indicates whether report d involves data items governed by the provision. We collected 8 sections from GDPR related to personal data, data breach, and incident response, resulting in a total of 1,536 annotations.

3.3.2 ***Qualitative analysis.*** For the semi-structured interview data, we conducted thematic analysis to identify common themes and patterns in the participant responses. Specifically, we used inductive coding [30] to analyze participants' responses in the interviews, as elaborated below:

• **Transcript processing.** The primary coder transcribed each Zoom audio recording into text using a speech-to-text converter. The primary coder carefully reviewed and aligned the recorded voices with the generated transcripts. Each transcript was then saved as an individual text file. These text file transcripts were then uploaded to NVivo R.14.23.0 for coding [41].

• **Data conceptualization and segmentation.** The primary coder reviewed the transcripts at the sentence level and applied codes accordingly to create a codebook (see Appendix A.6 Table 3).

• **Codebook verification.** A second coder reviewed the coding and provided a summary with added disagreements and comments, which were discussed with the first and third authors. Next, the secondary coder coded 20% of the subsample for each topic. The coded results were then iterated with the primary coder until Fleiss' kappa, which represented the inter-coder agreement, was greater than 0.9. For questions with insufficient responses to calculate kappa reliably, the primary coder and secondary coder collaborated to assign codes. Coding conflicts were resolved through group discussions among coders, following the practices of other studies [29].

Note that we observed a robust data saturation for our user study on personal data identification and privacy compliance. Through the iterative examination of participant responses and interactions,

themes emerged consistently after the 29th participant across the dataset, suggesting thorough coverage of relevant dimensions. More specifically, the data saturation point regarding the **RQ1**, **RQ2**, and **RQ3** was reached at the 28th, the 28th, and the 29th participant, respectively. The saturation analysis revealed a convergence of insights regarding users' understanding of personal data and privacy risks, their approaches to compliance measures, and their suggestions to improve the quality of incident response reports.

3.4 Discussion

Ethical considerations. The study was conducted in accordance with the requirements for studies involving human participants set by the institution's IRB. Before launching surveys and starting interviews, all participants were briefed on the study's objectives and the rationale behind its conduct. They were informed about how to reach out to the researchers for any additional inquiries. The data collection consent form (see Appendix A.4), study sections, and the approximate time required to complete the survey were also presented. In addition, we informed participants that they had the option to turn off their cameras during the interview if they preferred to do so, to ensure their comfort and privacy.

During the analysis stage, a variety of tools, including otter.ai for audio interview transcript conversion, NVivo for analyzing the participants' responses and creating the codebook, and a text file for analyzing interview transcripts, were employed to manage and analyze the data. The primary coder ensured that any participant's PII was removed to protect their confidentiality and comply with ethical guidelines for research involving human subjects.

Limitations. Like most qualitative and exploratory human subject research, our study is subject to the following limitations. First, the self-reported data we rely on may be biased due to social desirability, availability bias, and incorrect recall or self-assessment. However, our aim was to gain an initial understanding of identifying private information in incident responses. Second, our participant recruitment focused on individuals knowledgeable about GDPR, potentially limiting the diversity of technical insights and perspectives across different backgrounds. Third, we did not aim to make generalizable claims as we did not conduct a large-scale randomized study with many participants. Our sample is biased towards young participants. Additionally, our sample is hardly representative of all legal professionals but rather serves as a first step in shedding light on this concept. Fourth, the study, which involved 33 legal professionals from diverse backgrounds, included only two participants from the EU. We recognize that this distribution might potentially introduce biases or gaps in understanding that are specific to the EU context. However, we have addressed this limitation through careful selection criteria. Most participants, regardless of their location, engage with the GDPR in their professional practice. This ensures that even non-EU participants are likely well-versed in the nuances of the regulation. While the geographic diversity of the participants could theoretically introduce some biases, their extensive professional experience with GDPR is expected to mitigate this effect significantly. Therefore, although the limitation is recognized, it is unlikely to substantially undermine the validity of the study's conclusions regarding GDPR compliance practices among legal professionals.

Despite the limitations, we believe that this study is important for the cybersecurity and privacy landscape as it shows how legal professionals handle incident reports and how incident responses inform the development of better guidelines and policies that enhance the overall cybersecurity infrastructure. Especially, academics involved in policy-making, education, and cybersecurity frameworks can find these insights valuable.

4 RESULTS

In this section, we report all significant results in both survey answers and interviews regarding our three main research questions.

4.1 RQ1: Privacy-related Items

In response to **RQ1**, *what are the criteria by which legal professionals can identify privacy-related items (e.g., personal information, Apple advertising ID) under data protection laws and regulations from incident responses?*, we investigated our participants' criteria and strategies to recognize privacy-related items.

4.1.1 ***Criteria of privacy-related item recognition.*** In our study, participants (n=33) identified privacy-related items under the GDPR definition in 9 incident responses. On average, each participant annotated seven privacy-related items per repo. The observed proportion of agreement (i.e., average P_i on all data items) regarding whether a data item was privacy-related among all 108 candidate data items is 0.647. We observed that participants have generally higher agreement on news articles than on data breach incident reports. More specifically, the $\bar{P}_d$ of all incidence responses were below 0.65, ranging from 0.503 to 0.645, with an average of 0.546 and a standard deviation of 0.042; while the $\bar{P}_d$ of news articles were above 0.8, ranging from 0.804 to 0.875 with an average of 0.844 and a standard deviation of 0.037. The proportion of agreement for each response or news (i.e., $\bar{P}_d$) can be found in Appendix A.8.

The report with the lowest $\bar{P}_d$ (0.503) is an incident response describing security weaknesses in an Internet of Things (IoT) smart plug that led to home safety concerns, including unauthorized disclosure of PII. While the response extensively discussed technical details, such as the threat model, and identified security weaknesses, it may be challenging for those without technical backgrounds to understand (see §4.3). As a result, there was a significant disagreement among participants in regards to the identification of privacy-related items, such as user certificate, device MQTT topics, device mode, and device shadow. These involved data items fall into blurred cases and require a comprehensive understanding of their context, including the usage scenario and the risk of personal data leakage, in order to be properly evaluated. However, some participants found it confusing to identify these items as privacy-related. For example, P28 said,

> *"I don't think there is privacy risk in user certificate and user right. But so while I was reading those I'm trying to think of analogues that I have seen, or any place that seems unreasonable, and one of those in the study would be most likely the ability for the alarm status to be changed through MQTT messages to me, that seems like, okay, that's the purpose of the system is to provide security and notify customers when the alarm is happening. So any sort of breach that happens*

to the central purpose of the technology that seems like that would be an unreasonable risk. So kind of privacy related."

Meanwhile, the report with the highest $\bar{P}_d$ (1) is a news article about an incident involving relatively easily recognizable personal information such as the user's age, audio recordings, and location.

We summarized participants' criteria of privacy-related item recognition as elaborated below:

• **Personal information defined in laws and regulations.** In our study, 15 out of 33 participants recognized privacy-related items that are listed as personal data in GDPR. As discussed in Section 2.1, GDPR defines personal data and provides some examples, including name, identification number, location data, telephone, and credit card information. We found that geographical location, age, and email address were the top three privacy-related items all participants considered sensitive, which are all classified as personal data under the GDPR. Beyond the items specified in GDPR, the most common strategies are simply matching whether the data item is related to an individual, as P7 mentioned,

"I determine the personal data by considering whether the data is able to link to individual identify if it is then that kind of data falls under that with GDPR and considered as personal data."

However, some participants (n=8) pointed out that GDPR has different definitions and updates for PII, personal data, and sensitive data, making it difficult to align them. For instance, P27 said,

"To be honest, over the last several years, or even more, the definition of personal information has evolved. There have been additions to sensitive information under GDPR and other laws. However, I've always believed that certain types of data were inherently sensitive, regardless of legal definitions. These categories, now highlighted by various governments, were always considered sensitive information by us, regardless of whether US law recognized them as such."

5 participants excluded data indistinguishable between individuals. When asked whether the shared user certificate/credential of Switchmate users belongs to personal data, P21 commented

" I do not determine the shared credential as personal data if it is the same among all users, even if the credential enables communication with the system or login to the account since there is no association or distinguishable signals between the password and the user."

8 participants with Juris Doctor (JD) and Bachelor of Law (LLB) degrees had a narrower understanding of privacy-related items and considered only PII as privacy-related. They excluded device data such as device serial numbers or status, which they believed were relevant only to the device and not to the user.

• **Data ownership.** 9 participants examined the role of the data owner when identifying personal data. They checked whether the data owner is a data subject defined in the GDPR and is protected by the data subject rights. A data subject refers to an identifiable natural person whose personal data is being collected, held, or processed. Under the GDPR, data subjects have specific rights, including data portability and rectification, and the right to access, erase, restrict processing, and object to the processing of their data. When asked whether a secret account and its certificate of a malicious adversary are personal data, P26 commented

" In the context of data protection regulations, an attacker or adversary of a software application would not typically be considered a data subject. If a hacker created an account of his own for the attack, such information, though related to the attacker himself, is not concerned by the data subject rights. "

They also associated the level of sensitivity of the personal data with the role of the data owner, as P26 commented

" From the perspective of GDPR enforcement, data subjects concerned by the data protection agencies can be the end-users of the application, the employees of the company, and the customers. Among them, the affected end-users are mostly concerned and the violation may impose severe penalties such as fines and business disruption. "

• **Associated with security breaches.** 11 participants also responded using existing security breach responses to identify privacy-related items. When reviewing incident responses, they examined whether there are specific indicators that suggest a privacy data breach has occurred due to unauthorized access, use, disclosure, or transfer of data in the system. They determined the level of sensitivity of the data by assessing the severity of the consequences resulting from a data breach, including financial loss, identity theft, reputational damage, and regulatory fines. P19 mentioned that

" The leakage of personal data is generally measured from the point of view of infringement. For example, when a user's account can be logged in without authorization due to a data breach, we further examine the potential risk or loss such as the damage to property rights if money can be transferred through the account or the damage to reputation if false statements can be published through the account. "

For example, P15 identified privacy-related items by determining what type of data (e.g., page likes, followed groups, preference settings, contact information) the attackers can steal once they have gained unauthorized access to a user's social media account. P2 and P14 defined their criteria to identify privacy-related items as *"analyzing what type of data can be exposed once the device gets hacked"*. P2 also said

"So, the responses are saying that an unauthorized user can take control of IoT smart home devices remotely by exploiting a weakness in the device and I think any data stored in the device that has the risk to be exposed should be considered as private data."

3 participants also considered the timeliness when identifying privacy-related items. They claimed that time-related information is necessary to determine the sensitivity of data such as temporary passwords. For example, P23 said

" To determine this, I will ask whether this PIN code is still in force when it's disclosed since it serves as a temporary passcode. If the PIN code has expired, such leakage will not result in any privacy risks. In such

> *cases, I will consider it as not sensitive. Complementary information such as the period of data retention, timestamp of the data leakage, the expiration date, can be helpful in the determination. It would be beneficial to see such information in the incident reports. "*

• **Elements of privacy design/practice principles.** 3 participants identified personal information through privacy practice principles, which recommended a set of data processing actions, (e.g., minimization, access and control, transparency, and aggregation/anonymization) that organizations could take to ensure the protection of personal data and respect for privacy rights. If data was processed by those the best practices in privacy design, the participants would have considered the data to be personal information as it is protected. Regarding to the criteria, in our study, P9 mentioned these practices as below,

> *"When I was reading the mitigation plan, I noticed that social media data such as posts, likes, comments, friend lists can be considered to be de-identified meaning that they can no longer be linked to an individual. While I am unsure of the specific privacy impact of this data, I believe it is considered sensitive and is protected by privacy principles."*

• **Others**. 7 participants did not use the aforementioned criteria to identify privacy-related items, instead relying on their sensitivity including subjective preferences, understanding of the data properties, and surrounding context. P1 considered all data related to sexual activity as extremely sensitive, including the *sex time* recorded by an IoT sensor. P5 identified the *secret key* as sensitive because the modifier "secret" indicated confidentiality. P10 considered the *account topic* to be sensitive as it could be used to infer user identity based on the context. When encountering unknown technical jargon, 5 participants analyzed the applicability of the above criteria with the help of contextualized information. For example, given the context "*a user uses Govee app to control the device remotely by publishing a message to the device topic when the message contains a field which is called 'accountTopic' to express the sender identity.*", P21 commented

> *" I think accountTopic is also a personal data, although I do not know the actual meaning of this term. In the context of this report, the account topic is used to express the sender identity, which is obviously sensitive information. So I think the account topic here is also sensitive. "*

2 participants also assessed privacy-related items according to whether they were *public* or *private*. For example, P26 considered *Ed25519 public key records* as personal data since it is public although *the corresponding private key* was considered as privacy-sensitive. Interestingly, there were no significant differences in identifying privacy-related items among our participants based on their region and GDPR practices responses from different locations. However, the variations in each legal professional's interpretation also showed that there is no specific identification measure for privacy-related items.

4.2 RQ2: Privacy Compliance

In response to **RQ2**, *how do legal professionals strategize to map privacy risks in incident responses to ensure compliance with the specific sections of relevant laws and regulations?*, we queried the criteria and process of such alignment, investigated the consistency or discrepancies in their legal assessments, and examined the challenges they face in the legal compliance process.

4.2.1 ***Strategies for aligning legal sections.*** We found 1 incident response and 3 news articles have Fleiss' kappa scores above 0.6. The Fleiss' kappa of privacy compliance analysis for each response or news (i.e., κ) can be found in Appendix A.8. The high Fleiss' kappa scores in news articles might be because those articles included clear statements of privacy impacts associated with the reported data breaches and were written in an accessible and easy-to-understand language, avoiding technical jargon, making the information understandable for a less technically-savvy audience. The incident response with high Fleiss' kappa was thoroughly documented, effectively illuminating the privacy implications, such as the extent of data leakage and affected individuals, as well as potential compliance issues.

We investigated the common legal compliance strategies among participants for these incident responses. Particularly, 19 participants aligned privacy risks in incident responses with the specific sections of related laws and regulations by examining whether the obligation described in the articles can be triggered by the incident. This requires that the subject/region of the incident is regulated by the article, and the conditions to trigger the obligation are met by the incident. For example, P18 commented

> *" The region or country of the incident is one of the condition to check whether the law is applicable."*

When aligning an incident response to GDPR Article 33[2], P19 mentioned

> *"This article imposes obligations to the data processor. Hence, when aligning any incident to this article, I will first check whether the subject of the incident response is a data processor. If not, this article will not be applicable to the incident with no need to check other information."*

When aligning an incident response to the GDPR Article 83 [3], P25 analyzed it as

> *" It is also in my workflow to check whether the precondition of triggering the obligations has been illuminated by the actions taken. If the company has already taken appropriate measures to mitigate the risks, the requirements of this article will not be met to trigger the obligation. "*

However, some participants (n=17) wanted to learn additional information and assistance regarding incident response and personal

[2]The processor shall notify the controller without undue delay after becoming aware of a personal data breach

[3]In order to maintain security and to prevent processing in infringement of this Regulation, the controller or processor should evaluate the risks inherent in the processing and implement measures to mitigate those risks, such as encryption. Those measures should ensure an appropriate level of security, including confidentiality, taking into account the state of the art and the costs of implementation in relation to the risks and the nature of the personal data to be protected.

data in this phase. They wanted to make sure to enhance their understanding of incident responses and assess the impact on personal data. Moreover, these participants reported that assessing the size of impacted communities and coordinating with relevant teams is essential for effective incident response. Legal counsel or team can collaborate with counsel, incident response teams, compliance officers, cybersecurity, and IT experts to address privacy risks comprehensively. They reported this may involve collaboration with cybersecurity experts, compliance officers, and IT teams. P29 said

> *"Yeah, I might want to make sure that someone I can't. Just give them a call to ask. And yeah, and I think that the subscribing to, if I were like in an an organization, I would subscribe to the vendors the compliant vendors as they can help and save the time, and can imagine there are a lot of reports to go through and it's due to like, not to spend too much time on one report, not the others."*

Furthermore, 3 participants also preferred finalizing compliance for incident responses by conducting audits, ensuring that no aspect of incident response procedures was overlooked. These audits encompassed a comprehensive review of organizational policies, procedures, and engineering practices, with a particular focus on identifying areas for improvement and mitigating potential privacy risks proactively. Also, 2 participants highlighted the significance of ongoing compliance with legal requirements, including those stipulated by the GDPR and other relevant laws and regulations regarding the subject and region.

4.2.2 ***Challenges.*** We found 5 reports with low Fleiss' kappa scores (below 0.1). We summarized the underlying reasons for the substantial disparities in the privacy compliance analysis as follows:

• **Gap between system context and legal context.** Incident responses written by software engineers or security experts focus on the technical aspects of a system, providing detailed information about the components and tactics used by attackers. These descriptions could be challenging for legal professionals to understand, leading to disagreements when applying legal context to the incident report.

P9, P17 and P11 said respectively,

> *"I think there is a huge gap between those who write the incident responses and those who review them. The way the attack scenario is depicted is very technical, using complex protocol graphs, and I think that reviewer with different levels of technical expertise can lead to varying legal interpretations"*

> *"There seems to be a significant disparity between the technical language used in incident responses and the level of expertise possessed by the individuals responsible for reviewing them. This creates a risk of varying legal interpretations, as the complex technical jargon may be misinterpreted or not fully understood by some reviewers."*

> *"Well, it's hard because I just don't have the technical knowledge to understand the reports, so I would definitely have to reach out to others to understand them, so that I could know what personal information, if any, was being leaked, and then work towards the I notification of the parties, etc."*

• **Alignment with pertinent laws and regulations**. There were concerns regarding alignment with pertinent laws and regulations, particularly at the federal or state level in relation to United States laws. P28, a privacy and cybersecurity law expert, specifically expressed their concerns when implementing their national compliance strategy for incident responses.

> *"...in the US, we have more of a sectoral approach to privacy and cybersecurity. So the first thing I'm looking at is okay, what kind of data is this? If it's health data? Alright, I need to go look at HIPAA and HITECH and see what's going to be covered. If we're dealing with financial data, I know I need to go to standards of Gramm-Leach-Bliley. If I'm looking at insurance and we operate in New York, I'm going to be looking at NYDFS. So, there's just an endless number of regulations that you may need to bring in. "*

4.3 RQ3: Quality Improvement and Barriers

In this section, we discussed **RQ3**: *how can the quality of incident responses be improved so that legal professionals can better understand them?* We also detailed plausible writing styles and quality concerns in the incident responses.

4.3.1 ***Low Quality Features.*** The term "low quality" referred to participants' lack of engagement with the incident responses, missing items in the reports, and limited ability to comprehend the content of the reports. Based on the study participants' feedback, we presented the following low quality features in incident responses.

• **Lack of organized structure.** 11 participants reported that some incident responses were poorly organized and that they were confused by the structure of the incident responses. Participants expected a more professional report with a well-organized structure. For example, P7 and P13 said

> *"I just felt a little bit confused. I did not understand exactly what's happening there. I did not encounter those problems when I'm reading an incident from an incident tree. But I feel like there are some incomplete questions or something I don't know."*

> *"I would have expected it like a professional report. I would have expected a lot more details, purpose of the report and kind of just better grammatical structure, and it just kind of felt like they took a bunch of information and just threw it onto a page. There wasn't a whole lot of like organizational structure to it. I was very confused."*

P4 also mentioned that the attack scenarios are opaque with the background and reference sections missing in the report; and the lack of cross-check references makes it challenging to understand the meaning of specific terminologies (e.g., Accessory Protocol):

> *"If we do not have any background like on the cyber security things then I think we cannot solve the issues or everything like APIs and everything in the report. I think we have to have a background section in this type*

of reports and then technical aspects. I think it's better to have background in the report."

• **Excessive technical details.** Most of the participants (n=10) mentioned that the reports contained excessive technical details, and they were not adequately prepared to understand such reports. They felt either confused with the content or misunderstood them. P15 mentioned after reviewing incident responses:

" Because there were too many technical terms involved, I couldn't quite understand some of the reports easily, especially as someone who is new to this field. It's like trying to make sense of something you've never encountered before."

P2 and P6 also stated respectively

"The first two are really technical, especially when they use complex graphs to show information flow between different vendors in a specific system architecture."

"I got the whole security vulnerability idea but I did not catch any technical stuff they were mentioning in the reports. And, there was no description on that."

P11 said

"So I felt like I didn't have like the technical background to understand some parts of the report. There is like a lot of technical jargon that I didn't understand. I don't. I don't know coding, so I didn't understand those portions, and then it was hard for me to wrap my head around like I could understand that there had been an incident, or that there was like a breach, and I could even sometimes piece together what was happening. but there was no indicator of what personal information may have been at risk so trying to keep that together was a little bit difficult."

• **No description of violations of laws.** 12 participants expressed the need for incident responses to include a clear description of any legal violations or potential compliance risks. The participant P2, whose professional title was *lawyer*, mentioned that *legal department in an organization cannot internally handle those reports because there is no potential privacy compliance risks mentioned in the report.* P9 and P17 similarly agreed that if they were working in an organization, they would not send those reports to legal department due to the lack of identifying compliance risks:

"This could possibly apply for, you know potential liabilities here. I would check to see the cyber security and privacy engineer to understand the report. What data do we have stored? How much of that data is personally identifiable, or you know, confidential, whatever the case may be something that. If it was late, would lead to legal ramifications. Otherwise, I don't think legal department can handle it at first."

P7 wanted a more detailed description by saying,

"There is no detailed description for laws in the reports. They would have mentioned critic sections of laws at least."

• **Lack of scope of affected users**. 9 participants expressed the importance of presenting the scope of the affected users in the incident response. The distribution region of the application, the number of users affected, and the role of users are specifically concerned when determining the applicability of specific laws or regulations. P25, P26, and P18 stated respectively,

"It is important to include the countries and regions the application is sold to, especially in the cross-border cases. This is crucial for determine whether it is regulated by the GDPR or other data protection laws."

"We might take different actions regarding the roles fo the end users: are they downstream customer companies, or the end-users, or belong to specially protected population such as children? Will this potentially result in class actions, if a large population are affected?"

"Knowing how many users are affected is also helpful to assess the severity of the incident, for this will result in different degrees of litigation and different levels of penalties; then we could decide how to prioritize the following actions in response to the incident."

4.3.2 ***High Quality Features.*** We summarized high quality features as below.

• **Inclusion of technical details.** 7 participants with a strong background in cybersecurity and risk management rated "technically detailed" as a high-quality feature because it covered all bases, such as steps to reproduce the attack and mitigation solutions. For example, P4 mentioned,

"I think those reports are high quality because with pretty technically details. With a detailed PoC description, the engineers can reproduce them and better assess its severity confirm the privacy data exposure."

• **Inclusion of privacy concerns and impacts.** 12 participants praised the incident responses that comprehensively cover the fundamental aspects for legal compliance analysis, such as the extent of data leakage and affected parties, as well as privacy risks. They appreciated the reports that identified and listed privacy and security concerns and their potential impact on individuals and organizations. For instance, a participant (P11) did not like excessive technical details but appreciated how the report *"spoke to privacy and security concerns"*. Here, we also quoted a participant P16's response how they found the report's quality with regard to privacy concerns:

"They made it very clear that the companies were intentionally sharing that information with, you know, third parties. Privacy part is more useful."

4.3.3 ***Recommendations.*** The participants of this study were asked how they could improve the effectiveness of privacy compliance analysis in incident responses. The majority of participants (n=19) expressed that a privacy compliance analysis would require information presented in a more generic manner, rather than being too technical and concise. While P28 mentioned specific missing items in these incident responses such as "what data is at risk", P11 recommended adding the following items in incident responses to bridge the ambiguities in tech and legal:

> *"It's nice to see how many people were impacted and then a layout of what exact information either was compromised or could potentially be compromised due to this, and seeing how far spread it it is. Those are the kinds of things that are important. Also, who exactly was impacted? Who might be impacted? What sort of data could be taken? A clear and simple explanation of what occurred, who it impacted, and what preventive measures could be taken in the future or during the incident. These pieces..."*

P29 also reiterated the importance of understanding the context of incident responses and added their recommendations to enhance its quality for legal professionals:

> *"I couldn't like understand the context even clearly when I tried to read them so in terms of the presentation. I think the reports could like make clear what the affected devices are, and in what context they can be used? Or is it like of vulnerability in general? Or is it based on a case that has already happened? This contextual information would be helpful."*

Additionally, some emphasized (n=7) that incident responses rely on the context and background and specific needs of legal professionals regarding regulatory compliance and addressing privacy risks. While they highlighted the need of a few technical details but providing high-context information would be beneficial to understand the potential challenges comprehensively.

17 participants also believed that a technical expert-in-the-loop privacy assessment and privacy compliance analysis would be beneficial.

Particularly, participants P3, P8, and P9 recommended to improve the presentation of incident responses, respectively:

> *"The reports would be more specific on the steps they have used. The reports should answer why is this a problem?, how to deal with it?, what should you do about it?, and how would you prevent individuals from those vulnerable people or would you prefer preventing them?. And you would probably need to address what sections of laws can be violated or what security or confidential issues are under an act. Issues and backgrounds should also be well-written."*

> *"I think a good incident response should satisfy the expectations from three main readers. First, it should have technical details to help engineers understand the steps involved in the attack and how to recreate it. Second, it should highlight any privacy risks or harm that could arise, to help lawyers assess any legal implications. And third, it should show the impact on real-world scenarios to help product stakeholders adjust their marketing and PR strategies."*

> *"It might be useful for the preparer of the attack report to generate a glossary of terms bu this could run into problems because it's challenging for IT and cybersecurity professionals to automatically know which terms might be unfamiliar to others. Creating a perfect glossary could be difficult, but collaboration could help. Security professionals could work with legal professionals to synthesize the information before the attack report is released but again such partnerships may not always be feasible due to time constraints."*

3 participants recommended aligning the key performance indicators (KPIs) from the perspective of company policy to motivate the collaboration between legal professionals and technical experts toward a smoothier privacy compliance analysis. For example, P18 commented

> *" The most challenging thing is not about techniques or skills; it's about the motivation. Under the current corporate system, legal compliance analysis is not in the job duty of engineers and they are thus not inspired to assist. To fundamentally address the communication gap between legal and technical teams, companies are responsible for proposing new policies to quantify and credit such workload. "*

5 RECOMMENDATIONS

A high-quality incident response should fulfil the expectations of all stakeholders, e.g., security/privacy engineers, legal professionals, IT administrator, by (1) striking a balance between technical and general information to satisfy varying expectations, (2) featuring comprehensive key sections, and (3) having a well-structured and organized format.

Balance between technical and legal expectations. For cybersecurity and privacy engineers who have a strong technical background, the response should include a comprehensive proof-of-concept, which provides a clear step-by-step guide to show how to replicate the attack. IT administrators responsible for fixing issues should be provided with actionable details of the mitigation plan, including updates to software, stronger passwords, or changes to security settings. However, for legal professionals who may lack technical knowledge, the response should be written in a more general and concise manner, avoiding technical jargon. This would allow them to effectively identify any data compliance issues. P2 and P8 also said the followings respectively,

> *"To make sure that lawyers can understand it, incident responses need to be clearer about technical terms and presented in a way that all professionals can understand. It may also help us to comply with data privacy laws and regulations."*

> *"It is challenging to achieve complete compliance with incident responses that cater to both technical and legal audiences, but it is imperative that these responses are designed to meet the needs of all relevant stakeholders. Doing so ensures maximum efficacy in preserving data privacy and security standards."*

Therefore, an incident response must be designed to meet the unique needs of each stakeholder while effectively communicating critical information.

Inclusive key sections. Incident responses are a critical tool in the realm of cybersecurity for effectively communicating findings related to cybersecurity breaches and attacks. However, ensuring that the response is easily comprehensible and accessible to all

relevant stakeholders can be challenging. According to feedback from study participants, one key recommendation to improve the responses is the inclusion of an executive summary that provides a high-level overview of the assessment for non-technical executives. This summary should highlight any critical issues that might impact corporate cybersecurity or regulatory compliance. Doing so allows executives to grasp the essential findings and understand the impact of data breaches. Second, including a background section in the incident response that contextualizes the privacy risk and helps readers understand its significance is crucial. Providing clear and concise explanations of any relevant technical terms or jargon is also recommended, particularly in cases where the response covers new or emerging fields or paradigms. Lastly, incorporating a clear and concise conclusion section that summarizes the response's findings and highlights potential areas of improvement is essential. For example, P10 suggested that

> *"I could have understood better if the responses summarized most important issues that the company should take an action after they received the response. It should have been included which regulation also might be violated through the potential risk."*

Therefore, those recommendations can enhance readers' understanding of the response's findings and improve its overall effectiveness.

Well-organized structure. We suggest an incident response should be organized with the elements of summary, background, reproducing steps, privacy implication and impacts, recommended mitigation, recommended legal actions, and reference. It should also be presented using bullet points and graphical elements such as flow charts, diagrams, or graphs, to enhance its readability and facilitate understanding of its content. For example, here are the most coherent quotes from participants, P17 and P14,

> *"In those responses, there was no specific structure that I could have followed like an essay. A general structure would have been better everyone to understand."*

> *"It could have been included statistical or graphical responses that can indicate how the potential risk might impact the company. It would have made the readers take an action immediately before trying to understand the response itself."*

6 DISCUSSION

Deficient assistant tool. To the best of our knowledge, there is currently no tool that bridges the knowledge gap between technical and legal professionals in regards to the translation of technical terminology into legal definitions. To address this problem, a promising solution would be to develop a legal compliance assistant tool that provides an explainability feature to translate technical terminology into legal definitions. Specifically, this tool should enable users to understand the relationship between potential privacy risks associated with a data item and the relevant citations or sections in laws and regulations that govern it. Such an assistant tool will aid both technical and legal professionals in understanding the privacy risks associated with technical data, enabling them to take proactive measures to mitigate these risks and ensure compliance with applicable laws and regulations.

Contextual factors in privacy risk assessment. In our study, we discovered that the challenge of privacy risk assessment and privacy compliance analysis was intensified by the complex contexts surrounding privacy-related items. We summarized those contextual factors below.

- ***Data source***. The source of data can significantly impact an individual's privacy. Data generated by regular users may be considered privacy-sensitive, whereas data derived from publicly accessible sources may be seen as less sensitive. For example, data obtained from openly accessible sources like public records is generally regarded as having lower privacy sensitivity. In contrast, data collected from children is subject to stricter privacy regulations in many jurisdictions.
- ***Data characteristic***. The characteristics of data, such as its intended sharing status, play a pivotal role in determining its privacy sensitivity. For example, user certificates are considered sensitive because they serve as credentials. However, shared user certificates are less sensitive since they are intended for sharing among users. Similarly, PIN codes are generally seen as sensitive, but the default factory PIN codes for devices are less so. These examples underscore the importance of considering data characteristics when assessing its privacy sensitivity.
- ***Data processing***. The privacy risk associated with data can also be influenced by the degree of data processing and encryption. Data that has been anonymized and cannot be linked to a specific user poses a lower privacy risk compared to unprocessed or inadequately encrypted data.
- ***Data usage purpose***. The manner in which data is used also plays a crucial role in determining its privacy sensitivity. Data shared with data brokers or advertisers poses a higher risk to privacy because this data can be monetized through user profiling for targeted advertising. This risk is further amplified if the data is sold on underground markets, where malicious actors can use it for spamming. On the other hand, if the data is only used for app functionality and not shared with any third parties, it poses a lower risk to privacy. It is crucial to consider the data usage in order to accurately assess its sensitivity and privacy impact and to ensure that appropriate measures are in place to protect individuals' privacy.

Interdisciplinary training opportunities. In this study, we found that 27 participants expressed the need for additional and relevant interdisciplinary training to enhance their understanding of privacy-related items and privacy laws. Legal professionals for efficient incident responses need to be skilled in both technical and legal issues to comply with laws. In real-world scenarios, these professionals must interact with both legal and engineering fields and possess a harmonious balance of technical skills in computer science/information technology and legal expertise in privacy compliance. The NIST's NICE Workforce Framework [9] outlines the responsibilities of professionals including non-technical knowledge of laws, regulations, policies, and ethics, as well as technical skills. We also believe that legal professionals need a comprehensive training or education in privacy to develop a better understanding of privacy-related topics, including the definition of privacy data.

7 RELATED WORK

Privacy compliance risk. In recent years, the literature has extensively measured privacy compliance risks [23, 24, 27, 28, 31, 44, 45, 52, 58, 63–65] under various consistency models across various systems, revealing the widespread prevalence and severity of privacy non-compliance. PoliCheck [23] found that 42.4% of 13,796 Android apps have inaccurate or omitted privacy policy disclosures for privacy-sensitive data flow, using an entity-sensitive flow-to-policy consistency model. PurPliance [28] revealed a higher prevalence of inconsistencies in data practices among Android apps (69.66% non-compliant in a sample of 23.1k apps) by incorporating data usage purposes into its consistency model. MAPS [64] performed a large scale analysis of 1,035,853 Android apps and found 12.1% of apps have at least one location-related potential compliance issue. Even worse, the issue of privacy violations is rampant among children's apps. Reyes et al. [49] found that a majority of 5,855 popular free children's apps collected children's personal data for purposes of behavioral tracking and advertising, potentially violating the COPPA. Further, the researchers [36, 49, 55, 57] pointed out that the use of software development kits (SDKs)/libraries plays a crucial role in perpetuating privacy violations. Wang et al. [57] conducted a compliance check between cross-library interactions and the terms of service of SDKs, leading to a discovery of 42 distinct libraries stealthily harvesting data from 16 popular SDKs, affecting over 19K apps with a total of 9 billion downloads. In an effort to mitigate the serious privacy risks posed to end-users, Apple introduced its app privacy labels, which mandate that app developers disclose the data collection practices of the entire app, including those conducted by third-party partners [18]. However, recent research [40, 62] has highlighted the widespread and significant issue of privacy label non-compliance (67% in [62], 16.36% in [40]). These findings indicated significant privacy compliance risks even on the iOS platform, which is widely considered to be more trustworthy, reliable, and privacy-conscious [19], compared to Android.

Compared to prior work, we conducted a human subject study, towards a deeper understanding of why compliance audits/enforcement fails from legal professional's point of view and reveal the problems and challenges faced by stakeholders who could not achieve their privacy and accountability goals.

Software engineer and legal professionals' perception of privacy compliance. There are some empirical studies that focus on software engineers' perception of privacy compliance. [42] reported that software engineers can identify ambiguities in regulatory norms but still require further support or legal expertise to resolve. [43] measured the ability of software practitioners to classify legal cross-references and discovered that software practitioners are not well-equipped to understand the impact of legal cross-references on their software. [21] explored the problems developers encounter when including privacy concerning all the GDPR principles and revealed that software engineers believe that data privacy compliance is not their responsibility. [26] analyzed the reasons why software products failed to comply with laws and purposed a new software quality, called Legal Accountability, which can be evaluated alongside other qualities, such as usability, modifiability, performance, and testing. [46] conducted a review of best practices in system development and presented a new methodology that combines and enhances these practices. [60] revealed that during incident response, legal professionals' advice leads to lessons lost rather than lessons learned, such as by obfuscating root causes and corrective actions that could have prevented the incident.

To the best of our knowledge, limited work has been conducted on legal professionals' perceptions of privacy legal compliance. The closest work to our research is [39], which explores legal and engineering practitioners' understandings of the concept of *technical measures*, i.e., the ways in which technology can be used for data protection and compliance. Different from this work focusing on *technical measures*, our study aims at understanding legal professionals' procedures and strategies for privacy risk assessment and legal compliance analysis.

8 CONCLUSION

Our study focused on the privacy risk assessment and privacy legal compliance in incident response, which has become increasingly important due to rising data breach incidents. We conducted a survey and semi-structured interviews with 33 legal professionals with varying technical/legal expertise, educational backgrounds, and industry experiences. We aimed to answer three research questions. For **RQ1**, we found a significant discrepancy (observed proportion of agreement of 0.546 and 0.844 in incident reports and news articles) in the labeling of privacy-related items and identified the criteria used by legal professionals to identify privacy items. For **RQ2**, we observed a significant disagreement in legal assessment results and revealed three challenges in the legal compliance process: gaps in data definition across various sources, gaps in technical and legal context, and varying levels of expertise. For **RQ3**, we identified three low-quality features in incident response that hinder legal assessments, two high-quality features that facilitate legal assessments, and summarized recommendations from participants on how to improve the quality of the responses. In addition, we discussed the challenges involved in implementing privacy-compliant incident response procedures and suggested appropriate interdisciplinary training for legal professionals to enhance their expertise in both law and engineering fields. We also provided recommendations for improving incident response.

ACKNOWLEDGEMENT

We would like to thank the anonymous shepherd and reviewers for their insightful comments. This work was supported in part by the National Science Foundation (CNS-2343618) and Luddy Faculty Fellowship.

REFERENCES

[1] 2019 Capital One Cyber Incident | What Happened | Capital One. https://www.capitalone.com/digital/facts2019/.
[2] Cisa cybersecurity alerts & advisories.
[3] Codebook. https://docs.google.com/spreadsheets/d/1UiZJGsbFU0GPUvPkjWtjMGK4rxGeaE92/edit?usp=drive_link.
[4] Contact Survey Questions. https://docs.google.com/document/d/126W6CDR4bkECknHC7Nvg8Pu8QdyswEFyc8nGf2pnxxY/edit?usp=drive_link.
[5] GDPR Enforcement Tracker - list of GDPR fines. https://www.enforcementtracker.com.
[6] Informed Consent Statement. https://docs.google.com/document/d/17Ze5ryrX1WcljYyRArz_XBDVTijFvmXVo9oudfOurFk/edit?usp=drive_link.
[7] Inter-annotator Agreement. https://docs.google.com/document/d/1C-BbIZo6vyjgybfoU5tka1LtXnnGAuuDfeOqdXrefdk/edit?usp=drive_link.

[8] Interview Script. https://docs.google.com/document/d/1Tdpl4Wg9jhVgEY_OEgjneeyAJ1CJha6arOETgjjPUq4/edit?usp=sharing.

[9] National Initiative for Cybersecurity Careers and Studies Workforce Framework for Cybersecurity (NICE Framework). https://niccs.cisa.gov/workforce-development/nice-framework.

[10] NIST. https://nvlpubs.nist.gov/nistpubs/specialpublications/nist.sp.800-61r2.pdf.

[11] Participants Demographics. https://docs.google.com/document/d/1ulB7LnEWnlG9LwoGhw69GxWFTlGtbdfB/edit?usp=drive_link.

[12] Post Interview Questions. https://docs.google.com/document/d/11xe0FPQ_13NmGoLDP8KtjTqueeP1pXv8n6dFgdZVcko/edit?usp=drive_link.

[13] Recruitment Advertisements. https://docs.google.com/document/d/16xRREevCrLrhTpLTGciLUXrHmMiGeze4L6oc1JF3tVE/edit?usp=drive_link.

[14] Screening Interview Questions. https://docs.google.com/document/d/1c7uTArOVX12V0r7cuYXExOFUIeQDs_3lf5MvbSjijoo/edit?usp=drive_link.

[15] Survey Questions. https://docs.google.com/document/d/1n8f1p1atPW04HqLWhsra-7VOm0yYGDMKA6tRtYriN24/edit?usp=sharing.

[16] Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. *The New York Times*, 2018. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.

[17] Equifax Data Breach Settlement 2019. *Federal Trade Commission*, Jul 2019. https://www.ftc.gov/enforcement/refunds/equifax-data-breach-settlement.

[18] App privacy details on the App Store, 2021. https://developer.apple.com/app-store/app-privacy-details/.

[19] Data privacy day at Apple: Improving transparency and empowering users, 2021. https://www.apple.com/newsroom/2021/01/data-privacy-day-at-apple-improving-transparency-and-empowering-users/.

[20] FTC Report to Congress on Privacy and Security. *Federal Trade Commission*, Oct 2021. https://www.ftc.gov/reports/ftc-report-congress-privacy-security.

[21] Abdulrahman Alhazmi and Nalin Arachchilage. I'm all ears! Listening to software developers on putting GDPR principles into software development practice. *Personal and Ubiquitous Computing*, 25:1–14, 10 2021.

[22] Micah Altman, Aloni Cohen, Kobbi Nissim, and Alexandra Wood. What a hybrid legal-technical analysis teaches us about privacy regulation: The case of singling out. *Boston University Journal of Science and Technology Law*, 27(1):1–63, 2021.

[23] Benjami Andow, Samin Yaseer Mahmud, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Serge Egelman. Actions speak louder than words: Entity-sensitive privacy policy and data flow analysis with policheck. In *Proceedings of the 29th USENIX Security Symposium (USENIX Security'20)*, 2020.

[24] Benjamin Andow, Samin Yaseer Mahmud, Wenyu Wang, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Tao Xie. Policylint: Investigating internal privacy policy contradictions on google play. In *USENIX Security Symposium*, pages 585–602, 2019.

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

[26] Travis D Breaux and Thomas Norton. Legal accountability as software quality: A us data processing perspective. In *2022 IEEE 30th International Requirements Engineering Conference (RE)*, pages 101–113. IEEE, 2022.

[27] Travis D Breaux and Ashwini Rao. Formal analysis of privacy requirements specifications for multi-tier applications. In *2013 21st IEEE International Requirements Engineering Conference (RE)*, pages 14–23. IEEE, 2013.

[28] Duc Bui, Yuan Yao, Kang G Shin, Jong-Min Choi, and Junbum Shin. Consistency analysis of data-usage purposes in mobile apps. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, pages 2824–2843, 2021.

[29] Karen A. Campbell, Elizabeth Orr, Pamela Durepos, Linda Nguyen, Lin Li, Carly Whitmore, Paige Gehrke, Leslie Graham, and Susan M. Jack. Reflexive Thematic Analysis for Applied Qualitative Health Research. *The Qualitative Report*, 26(6):2011–2028, 06 2021.

[30] Yanto Chandra and Liang Shang. *Inductive Coding*, page 91–106. Springer Nature Singapore, Singapore, 2019.

[31] Yi Chen, Mingming Zha, Nan Zhang, Dandan Xu, Qianqian Zhao, Xuan Feng, Kan Yuan, Fnu Suya, Yuan Tian, and Kai Chen. Demystifying hidden privacy settings in mobile apps. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 570–586. IEEE, 2019.

[32] Andrew Clearwater and J. Trevor Hughes. In the Beginning - An Early History of the Privacy Profession Symposium: The Second Wave of Global Privacy Protection. *Ohio State Law Journal*, 74(6):897–922, 2013.

[33] Jacob Cohen. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4):213, 1968.

[34] Michael Colesky, Jaap-Henk Hoepman, and Christiaan Hillen. A Critical Analysis of Privacy Design Strategies. In *2016 IEEE Security and Privacy Workshops (SPW)*, pages 33–40, 2016.

[35] United States. Federal Trade Commission. *Privacy online: a report to Congress*. The Commission, 1998.

[36] Soteris Demetriou, Whitney Merrill, et al. Free for all! assessing user data exposure to advertising libraries on android. In *NDSS*, 2016.

[37] EU. General Data Protection Regulation (EU) 2016/679. *Official Journal of the European Union*, 2016.

[38] Seda Gürses and Jose M Del Alamo. Privacy engineering: Shaping an emerging field of research and practice. *IEEE Security & Privacy*, 14(2):40–46, 2016.

[39] Oleksandra Klymenko, Oleksandr Kosenkov, Stephen Meisenbacher, Parisa Elahidoost, Daniel Mendez, and Florian Matthes. Understanding the implementation of technical measures in the process of data privacy compliance: A qualitative study. In *Proceedings of the 16th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement*, ESEM '22, page 261–271, New York, NY, USA, 2022. Association for Computing Machinery.

[40] Simon Koch, Malte Wessels, Benjamin Altpeter, Madita Olvermann, and Martin Johns. Keeping privacy labels honest. *Proceedings on Privacy Enhancing Technologies*, 4:486–506, 2022.

[41] Patricia Bazeley Kristi Jackson. *Qualitative Data Analysis with NVivo*. SAGE Publications Inc, 1981.

[42] Aaron Massey, Richard Rutledge, Annie Antón, and Peter Swire. Identifying and classifying ambiguity for regulatory requirements. pages 83–92, 08 2014.

[43] Jeremy C. Maxwell, Annie I. Antòn, and Julie B. Earp. An empirical investigation of software engineers' ability to classify legal cross-references. In *2013 21st IEEE International Requirements Engineering Conference (RE)*, pages 24–31, 2013.

[44] Yuhong Nan, Min Yang, Zhemin Yang, Shunfan Zhou, Guofei Gu, and XiaoFeng Wang. Uipicker: User-input privacy identification in mobile applications. In *24th USENIX Security Symposium (USENIX Security 15)*, pages 993–1008, 2015.

[45] Yuhong Nan, Zhemin Yang, Xiaofeng Wang, Yuan Zhang, Donglai Zhu, and Min Yang. Finding clues for your secrets: Semantics-driven, learning-based privacy discovery in mobile apps. In *NDSS*, 2018.

[46] Nicolás Notario, Alberto Crespo, Yod-Samuel Martín, Jose M Del Alamo, Daniel Le Métayer, Thibaud Antignac, Antonio Kung, Inga Kroener, and David Wright. Pripare: integrating privacy best practices into a privacy engineering methodology. In *2015 IEEE Security and Privacy Workshops*, pages 151–158. IEEE, 2015.

[47] Nicolas Notario, Alberto Crespo, Yod-Samuel Martin, Jose M. Del Alamo, Daniel Le Metayer, Thibaud Antignac, Antonio Kung, Inga Kroener, and David Wright. PRIPARE: Integrating Privacy Best Practices into a Privacy Engineering Methodology. In *2015 IEEE Security and Privacy Workshops*, pages 151–158, 2015.

[48] CHAPTER I GENERAL PROVISIONS. Directive 95/46/ec of the european parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal L*, 281(23/11):0031–0050, 1995.

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

[50] Johnny Ryan. Europe's enforcement paralysis: ICCL's 2021 GDPR report. *Irish Council for Civil Liberties*, Sep 2021. https://www.iccl.ie/digital-data/2021-gdpr-report/.

[51] Natasha Singer and Kate Conger. Google Is Fined $170 Million for Violating Children's Privacy on YouTube. *The New York Times*, Sep 2019.

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

[53] Mark Smith and Jackquelyn Palmer. ANALYSIS: Three Years Later, GDPR Compliance Still a Challenge. *Bloomberg Law*, 2021. https://news.bloomberglaw.com/bloomberg-law-analysis/analysis-three-years-later-gdpr-compliance\-still-a-challenge.

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

[55] Kurt Thomas, Elie Bursztein, et al. Ad injection at scale: Assessing deceptive advertisement modifications. In *2015 IEEE Symposium on Security and Privacy*, pages 151–167. IEEE, 2015.

[56] Muhammad Usman, Michael Felderer, Michael Unterkalmsteiner, Eriks Klotins, Daniel Mendez, and Emil Alégroth. Compliance requirements in large-scale software development: An industrial case study. In Maurizio Morisio, Marco Torchiano, and Andreas Jedlitschka, editors, *Product-Focused Software Process Improvement*, page 385–401, Cham, 2020. Springer International Publishing.

[57] Jice Wang, Yue Xiao, Xueqiang Wang, Yuhong Nan, Luyi Xing, Xiaojing Liao, JinWei Dong, Nicolas Serrano, Haoran Lu, XiaoFeng Wang, et al. Understanding malicious cross-library data harvesting on android. In *USENIX Security Symposium*, pages 4133–4150, 2021.

[58] Xiaoyin Wang, Xue Qin, Mitra Bokaei Hosseini, Rocky Slavin, Travis D Breaux, and Jianwei Niu. Guileak: Tracing privacy policy claims on user input data for android applications. In *Proceedings of the 40th International Conference on Software Engineering*, pages 37–47, 2018.

[59] Josephine Wolff and Nicole Atallah. Early GDPR penalties: Analysis of implementation and fines through May 2020. *Journal of Information Policy*, 11:63–103, 2021.

[60] Daniel W Woods, Rainer Böhme, Josephine Wolff, and Daniel Schwarcz. Lessons lost: Incident response in the age of cyber insurance and breach attorneys. In

32nd USENIX Security Symposium (USENIX Security 23), pages 2259–2273, 2023.
[61] Daniel W. Woods and Rainer Böhme. Incident response as a lawyers' service. *IEEE Security & Privacy*, 20(2):68–74, 2022.
[62] Yue Xiao, Zhengyi Li, Yue Qin, Jiale Guan, Xiaolong Bai, Xiaojing Liao, and Luyi Xing. Lalaine: Measuring and characterizing non-compliance of apple privacy labels at scale. *arXiv preprint arXiv:2206.06274*, 2022.
[63] Le Yu, Xiapu Luo, Xule Liu, and Tao Zhang. Can we trust the privacy policies of android apps? In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 538–549. IEEE, 2016.
[64] Sebastian Zimmeck, Peter Story, Daniel Smullen, Abhilasha Ravichander, Ziqi Wang, Joel R Reidenberg, N Cameron Russell, and Norman Sadeh. Maps: Scaling privacy compliance analysis to a million apps. *Proc. Priv. Enhancing Tech.*, 2019:66, 2019.
[65] Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven Bellovin, and Joel Reidenberg. Automated analysis of privacy requirements for mobile apps. In *2016 AAAI Fall Symposium Series*, 2016.
[66] Christian Zimmermann. Automation potentials in privacy engineering. 05 2020.

A APPENDIX

A.1 Recruitment Advertisements

Please see [13].

A.2 Contact Survey Questions

Please see [4].

A.3 Screening Interview Questions

Please see [14].

A.4 Informed Consent Statement

Please see [6].

A.5 Participants Demographics

The demographic information of participants are presented in [11].

A.6 Codebook

The codebook is presented in [3].

A.7 Post Interview Questions

Please see [12].

A.8 Inter-annotator Agreement

The inter-annotator agreement regarding privacy item recognition and privacy compliance of each report is shown in [7].