GDPR-Relevant Privacy Concerns in Mobile Apps Research: A Systematic Literature Review

Orlando Amaral Cejas[1], Nicolas Sannier[1], Sallam Abualhaija[1],
Marcello Ceci[1], and Domenico Bianculli[1]
[1]SnT - University of Luxembourg
firstname.lastname@uni.lu

November 28, 2024

Abstract

The General Data Protection Regulation (GDPR) is considered as the benchmark in the European Union (EU) for privacy and data protection standards. Since before its entry into force in 2018, substantial research has been conducted in the requirements engineering (RE) literature investigating the elicitation, representation and verification of privacy requirements in GDPR. Software systems deployed anywhere in the world must comply with GDPR as long as they handle personal data of EU residents. Mobile applications (apps) are no different in that regard. With the growing pervasiveness of mobile apps and their increasing demand for personal data, privacy concerns have acquired further interest within the software engineering (SE) community at large. Despite the extensive literature on GDPR-relevant privacy concerns in mobile apps, there is no secondary study that describes, analyzes, and categorizes the current focus. Research gaps and persistent challenges are thus left unnoticed.

In this article, we aim to systematically review existing primary studies highlighting various GDPR concepts and how these concepts are addressed in mobile apps research. The objective is to reconcile the existing work on GDPR in the RE literature with the research on GDPR-related privacy concepts in mobile apps in the SE literature.

Our findings show that the current research landscape reflects a rather shallow understanding of GDPR requirements. The GDPR concepts investigated in the majority of the studies include: (i) the sharing of personal data with third-party libraries, mainly for the purpose of identifying data leaks; (ii) different mechanisms for acquiring explicit consent from users; and (iii) data collection involving various personal data categories that are often obtained directly from the users. While such GDPR concepts are indeed of significant importance, other topics such as data subject rights (i.e., the rights of individuals over their personal data) are fundamental to GDPR, yet under-explored in the literature. In this article, we highlight future directions to be pursued by the SE community for supporting the development of GDPR-compliant mobile apps.

Keywords: Systematic Literature Review (SLR), Requirements Engineering (RE), Regulatory Compliance, The General Data Protection Regulation (GDPR), Privacy Requirements, Mobile Apps.

1 Introduction

With the widespread adoption of mobile applications (apps) in various domains, e.g., language learning [1], healthcare [2–4], and finance [5], regular sharing of personal data has become the norm. Individuals use mobile apps but have concerns and expectations on how their personal data is handled, i.e., collected, processed, or shared. Mobile app development companies, on the other hand, aspire to meet their users' expectations by providing effective data protection mechanisms and fulfilling their commitments, as agreed upon in their privacy policies. In a privacy policy, a user is informed of (and agrees on) a commitment by the company to adopt privacy-related measures, including data collection and processing details [6].

In response to growing privacy concerns, regulations—such as the General Data Protection Regulation (GDPR) [7] issued by the European Union (EU)—introduce various privacy requirements that should, if implemented appropriately, guarantee the protection of personal data of individuals throughout the data processing chain. In relation to mobile apps, GDPR imposes obligations onto development companies, whether EU-based or not, as long as they would collect or process personal data of EU residents in any way. GDPR further levies hefty fines on companies that violate such obligations.

Requirements engineering (RE) is an essential step in software engineering (SE), involving the elicitation, representation and verification of software requirements [8]. To develop legally-compliant software, requirement engineers must specify compliance requirements according to the applicable legal sources.

The RE community has a long-standing interest to elicit and specify requirements pertinent to privacy, as well as to analyze relevant legal documents such as privacy policies [2, 9–13].

Being a fundamental legal source for privacy requirements since even before its entry into force, GDPR has been extensively investigated in both the legal and technological domains. The RE literature has contributed to this investigation with work addressing a wide variety of GDPR-related challenges, including modeling (or representing) legal requirements [14, 15], categorizing and analyzing legal documents [6, 16–18], investigating the users' awareness of privacy [19, 20], and assessing privacy concerns in software applications [21, 22].

Privacy concerns in mobile apps research have also been investigated to a large extent [12, 23]. By downloading a mobile app, users are automatically sharing data of various forms, e.g., the "click to download" button may contribute not only to the total number of downloads shown as statistical metadata in the app, but also to capturing the user's preferences, to be further used by, e.g., recommender systems. This data collection procedure can pose a threat to the user's privacy, and is therefore prohibited unless the user explicitly agrees to it. Mobile apps that do not appropriately implement (i.e., are non-compliant to) GDPR privacy requirements are more prone to posing this threat. Several studies in the literature [10,11, 13, 24, 25] have acknowledged the likelihood of such a privacy threat, highlighting how mitigating privacy threats in mobile apps requires more in-depth understanding of GDPR.

While GDPR privacy requirements have been extensively studied in RE literature and SE at large, there is no secondary study that gathers and consolidates the wide spectrum of efforts conducted in the mobile apps research, studied through the lens of GDPR-relevant privacy concepts.

In this article, we perform a systematic literature review (SLR) and provide a comprehensive overview of the research on privacy in mobile apps, with an emphasis on GDPR. To do so, we analyzed a total of 60 primary studies vis-à-vis an existing comprehensive conceptual model which characterizes all possible privacy-relevant information types according to the GDPR provisions. The conceptual model, proposed by Amaral et al. [14], consists of 56 information types pertinent to GDPR privacy policies. Examples of information types include: *DATA SUBJECT RIGHT*, referring to the individuals' rights over their personal data; *LEGAL BASIS*, referring to the legal basis under which personal data is collected (e.g., through explicit consent obtained from individuals); and *TRANSFER OUTSIDE EUROPE*, encapsulating the necessary mechanisms required for transferring personal data outside the EU.

In this SLR, we categorize existing work in the SE literature according to the information types presented in Amaral et al.'s conceptual model. By doing so, we identify the main topics highlighted in the research as well as the research gaps that should be further investigated by the community.

Our study focuses exclusively on GDPR. We thus restrict the time span of the primary studies to 2016–2023, considering that discussions about GDPR commenced in 2016, before it came into force in 2018. Prior to 2016, personal data protection in the EU was based on national transpositions of the 1995 Directive on personal data protection [26]. We cover work across different domains and venues, targeting research studies that investigate GDPR-relevant privacy concerns in mobile apps. Our search process resulted in 484 primary studies that were reduced to 60 studies [S1–S60], those deemed relevant to our analysis. To conduct our review, we extracted information from all relevant papers, concerning eight categories, namely (1) the GDPR concepts and principles covered in the primary studies; (2) the traced personal data; (3) the targeted app stores and operating systems; (4) the goals and contributions of the primary studies; (5) the applied data analysis methods; (6) the proposed solutions and employed technologies; (7) the data used; and (8) whether any material is made publicly available.

Our results indicate that existing work focuses on a subset of the GDPR-relevant privacy concerns in mobile apps. Specifically, the following GDPR concerns have been prominently investigated:

1. personal data sharing with third-party libraries, mainly for the purpose of identifying data leaks;
2. mechanisms for obtaining explicit consent from users;
3. data collection involving various personal data categories that are often collected directly from the users.

While such GDPR concepts are indeed important for SE, other fundamental concepts in GDPR remain under-explored in the literature, e.g., the data subject rights describing the rights of individuals over their personal data. In this work, we highlight future directions to be pursued by the SE community for supporting the development of GDPR-compliant mobile apps.

Contributions. The paper makes the following contributions:

(1) We present a meta-study that systematically surveys the existing work on mobile apps research, investigating various GDPR-relevant privacy concepts. We conduct our systematic literature review following the best practices and guidelines in the literature [27,28]. Our study collects and consolidates the different research contributions in the SE literature. We analyze the different primary studies according

to an existing comprehensive conceptual model that characterizes the information types pertinent to GDPR privacy requirements.

(2) We identify the research gaps and discuss future research directions to advance the SE research on GDPR privacy in mobile apps. By projecting existing work against the same conceptual model, we categorize the research goals and identify the privacy concepts that, despite their importance for GDPR compliance, received little attention from the SE community. With the rapid growth of the mobile apps market, it has become essential to investigate methods for developing GDPR-compliant apps and checking their compliance against GDPR.

Structure. The remainder of this paper is structured as follows: Section 2 provides background information. Section 3 surveys the state of the art. Section 4 presents the design of our systematic literature review. Section 5 reports on our findings. Section 6 discusses threats to validity. Section 7 concludes the paper.

2 Background

GDPR has been widely investigated by the RE community. For our SLR, we leverage the existing work that focuses on representing the GDPR-related privacy concepts. These concepts are mainly derived from the legal provisions of Chapters III-V of GDPR, and must be distinguished from the principles stated in the preamble and in Chapter II (e.g., the principle of "data minimisation" in preamble 156 and Art. 5(1)(c)). While principles are important in legal procedure and literature, compliance of software is rather ensured by extracting requirements from the directly applicable parts of regulations: therefore, *this research sees privacy concerns through the lens of the information types representing (privacy) concepts that are relevant to legal requirements as explicitly stipulated in GDPR.* In that regard, several representations have been proposed in the RE literature (see, for example, [29–31]).

In this SLR, we utilize the conceptual model proposed by Amaral et al. [14] as a basis for reviewing the primary studies. The model characterizes GDPR requirements directed at data controllers (i.e., those determining the purposes and means of the processing of personal data) and pertinent to privacy policies. Privacy policies of mobile apps must be provided by the software house developing the app and contain legally-binding privacy requirements that are customized to specific apps' particularities. Since such policies are at the front end between users and apps, we believe that the conceptual model is a good instrument against which to analyze the existing work. Selecting this conceptual model is driven by the following reasons: first, the model is, to our knowledge, the most comprehensive in the RE literature and hence it provides us with a more complete view of what GDPR requires. Second, the model focuses on GDPR requirements specifically for privacy policies—the technical documents that are most relevant for mobile apps. Privacy policies can be used for eliciting privacy requirements with which the mobile app should comply [14]. Third, the model has been created in close collaboration with legal experts, and it thus integrates both the legal and requirements engineering perspectives.

The conceptual model, hereafter referred to as *reference model*, is depicted in Fig. 1. The model is composed of 56 information types organized in three hierarchical levels. Level 1, shaded teal-blue (solid-border boxes), captures high-level concepts such as *DATA SUBJECT RIGHT* that is mentioned earlier. Level 2, shaded light green (dashed-border boxes), and level 3, shaded gray (dotted-border boxes), capture the different specializations. For instance, *DATA SUBJECT RIGHT* has been refined into eight specializations, one for each of the individuals' rights, such as *ACCESS*, i.e., the right to access personal data. For more details, we refer the reader to the original paper [14].

3 State of the art

Despite the significant impact of GDPR on mobile app development and the risks for non-compliance, with hefty fines being imposed on major mobile apps development companies violating GDPR, there is no study to date that comprehensively reviews the research landscape on how GDPR-relevant privacy concerns are addressed in mobile apps. Below, we discuss secondary studies that survey relevant topics.

Fernández-Alemán et al. [2] studied security and privacy concerns in electronic health records (EHR) systems. Specifically, the authors analyzed 49 papers, of which only 26 use standards or regulations regarding privacy and security of EHR data. The authors reported the Health Insurance Portability and Accountability Act (HIPAA) [32] and GDPR [7] as the most widely used regulations. The authors concluded that, to implement secure EHR systems, extensive harmonization is first necessary to resolve discrepancies between the set of applicable regulatory texts and standards.

Similarly, Iwaya et al. [12] analyzed 52 papers to study the state-of-the-art on security and privacy in mobile health (mHealth) and ubiquitous health (uHealth) systems. Results suggested that existing work primarily focused on certain aspects that were not necessarily critical to the organizations employing

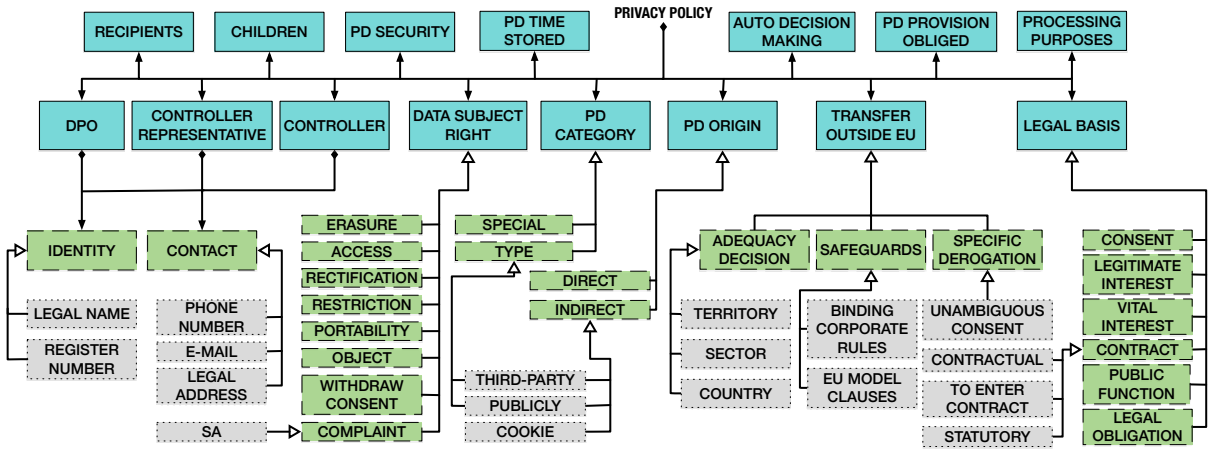

Figure 1: Overview of the *reference model* [14] used in this SLR for identifying GDPR-related privacy concepts

mHealth and uHealth systems, leaving out aspects such as data governance, security and privacy policies, or program management.

Martínez-Pérez et al. [10] surveyed privacy and security in mHealth apps. Among the 169 analyzed papers, the authors highlighted the lack of awareness about security and privacy laws and further presented some recommendations to serve as a quick guide for designers, developers, and researchers. Such recommendations provide a more detailed description of the legal text, facilitating compliance to the current security and privacy standards.

In a systematic mapping study, Morales-Trujillo et al. [11] analyzed *privacy by design (PbD)*, one of GDPR's main principles, in software systems. Their work aimed to identify relevant literature collecting PbD goals (e.g., data protection mechanisms) in software development and determining the extent to which PbD was present in the current software development practices. The authors analyzed 49 papers and concluded that the primary studies mainly focus on data minimization. The authors further suggested that PbD was still an underdeveloped concept in the SE literature and acknowledged the necessity for a framework to support the development of privacy-aware systems.

In a similar vein, Semantha et al. [24] analyzed PbD in the healthcare domain. Their work focused on producing recommendations to reduce personal data breaches. The authors targeted contemporary frameworks regularly applied for safeguarding data privacy; specifically, they examined seven PbD frameworks and identified key limitations in these frameworks that would allow for potential data breaches, e.g., inefficiencies in data managing jeopardizing the reliability of personal data. Finally, the authors advocated for refining and improving these frameworks as a way to reduce the rate of data breaches particularly in the healthcare domain.

More recently, Andrade et al. [13] investigated the misuse of personal data and what this entails in terms of individuals' privacy and software product's quality. The authors analyzed 75 papers to understand how PbD principles had been applied in the SE practices. The results highlighted the lack of specific methodology and tools for translating these principles into practical activities throughout the software development lifecycle. This concern has become even more relevant now that PbD is part of regulations such as GDPR.

More related to our work, Alloghani et al. [23] focused on identifying security and privacy issues in mobile devices and systems alongside existing methods for detecting and preventing such issues. The authors emphasized the importance of utilizing preventive, detective, and responsive methods for ensuring security and privacy in mobile applications. Such methods require the involvement of both mobile vendors and service providers.

Shrivastava et al. [25] analyzed privacy issues concerning permission requests in Android apps. The authors studied 110 research papers, identifying possible severe security implications of the Android permission protocol and further describing several limitations in permission checks.

Ebrahimi et al. [33] analyzed 59 papers related to privacy in mobile apps. They focused on the practice of obtaining explicit consent prior to accessing sensitive private information, in exchange for offering a more personalized user experience. The authors argue that such privacy-invading practices have led to major privacy concerns among app users. The authors also showed that existing literature had focused mainly on detecting data leaks, with little to no attention being given to the user's perspective.

Negri-Ribalta et al. [34] surveyed the impact of GDPR's on regulatory requirements pertinent to data

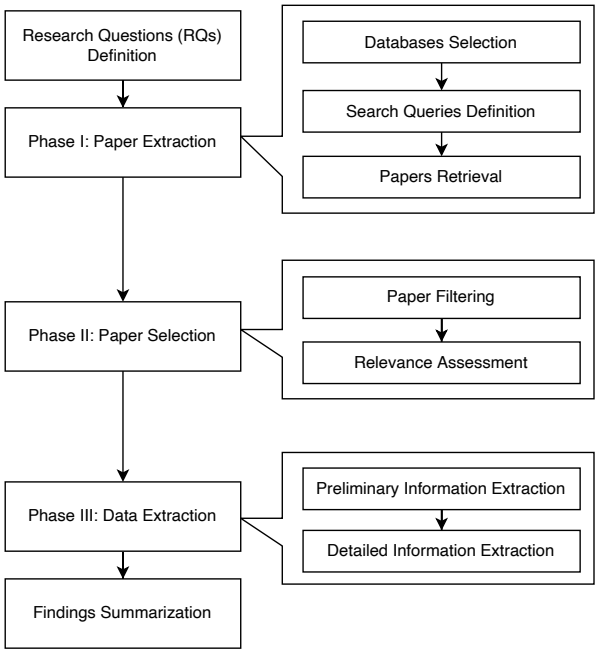

Figure 2: Overview of our Review Process

protection and how they are handled from an RE perspective. Their study involved 90 papers published between 2016 (i.e., right about when the regulation was voted) and 2022 (four years after GDPR was actually enforced). The work provides a generic view of the GDPR provisions. For the analysis, the authors created a taxonomy of GDPR privacy concerns according to an ad-hoc manual analysis of the regulation. In comparison, our work is driven by a comprehensive and detailed conceptualization of privacy-relevant concerns in GDPR. Mapping existing studies in the literature against this conceptual model provides guidance for future comparisons.

To conclude, in contrast with the aforementioned secondary studies, in this SLR we present a comprehensive overview of the existing research landscape in SE, focusing on privacy in mobile apps. Our SLR further projects the current research onto a *reference model* of GDPR privacy concepts.

4 Study Design

We conducted this SLR following the recommended guidelines by Kitchenham and Charters [28, 35, 36]. Fig. 2 shows the different steps that we followed to achieve our findings. In particular, our method spanned three phases, namely (i) searching for the literature (i.e., paper extraction), (ii) selecting the papers to analyze (i.e., paper selection), and (iii) carrying out the data extraction. To this end, the first author defined the review protocol, performed the search and selection of relevant papers, and performed the data extraction step. This process was closely supervised by the second and third authors. Findings were thoroughly discussed and validated across several sessions.

4.1 Research Questions (RQs) Definition

This study investigates four research questions (RQs):

RQ1. How well does the literature on privacy in mobile apps cover the GDPR privacy metadata types? This RQ explores the research landscape on mobile apps to identify the main topics in terms of GDPR privacy requirements. Drawing on the *reference model* (presented in Section 2), we answer RQ1 by mapping the topics investigated in existing work on mobile apps to the metadata types in the conceptual model. We further outline the GDPR principles that are investigated in the literature but not represented in the conceptual model.

RQ2. What are the main objectives pursued when investigating GDPR privacy concepts? RQ2 explores the main objectives of primary studies that investigated GDPR privacy in mobile apps. The intuition behind this RQ is to understand the *SE research problems* related to GDPR that were acknowledged or addressed by existing work. This helps to understand which problems or GDPR privacy metadata remain under-investigated, leading to potentially open directions for future research.

RQ3. What are the different types of research contribution? RQ3 explores the different types of research contributions made by the primary studies, distinguishing between *technical* and *non-technical* contributions, and analyzing the app stores and operating systems targeted in the literature. It provides a comprehensive view on the research contributions and further describes, in the case of technical solutions, the enabling technologies and the way such contributions were validated.
RQ4. Are the research artifacts accompanying studies on GDPR privacy concepts publicly available? This RQ sheds light on the publicly available artifacts that can be potentially leveraged in future research. RQ4 further reports on the licenses under which these artifacts have been shared.

4.2 Phase I: Paper Extraction

Ensuring the retrieval of a relatively representative sample of the existing literature is essential for conducting an SLR. This phase involves three steps: (1) selecting the right databases or search engines to query, (2) appropriately refining the query keywords, and (3) expanding our results by retrieving additional papers from indexing systems. Next, we elaborate on each step.

4.2.1 Databases Selection

For the search of papers in the different databases, we adopted common practices for and recommendations from the SLR literature [36–38]. Specifically, we retrieved primary studies from the following databases: ACM Digital Library[1], IEEE Xplore[2], ScienceDirect[3], Scopus[4], and SpringerLink[5]. These sources are typically selected in SLRs and systematic mapping studies [39, 40]. Additionally, we searched for papers in the following search engines or indexing systems: DBLP[6], Google Scholar[7], and Web of Science[8]. Our motivation to include these additional sources is driven by two reasons. First, we enlarged our search process to ensure a complete coverage of the SE and RE literature, assuming that these sources likely contain other journals and conference proceedings that might be missing from the regular databases. Second, our paper search using regular sources might miss papers due to being, for instance, not yet in their final publisher's form. We therefore scoped our search against these indexing systems by restricting it to only years 2022 and 2023.

4.2.2 Search Queries Definition

We iteratively refined our search queries by experimenting with several terms targeting papers that address *GDPR privacy in mobile apps*.

Initially, we tested the following query, which combines several keywords: *("mobile app*" AND "data protection" AND "data privacy")*. We then created variants where we used single terms as keywords, e.g., *("mobile" AND "app*")* instead of *("mobile app*")*. We queried all databases with the same pre-defined set of keywords except for ScienceDirect, since it does not allow defining queries with special characters such as wildcards(`*`). In this case, we used the following combination: *("mobile app" OR "mobile apps" OR "mobile application" OR "mobile applications")*.

Following this, we iteratively refined our queries. To ensure that the retrieved papers were within the scope of our study, our final queries comprised the following keywords: *"mobile app"*, *"privacy"*, and *"gdpr"*. The final queries are listed in Table 1.

4.2.3 Papers Retrieval

In this step, we queried the various databases using our defined queries. All queries were applied on the full text of the papers. Additionally, we extracted papers while directly integrating the following inclusion criteria:

I1 *Time span:* To fully capture primary studies focusing on GDPR, we defined the time span 2016–2023 early in our search process. The GDPR was adopted in 2016 after passing European Parliament and entered into force on May 25, 2018. Using this time span, we targeted research work that investigated GDPR not only after its entry into force, but also its final public version shortly before.

[1] `https://dl.acm.org/`
[2] `https://ieeexplore.ieee.org/`
[3] `https://www.sciencedirect.com/`
[4] `https://www.scopus.com/`
[5] `https://link.springer.com/`
[6] `https://dblp.org/`
[7] `https://scholar.google.com/`
[8] `https://www.webofscience.com/wos/`

Table 1: Final Search Queries and Respective Number of Papers Retrieved per Database.

Query	Database	N
"mobile apps" AND "privacy" AND "gdpr"	ACM Digital Library	181
"mobile apps" AND "privacy" AND "gdpr"	IEEE Xplore	70
"mobile apps" AND "privacy" AND "gdpr"	ScienceDirect	129
"mobile apps" AND "privacy" AND "gdpr"	Scopus	26
"mobile apps" AND "privacy" AND "gdpr"	SpringerLink	78
"mobile apps" AND "privacy" AND "gdpr"	Web of Science	17
"mobile apps" AND "privacy" AND "gdpr"	DBLP	15
"mobile apps" AND "privacy" AND "gdpr"	Google Scholar	100
Total	All	616

[*] *N: Number of retrieved papers.*

I2 *Language:* We exclusively considered primary studies written in English.

I3 *Domains and paper types:* We scoped our search to subject area of *Computer Science* OR *Engineering*, and we further limited the search to *Conference papers* OR *Journal Articles*.

Querying the regular databases, as shown in Table 1, resulted in a collection of 484 papers, distributed across the different databases as follows: 181 from ACM Digital Library, 70 from IEEE Xplore, 129 from ScienceDirect, 26 from Scopus, and 78 from SpringerLink.

As noted earlier, we also queried indexing systems for additional potentially missing papers. Consequently, we retrieved 132 additional papers, including 100 papers from Google Scholar, 17 papers from the Web of Science, and 15 papers from DBLP. As for Google Scholar, we considered only the first 10 pages of results, assuming that papers appearing after that will not likely be relevant to the objectives of our study.

Our paper extraction phase resulted in a total of *616 papers*, which were then passed on to the next phase for further analysis.

4.3 Phase II: Paper Selection

In this phase, we selected papers that are relevant for our study, following three steps including: (1) removing duplicate papers and cleaning the results, (2) excluding papers according to certain criteria, and (3) identifying the papers within the scope of our study. We explain these steps below.

4.3.1 Paper filtering

We applied duplicate removal on the entire collection of the 616 papers. To identify duplicates, we defined a case-insensitive *Excel macro* that compares the titles of the papers and determines possible duplicates. The output of this macro indicated which paper metadata details (particularly the title) were actually duplicates. This output was then manually vetted by the first author of this paper. As a result, we filtered out 204 papers which were deemed genuine duplicates, thereby reducing our collection to 412 papers.

To further refine the retrieved primary studies, we applied two *exclusion criteria*, described below.

E1 *Paper length:* We defined a minimum paper size of six pages (including references). The rationale behind this threshold was to discard short papers such as position/vision/demo/new-idea/early-research-achievements papers with only preliminary or partial results, as commonly done in relevant literature [39, 40].

E2 *Publication venues:* Following the guidelines of Kitchenham and Brereton [27], we filtered out papers that were not relevant for our study. Specifically, we excluded papers published in journals or conferences clearly unrelated to SE and security and privacy since privacy has long been studied from the software security perspective. In this way, we removed papers focusing on the social aspects of privacy/data protection or investigating specialized topics. In case of uncertainty (e.g., when a paper title appeared relevant) we retained the venue and related paper(s) for further analysis

Table 2: Journals and Conference Venues with Number of Papers Deemed Relevant in this SLR

ID	Venue	N
J1	Wireless Networks	1
J2	Frontiers of Computer Science	1
J3	Software and Systems Modeling	1
J4	Computing	1
J5	Empirical Software Engineering	2
J6	AI & Society	1
J7	International Journal of Information Security	1
J8	Personal and Ubiquitous Computing	1
J9	Internet Interventions	1
J10	Computers & Security	4
J11	Journal of Systems and Software	1
J12	Information and Software Technology	1
J13	Computer Law & Security Review	4
J14	ACM Transactions on Computer-Human Interaction	7
J15	ACM Transactions on Software Engineering Methodology	1
J16	IEEE Transactions on Software Engineering	2
V1	IEEE International Conference on Software Analysis, Evolution and Reengineering	1
V2	The Web Conference	1
V3	IEEE/ACM International Conference on Automated Software Engineering	5
V4	Workshop on Privacy in the Electronic Society	1
V5	ACM Conference on Security and Privacy in Wireless and Mobile Networks	1
V6	CHI Conference on Human Factors in Computing Systems	1
V7	ACM/IEEE International Conference on Software Engineering	6
V8	ACM Web Conference	1
V9	Annual Computer Security Applications Conference	1
V10	International Conference on Availability, Reliability and Security	1
V11	ACM SIGSAC Conference on Computer and Communications Security	2
V12	Annual IEEE/IFIP International Conference on Dependable Systems and Network	1
V13	USENIX Security Symposium	2
V14	IEEE International Requirements Engineering Conference	3
V15	IEEE European Symposium on Securityand Privacy Workshops	2
V16	IEEE Annual Computer Software and Applications Conference	1
	Total	60

[*] *N: Number of papers deemed relevant in this SLR.*

(see below). Examples of excluded venues are the *International Journal of Medical Informatics*, *Blockchain: Research and Applications*, and the *EAI International Conference on Smart Objects and Technologies for Social Good*.

After applying the paper filtering, our collection was reduced to 168 papers which were further processed in the successive step.

4.3.2 Relevance Assessment

In this step, we analyzed the remaining papers and identified those relevant for our SLR. The first three authors collectively participated in this process. Driven by the motivation of our SLR, the relevance of the papers was assessed according to their main focus being privacy. Relevant papers should clearly discuss GDPR-relevant privacy concepts and how these were handled in mobile apps. For instance, a paper that discusses users practices in mobile apps, such as subscription cancellation and asking for refunds, or users providing inaccurate information were considered as not relevant to our SLR. Similarly, papers discussing dark patterns that attempt to manipulate users' behaviors are not relevant to our SLR. While these papers are potentially relevant to the overall privacy concept, they are not particularly informative about how GDPR privacy concepts on handling and processing personal data were addressed in mobile apps. This step spanned three iterations, described below.

In the first iteration, the first author read through all abstracts and labeled the respective papers according to three categories, namely "relevant" when the paper was relevant to our SLR, "not relevant"

when it was not relevant, and "possibly relevant" when the author was in doubt or the abstract was not detailed enough for a conclusive decision. As a result, 52 papers were labeled as "relevant", 94 as "not relevant", and 22 as "possibly relevant".

The second iteration aimed to confirm this labeling, focusing on the 52+94=146 papers deemed as "relevant" or "not relevant". Specifically, we randomly split the papers into two equal subsets, of 73 papers each. Subsequently, the second and third authors independently cross-checked all papers in these two subsets. This validation process confirmed the decisions made in the first iteration. All disagreements were thoroughly discussed by the three authors.

In the last iteration, the first author examined the remaining 22 "possibly relevant" papers by screening the full paper and carefully considering the relevance of its content to the objectives of our SLR. As a result, eight papers were deemed "relevant", and the remaining 14 were deemed "not relevant". The second and third authors confirmed the decisions on the papers in this iteration as well.

This step allowed us to perform a further selection of papers as well as of the relevant journals and venues (which were reduced to 32, equally split—see Table 2). The output of this step is a final set of 60 papers (corresponding to 35.7% of the 168 analyzed papers) on which we conducted our data extraction phase.

4.4 Phase III: Data Extraction

Our data extraction phase targeted certain *relevant information*, listed in Table 3. In addition to the metadata about the paper (prefixed by C0 in the table), e.g., its title and the authors name, we extracted the following information: the GDPR concepts and principles covered in the primary studies (C1), the traced personal data (C2), the targeted app store(s) and operating system(s) (C3), the goals and contributions of the primary studies (C4), the applied data analysis methods (C5), the proposed solutions and their enabling technologies (C6), the data used (C7), and whether any artifact was made publicly available (C8). We extracted this information by thoroughly reviewing the 60 papers marked as relevant in the previous phase. Our data extraction process involved two steps, described below.

4.4.1 Preliminary Information Extraction

In this step, we primarily reviewed the abstract and the first sections (usually called "introduction", "background", and "related work"). The purpose of this step is to extract generic *relevant information* related to the main focus of the primary study. Concretely, we extracted the main goal, the contributions, the research questions, the outcomes, and information on online availability of resources, if applicable. We also identified the covered GDPR principles and privacy concepts according to our *reference model*. We noted that some papers lacked details concerning, e.g., contributions or research questions. When such details were not clearly stated in the analyzed papers, we labeled the respective information as "not stated".

4.4.2 Detailed Information Extraction

In the last step, we analyzed the remaining sections of the papers. We reviewed the details related to the proposed approaches or automated solutions, experimental design, implementation, evaluation, and results. From these sections, we extracted additional *relevant information*. In the cases where a type of *relevant information* was not applicable or not clearly found in the analyzed paper, we labeled it as "not stated".

The extracted *relevant information* from all relevant papers are the basis for our findings and for answering our RQs next.

5 Findings

In this section, we answer the RQs outlined in Section 4.

5.1 RQ1. How well does the literature on privacy in mobile apps cover the GDPR privacy metadata types?

To answer RQ1, we mapped the privacy concepts highlighted in the primary studies to the information types in our *reference model* (see Section 2). This mapping activity involved labeling the primary study with the specific information types, if the study mentions or discusses that information type in any way. Multiple information types can obviously be assigned to a single analyzed study, i.e., the mapping is a multi-labeling task. The mapping was primarily conducted by the first author. To ensure the quality

Table 3: Data Extraction Fields

ID	Label	Description
C0.1	Paper ID	Each paper is assigned a unique identifier.
C0.2	Paper Type	Whether the paper is a conference paper or a journal article.
C0.3	Title	The title of the paper.
C0.4	Authors	The authors names.
C0.5	Venue	The name of the journal or conference where the paper is published.
C0.6	Publication Year	The year when the paper is published.
C0.7	Pages	The number of pages in the paper.
C0.8	Publisher	The name of the publisher.
C1.1	GDPR Principles	Key GDPR principles mentioned in the paper
C1.2	Privacy Concepts	Key GDPR privacy concepts mentioned in the paper according to our *reference model* [§]
C2.1	Personal Data	The list of personal data that is mentioned and which the mobile apps tracks.
C3.1	OS	The operating system (OS) investigated by the paper.
C3.2	App Store	The app store investigated by the paper.
C4.1	Main Goal	The main goal of the paper.
C4.2	Contributions	The key contributions of the paper.
C5.1	Study Type	The study type conducted in the paper.
C5.2	Research Approach	Whether the paper conducts qualitative or quantitative analysis, or a mix of both.
C5.2.1	Participants	The number of participants in the qualitative study.
C6.1	Solution Type	Whether the solution is purely manual, semi-automated, or automated.
C6.2	Enabling Technology	The enabling technologies used to build the proposed solution in the paper.
C6.3	Program Analysis	Whether the paper employs program analysis techniques.
C6.3.1	Method	The concrete program analysis method(s) applied in the paper.
C6.4	Outcome	The final outcome of the paper.
C7.1	Data Source	The source of the data which is used in the paper.
C7.2	Domain	The domain of the data used in the paper.
C7.3	Dataset Size	The size of the dataset or corpus used in the paper.
C7.4	Dataset Categories	The categories of the data and mobile apps collected and/or used in the paper.
C8.1	Availability	Any links provided with the paper with shared material.

[§] See Section 2.

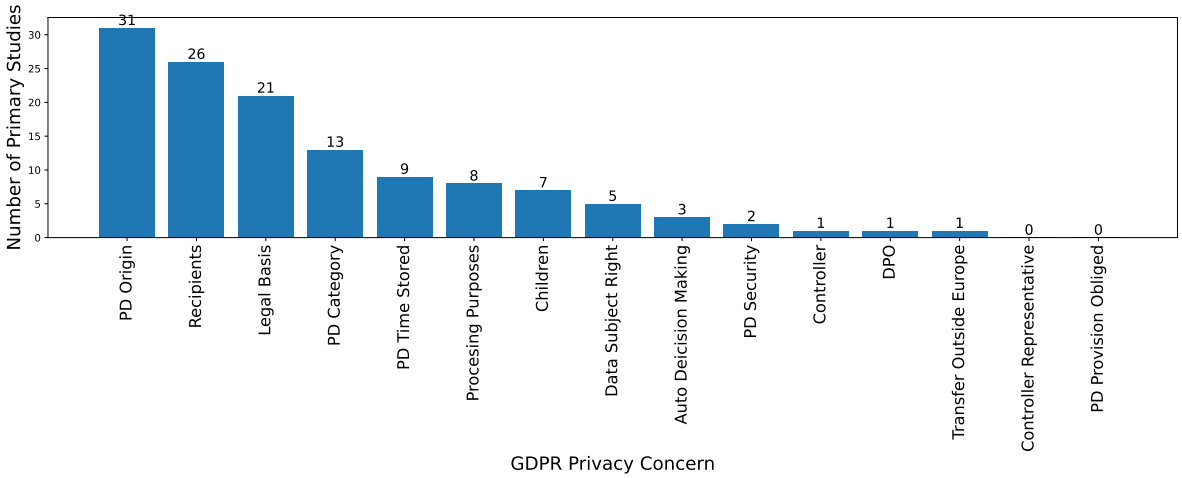

Figure 3: Coverage of GDPR Privacy Concerns in the Literature (RQ1).

of the resulting labeling, the second and third authors independently validated a subset of 10% primary studies each. We then measured the agreement among the information types identified by the first author and the types identified by the second or the third author using the Krippendorff's alpha coefficient [41] over this subset. We obtained an agreement of 0.94, indicating a high agreement.

According to our results, 49 out of 60 papers (≈82%) mention one or more GDPR privacy concepts and 25 papers (≈42%) mention one or more GDPR principles. 51 out of 60 papers (≈85%) have been published after 2020, with half of them published in 2023. These percentages indicate the increase of interest in investigating the GDPR privacy concepts in various SE contexts, as we will show later in this section. Table 4 presents the information types corresponding to the GDPR concepts, alongside the analyzed papers discussing them. Note that the total number of papers exceeds 49, since the same paper could discuss multiple information types. Fig. 3 demonstrates the attention given to these information types by showing the number of papers that mention them.

In our analysis, we observed that some primary studies investigate generic GDPR principles. For example, the *data minimisation* principle refers to collecting only the data necessary to run a service (Article 5(1)(c) of GDPR). We observe that the SE literature often discusses the following GDPR principles: (1) lawfulness, fairness, and transparency; (2) purpose limitation; (3) data minimisation; (4) accuracy; (5) storage limitation; (6) integrity and confidentiality; and (7) accountability [42, 43]. To this end, we note that the *reference model* we use to do this review does not explicitly refer to GDPR principles. Instead, the model focuses on obligations and rights stipulated in the GDPR that are directly relevant to compliance. We still decided to report on GDPR principles in addition to the information types identified according to the *reference model*. Our findings are therefore informative to the research community also with respect to GDPR principles in mobile apps literature.

We observe that the literature is dominated by three information types. The most investigated information type (in 31 primary studies) is the *PERSONAL DATA ORIGIN*, which refers to the personal data collection mechanisms, e.g., whether the personal data is collected through cookies. The second is *RECIPIENTS*, which refers to the sharing of personal data with entities or organizations other than the controller, e.g., with third parties. Finally, *LEGAL BASIS* is another well-investigated information type, where the literature exclusively focus on *CONSENT*.

We note that, while *CONSENT* has a significant value since it authorizes a software system (or mobile app in this case) to process personal data of individuals, other legal bases such as *LEGITIMATE INTEREST* can impact the implementation of certain actions related to users' data. For instance, the deletion of personal data upon the users' request can be postponed or declined because of the *LEGAL OBLIGATION* based on which personal data was collected in the first place. Personal data can further be processed based on lawful *PUBLIC INTEREST* (e.g., public health or historical research purposes).

Another key observation is that, despite the significant attention given to *CONSENT*—since it concerns individuals (i.e., users of a given mobile app)— very little attention has been given to *DATA SUBJECT RIGHT*, concerning individuals' rights over their personal data, another user-centered information type that is highly-relevant for the development of legally-compliant mobile apps. Examples of such rights include *WITHDRAW CONSENT* to enable the user to revoke their consent given in the first place, the right to *ACCESS* where users should be able to access their personal data collected over a certain period of time, or the right to *ERASURE* enabling users to request the deletion of their personal data. Allowing individuals to practice their rights is fundamental according to GDPR. If such legal

Table 4: Mentions of Information Types in the Literature (RQ1)

Metadata	Level[§]	Total	Papers
Auto Decision Making	L1	3	[S12, S14, S44]
Children	L1	7	[S2, S9, S10, S24, S32, S37, S40]
Controller	L1	2	[S30, S40]
Contact	L2	2	[S30, S40]
DPO	L1	1	[S30]
Contact	L2	1	[S30]
Data Subject Right	L1	5	[S37, S40, S44, S53, S56]
Withdraw Consent	L2	3	[S44, S53, S56]
Legal Basis	L1	21	[S1, S5, S7, S11, S14, S22, S32, S34, S38, S39, S41–S45, S48–S50, S53, S56, S58]
Contract	L2	1	[S49]
Public Function	L2	1	[S49]
Legitimate Interest	L2	1	[S49]
Vital Interest	L2	1	[S49]
Legal Obligation	L2	1	[S49]
Consent	L2	21	[S1, S5, S7, S11, S14, S22, S32, S34, S38, S39, S41–S45, S48–S50, S53, S56, S58]
Personal Data Category	L1	13	[S11, S14, S15, S41, S48, S51, S52, S54–S58, S60]
Type	L2	1	[S54]
Personal Data Origin	L1	32	[S2, S3, S7, S11–S15, S17, S23, S28–S31, S34, S35, S37–S41, S43, S44, S48, S51–S54, S56, S58–S60]
Direct	L2	30	[S2, S3, S7, S11–S15, S17, S23, S28–S31, S34, S35, S37–S41, S43, S48, S51–S54, S56, S58–S60]
Third Party	L3	2	[S53, S54]
Cookie	L3	1	[S44]
Personal Data Time Stored	L1	9	[S2, S7, S11, S14, S23, S30, S37, S40, S53]
Personal Data Security	L1	2	[S2, S40]
Processing Purposes	L1	8	[S2, S7, S11, S14, S23, S30, S37, S43]
Recipients	L1	26	[S2, S3, S7, S8, S12–S14, S17, S23, S28, S29, S31, S35, S38–S41, S44, S46–S48, S51–S54, S59]
Transfer outside Europe	L1	1	[S53]
Safeguards	L2	1	[S53]

[§] Level corresponds to the hierarchy in the *reference model* as shown in Fig. 1: L1 is the generic metadata (shaded teal-blue), while L2 (shaded light green) and L3 (shaded gray) provide specializations to L1.

Table 5: Overview of Goals Observed in the Literature (RQ2)

Theme	Total	Papers
T1	8	[S1, S28, S41, S46, S50–S52, S59]
T2	6	[S2, S3, S20, S25, S27, S57]
T3	3	[S9, S24, S32]
T4	11	[S5, S7, S11, S21, S39, S42, S44, S45, S48, S49, S56]
T5	10	[S16, S19, S22, S23, S26, S31, S33, S34, S36, S58]
T6	7	[S18, S29, S35, S37, S38, S55, S60]
T7	7	[S4, S8, S10, S12–S14, S43]
T8	8	[S6, S15, S17, S30, S40, S47, S53, S54]

requirements are well-captured and appropriately implemented in mobile apps, individuals can easily practice their rights through, e.g., direct interaction within the app.

Considering that the most well-covered information type is *PERSONAL DATA ORIGIN*, we looked closely at the types of personal data that are being discussed in the primary studies. Broadly, there are three categories of personal data types. The first category concerns data about individuals, which are not app-dependent, i.e., they do not change depending on which app is being used. Examples in this category include fingerprints, name, phone number, home address, gender, age, bank details, health information, photos, and salary. The data in this case are often collected directly from the user of an app, corresponding therefore to *DIRECT* information type. The second category concerns data about the user of the mobile app, e.g., user name, password, phone calls, SMS, contacts. Such data often varies across different apps, depending on, for example, the domain and the purpose of the app. The last category concerns data related to the mobile device, e.g., geolocation information, IP address, connectivity status, SIM card ID, addresses of nearby hotspots. We believe it is important to make this distinction at the time of developing the app, to have a better understanding of what data will be collected and hence what requirements are necessary to fulfill in order to develop apps that comply with GDPR. We remind the reader that the GDPR imposes obligations for protecting personal data. It requires more strict obligations (e.g., security mechanisms) when sensitive personal data (e.g., health information) is collected compared to less or non sensitive data, such as the user name invented by an individual for using a specific mobile app.

Requirements pertaining to some information types are likely not implementable, i.e., no software component is expected to deal with them. Such information types are expected to receive very little attention in the SE literature, due to the lack of software relevance. However, these information types often entail legal requirements affecting the involved parties in the data collection and processing activities, such as the mobile software development companies. To be fully GDPR compliant, these requirements must be comprehensively specified in legal agreements such as privacy policies or data processing agreements established among the involved parties. Examples of such information types include *PD PROVISION OBLIGED*, referring to the failure of providing some services as a consequence for not providing certain personal data to the controller. Another example is *CONTROLLER REPRESENTATIVE*, which defines a representative entity of the controller when the controller is located outside the EU.

5.2 RQ2. What are the main objectives pursued when investigating GDPR privacy concepts?

RQ2 investigates the main research problems related to GDPR compliance addressed by the SE community. The main objectives observed in our work can be summarized across eight themes, discussed below. Table 5 lists the primary studies under each theme.

5.2.1 T1. Tracing data through SOURCE and SINK methods

This theme is frequent in SE, as primary studies often trace certain personal data (SOURCE) to identify potential data leaks. Identifying which personal data to trace is a relevant research topic.

According to our *reference model*, the categories of personal data being collected (*PERSONAL DATA CATEGORY*) must be clearly specified by the controller and explicitly communicated to the data subject (i.e., the users of these mobile apps) through a privacy policy that must be agreed prior to using the app. In other words, no personal data should be collected without having been declared in the privacy policies upfront.

Fostering this research theme is essential for understanding whether the mobile app is GDPR compliant when collecting personal data from the user. Existing work under this theme focuses on analyzing

source or byte code of mobile apps and does not necessarily investigate privacy policies.

We believe that this theme crosses multiple information types, including *PERSONAL DATA CATEGORY*, *PD ORIGIN*, and *CONSENT*. Sharing data with third-party libraries is a closely related theme; in our *reference model*, this theme concerns the information type *RECIPIENT*. In brief, the GDPR requires all recipients to be listed in the privacy policies of mobile apps. If the app is sharing personal data with third-party libraries, then this has to be fully transparent.

5.2.2 T2. App permission requests

This theme investigates the access to personal data granted to the mobile app through permission requests. As mobile apps are becoming more pervasive, this theme is important for understanding what personal information the mobile app can access or collect indirectly through permission requests, e.g., to access information in other apps. Typical example is when the user ought to share location information to be able to use various apps such as navigation or fitness apps. This theme is relevant to the information types: *PERSONAL DATA SECURITY*, *RECIPIENTS* and *PD ORIGIN*.

5.2.3 T3. Security and privacy of minors.

This theme focuses on minors, with an emphasis on the threats, risks, and consent needed to ensure that their rights are adequately protected. This theme maps directly to the *CHILDREN* information type.

5.2.4 T4. Informed consent

Consent is a well-studied concept. The GDPR emphasizes on collecting explicit consent from individuals, enabling them to make informed decisions with regard to the handling of their personal data. However, according to the *reference model* we apply in this SLR, *CONSENT* is only one of the several possible legal bases for data collection and processing.

CONSENT is also the easiest to verify in software as it should be translated into interaction between the mobile app and the user. To investigate whether the practices in the mobile app code are GDPR compliant with respect to some other legal bases, a more comprehensive analysis of the app context must be conducted; such an analysis would not be necessarily limited to, e.g., the app's source code. We note that, despite not being popular in the SE literature, legal bases other than consent can play an important role in mobile apps development decisions.

5.2.5 T5. Security/privacy awareness

This theme surveys the users of mobile apps to study how well individuals are aware of their rights as well as common practices concerning privacy and security issues, often perceived in mobile apps' behaviors. While this theme highlights the importance of individuals' awareness about data protection, it cannot be directly mapped to our *reference model*.

5.2.6 T6. Consistency between apps and their privacy policies

Some papers investigate the consistency between the content of the privacy policy associated with an app against the actual behavior of that app. However, this research theme is often scoped to a use case such as the collection and sharing of sensitive personal data with third parties.

As pinpointed earlier, this theme is important to ensure that mobile apps are compliant with what is disclosed in their privacy policies, upon which individuals agree.

Advancing this theme requires in-depth analysis of at least the following information types: *PERSONAL DATA CATEGORY*, *LEGAL BASIS*, *RECIPIENTS*, *PD ORIGIN*, and *DATA SUBJECT RIGHT*. These information types facilitate the understanding of what personal data is collected, the source where it is obtained from, with whom it will be shared, what rights individuals can have on their personal data, and how these rights are implemented.

5.2.7 T7. Detecting violations of data minimization principle

Data minimization is an important guiding principle in GDPR: according to this principle, an app should not collect more personal data than necessary for its use.

Such a theme would require analyzing the context of the mobile app, specifically the *LEGAL BASIS* under which personal data is being collected and processed as well as the *PROCESSING PURPOSES* declared in the privacy policies of mobile apps.

To ensure that no personal data is unnecessarily collected, the findings related to these two information types should then be cross-checked against *PERSONAL DATA CATEGORY*, indicating the personal data types collected, and *PD ORIGIN*, indicating the source where the data is collected from.

5.2.8 T8. Summarizing privacy policies

This theme focuses on making privacy policies more accessible to individuals by providing informative summaries or categorization of the policies.

While the primary studies under this theme cannot be one-to-one mapped to information types in our *reference model*, we believe that the *reference model* can be useful to advance this research theme as it provides a structured hierarchy of the necessary details that must be identified in privacy policies.

5.3 RQ3. What are the different types of research contributions?

RQ3 explores the various contributions identified in the literature concerning GDPR. We distinguish in this SLR between *technical* and *non-technical* contributions: the former refers to proposing automated solutions for addressing GDPR privacy concerns, whereas the latter are concerned with sharing insights or recommendations derived from, e.g., surveys, aiming at advancing the research knowledge.

In total, 35 out of 60 primary studies have non-technical contributions, targeting various GDPR privacy concerns. The majority of them conducted empirical studies as well as literature or apps surveys to learn more about mobile apps' users and practices in different contexts. Concrete outcomes of such studies include, for instance, lessons learned and guidelines on the development of GDPR-compliant mobile apps (which also address users' privacy concerns).

Most of these studies are qualitative, relying on manual means for analyzing certain privacy concerns. However, some studies used simple tool support such as keyword-based or string search for extracting indicator terms, or applied third-party tools such as Mobile Security Framework (MobSF)[9], FlowDroid[10] or Wireshark[11] for static or dynamic code analysis. Such an automation was strictly used to support their empirical studies or surveys.

Primary studies with non-technical contribution often described user studies or app surveys. The number of participants involved in user studies performed under this category vary between eight to 6,124 participants. Similarly, the number of apps considered in apps surveys vary, ranging from small app datasets (27 in [S3], or 28 in [S2]) to more than 10,000 apps in large-scale studies such as [S1].

The remaining 25 primary studies make technical contributions. They proposed automation strategies by leveraging different enabling technologies: model-driven engineering was used to enhancing business process modeling notation with privacy considerations [S6], static and dynamic program analysis were used for detecting violations or vulnerabilities in mobile code [S35, S41, S46, S51, S59], NLP was employed for analyzing privacy policies and app reviews [S36, S44, S53, S57], machine learning (ML) and deep learning (DL) were used for classifying privacy policy provisions [S30] and detecting source and sink method detection [S1]. About half of the primary studies under this category leveraged a combination of these enabling technologies, particularly NLP, ML, DL, and static or dynamic program analysis. These studies pursued the following objectives:

- identifying data recipients [S8],
- detecting privacy issues in app reviews [S16],
- detecting sensitive information disclosure in mobile apps [S20],
- addressing privacy considerations entailed from using third-party apps or libraries [S28, S47],
- evaluating the consistency between privacy policies and actual app code [S37, S52, S54, S60],
- detecting abuse of data collection due to granted permissions [S42, S43, S49, S50]

We noticed that technical contributions were often built on existing tools. The novelty of these contributions resided in the integration of existing tools and processes into a comprehensive pipeline for answering certain research questions.

The aforementioned technical contributions were in all cases accompanied with validation elements such as, e.g., proof of concept examples [S6], user studies [S16, S59], or empirical validation using datasets. These datasets vary in size from small with less than 50 elements such as apps, privacy policies, app

[9] https://github.com/MobSF/Mobile-Security-Framework-MobSF
[10] https://github.com/secure-software-engineering/FlowDroid
[11] https://www.wireshark.org/

Table 6: Publicly Available Material

Source code	Datasets	Output files	Papers
✓	✓	✓	[S37, S50]
✓	✓	x	[S28, S46, S47]
x	✓	✓	[S2]
✓	x	x	[S1, S43, S48, S53, S60]
x	✓	x	[S8, S21]
x	x	✓	[S3]

reviews, or programming methods [S42]. Medium-sized datasets range from 51–500 elements [S16, S28, S30, S50, S52, S53, S57], while large datasets contain more than 500 elements [S1, S8, S20, S35–S37, S41, S44, S46, S49, S51, S53, S54].

As part of this RQ, we also looked into the app stores and operating systems targeted in the literature. Our results show that 31 papers considered Android only, eight considered both Android and iOS, one paper considered Android and Windows systems, and one paper focused exclusively on iOS.

By examining the 25 studies with technical contributions, we observe that all but four are exclusively focused on Android. Regarding the outliers, one utilizes Amazon Alexa (a virtual assistant) [S37], one is based on iOS [S59], and the last two mention Apple App Store as well as Google Play Store for work based on explicit consent (through GUI testing) and privacy policies [S49, S57]. The remaining 21 studies focusing on Android include: One study that investigates Xiaomi market Store [S43], five lack comprehensive details about app stores [S41, S42, S44, S50, S60], and the remaining ones utilize GooglePlay [S8, S16, S20, S28, S30, S35, S36, S41, S44, S46, S47, S51–S54].

5.4 RQ4. Are the research artifacts accompanying studies on GDPR privacy concepts publicly available?

Out of the 60 papers, only 14 papers (≈23%) have shared some artifacts associated with the conducted research; Table 6 lists these papers and indicates the type of artifacts (source code, datasets, output files) made available. Out of these 14 papers, only two papers released all three types of artifacts.

These results show that the open science practices in the SE community still need improvement in terms of adding appropriate licenses and sharing material through persistent repositories such as Zenodo or FigShare.

5.5 Research Gaps

Our findings indicate that despite the attention received by GDPR in mobile app research, existing work focuses only on certain aspects, such as consent or sharing data with third-party libraries. Leveraging the reference model by Amaral et al. [14], we identify research gaps that need yet to be covered, outlined below.

1. More research investigating personal data collection is needed. Table 4 shows that the primary studies have extensively investigated the collection of personal data and recipients of personal data sharing activities. These topics are often related to analyzing security issues concerning data leaks. To this end, we identify the following research gaps:

1.1 While there has been some work toward discovering sensitive personal data collected and processed in mobile apps, there is still ample room for tracing what data is being actually processed. According to Amaral et. al [14], privacy policies of mobile apps should explicitly list the personal data categories that are planned to be collected by the mobile app; sometimes the policy should also explicitly mention the origin from where such personal data is being collected. *Verifying whether the categories mentioned in the privacy policies are also the only ones being collected and processed by mobile apps is essential for checking the compliance of apps.*

1.2 Nowadays, personal data collection typically involves permission requests to access personal data collected by other apps on the same mobile device. *Another research gap concerns the investigation of indirect collection of personal data.* It is important to understand what personal data is collected indirectly, how the process is documented, and whether users are aware of that.

2. Exploring the legal bases other than "consent' can have an impact on mobile apps development. Our results show that most primary studies investigate the explicit consent of a user for data collection and processing. While consent is indeed an important legal basis, other legal bases are

as important. For instance, collecting and processing personal data based on legitimate interest for the purpose of investigating a security attack is a good example where consent might not even be asked. *The legal basis for handling personal data provides a complex, yet unexplored research landscape.*

3. The elicitation of comprehensive implementation details of data subject rights is paramount for developing legally-compliant mobile apps. Our SLR highlights that the existing work investigated data subject rights and their implementation in mobile apps only to a very limited extent. According to the GDPR, users (as data subjects) should be able to easily exercise their rights to access their personal data, lodge a complaint, etc. For instance, mobile apps should provide adequate and explicit information (ideally through direct interactions) for users to request account deletion or to access all their personal data collected during a specific time span [44]. *In-depth understanding of data subject rights should first be established in research to then make its way into practice.* Without such an awareness level on the different users' rights, developing better and legally-compliant apps is subject to the random, potentially correct interpretations of developers.

6 Threats to Validity

Internal Validity. The main consideration concerning the internal validity is selection bias. To mitigate the effect of this threat, we followed the guidelines reported in the literature for conducting systematic literature reviews. The results of our data extraction are also made publicly available and are thus open to scrutiny.

Construct Validity. The primary threat to construct validity is related to data extraction. To minimize this threat, the first three authors have conducted several online sessions to discuss the findings and ensure the mutual understanding of the extracted information categories. We further computed the inter-rater agreement on the mapping activity that was primarily performed by the first author, to ensure that the GDPR privacy concerns analyzed in this SLR are correctly mapped to the main topics discussed in the primary studies.

External Validity. We have explicitly defined the time range of interest for our SLR on the GDPR. We also iteratively refined our search query and retrieved the primary studies from multiple online repositories and search engines. While we are unable to cross out the potential threat of missing primary studies, we are confident that our findings are representative of the SE engineering research landscape and can thus be generalized.

7 Conclusion

In this paper we have presented our systematic literature review (SLR) concerning the research landscape in software engineering (SE) on addressing GDPR-relevant privacy concerns in mobile apps. We have comprehensively reviewed 60 primary studies, and analyzed them with respect to an existing comprehensive conceptual model which characterizes all possible privacy-relevant information types according to the GDPR provisions

We found that existing work focuses on three main GDPR concerns, related to: (i) sharing personal data with other entities (e.g., third parties) outside mobile apps, (ii) obtaining explicit consent from users, and (iii) collecting from the users directly certain categories of personal data. Additionally, our results show that the majority of existing work investigates problems such as data leaks, app permission requests, and obtaining users' consent. To address these challenges, existing work relies to a large extent on natural language processing, machine learning, and program analysis.

We further identified research gaps that are not yet well explored in the literature. These gaps include the necessity of: (i) consistency checking of what the apps disclose to users through privacy policies against what the apps' implementation details, (ii) investigating data subject rights (e.g., the possibility for users to request the deletion of their personal data in mobile apps), and (iii) studying the impact of legal bases other than consent on developing compliant mobile apps.

Acknowledgment

This research was funded in whole, or in part, by the Luxembourg National Research Fund (FNR), grant reference NCER22/IS/16570468/NCER-FT. For the purpose of open access, and in fulfillment of the obligations arising from the grant agreement, the author has applied a Creative Commons Attribution 4.0 International (CC BY 4.0) license to any Author Accepted Manuscript version arising from this submission.

Primary Studies

[S1] J. Samhi, M. Kober, A. K. Kabore, S. Arzt, T. F. Bissyandé, and J. Klein, "Negative results of fusing code and documentation for learning to accurately identify sensitive source and sink methods: An application to the android framework for data leak detection," in *2023 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 2023, pp. 783–794.

[S2] M. Hatamian, S. Wairimu, N. Momen, and L. Fritsch, "A privacy and security analysis of early-deployed covid-19 contact tracing android apps," *Empirical software engineering*, vol. 26, pp. 1–51, 2021.

[S3] L. H. Iwaya, M. A. Babar, A. Rashid, and C. Wijayarathna, "On the privacy of mental health apps," *Empirical Software Engineering*, vol. 28, no. 1, pp. 1–42, 2023.

[S4] I. Ullah, R. Boreli, and S. S. Kanhere, "Privacy in targeted advertising on mobile devices: a survey," *International Journal of Information Security*, vol. 22, no. 3, pp. 647–678, 2023.

[S5] S. J. De and A. Imine, "Consent for targeted advertising: the case of facebook," *AI & SOCIETY*, vol. 35, pp. 1055–1064, 2020.

[S6] P. Pullonen, J. Tom, R. Matulevičius, and A. Toots, "Privacy-enhanced BPMN: enabling data privacy analysis in business processes models," *Software and Systems Modeling*, vol. 18, pp. 3235–3264, 2019.

[S7] J. D. Fernández, M. Sabou, S. Kirrane, E. Kiesling, F. J. Ekaputra, A. Azzam, and R. Wenning, "User consent modeling for ensuring transparency and compliance in smart cities," *Personal and Ubiquitous Computing*, vol. 24, pp. 465–486, 2020.

[S8] D. Rodriguez, J. M. Del Alamo, M. Cozar, and B. García, "Roi: a method for identifying organizations receiving personal data," *Computing*, vol. 106, no. 1, pp. 163–184, 2024.

[S9] J. Fuster, S. Solera-Cotanilla, J. Pérez, M. Vega-Barbas, R. Palacios, M. Alvarez-Campana, and G. Lopez, "Analysis of security and privacy issues in wearables for minors," *Wireless Networks*, pp. 1–17, 2023.

[S10] J. Van Hoboken and R. Ó. Fathaigh, "Smartphone platforms as privacy regulators," *Computer Law & Security Review*, vol. 41, pp. 1–18, 2021.

[S11] A. Tsohou and E. Kosta, "Enabling valid informed consent for location tracking through privacy awareness of users: A process theory," *Computer law & security review*, vol. 33, no. 4, pp. 434–457, 2017.

[S12] I. Symeonidis, G. Biczók, F. Shirazi, C. Pérez-Solà, J. Schroers, and B. Preneel, "Collateral damage of facebook third-party applications: a comprehensive study," *Computers & Security*, vol. 77, pp. 179–208, 2018.

[S13] R. Binns and E. Bietti, "Dissolving privacy, one merger at a time: Competition, data and third party tracking," *Computer Law & Security Review*, vol. 36, pp. 1–19, 2020.

[S14] Z. He, "When data protection norms meet digital health technology: China's regulatory approaches to health data protection," *Computer Law & Security Review*, vol. 47, pp. 1–14, 2022.

[S15] M. B. Hosseini, T. D. Breaux, R. Slavin, J. Niu, and X. Wang, "Analyzing privacy policies through syntax-driven semantic analysis of information types," *Information and Software Technology*, vol. 138, pp. 1–18, 2021.

[S16] M. Hatamian, J. Serna, and K. Rannenberg, "Revealing the unrevealed: Mining smartphone users privacy perception on app markets," *Computers & Security*, vol. 83, pp. 332–353, 2019.

[S17] O. Akanfe, R. Valecha, and H. R. Rao, "Assessing country-level privacy risk for digital payment systems," *Computers & Security*, vol. 99, pp. 1–13, 2020.

[S18] K. O'Loughlin, M. Neary, E. C. Adkins, and S. M. Schueller, "Reviewing the data security and privacy policies of mobile apps for depression," *Internet interventions*, vol. 15, pp. 110–115, 2019.

[S19] B. Aljedaani, A. Ahmad, M. Zahedi, and M. A. Babar, "End-users' knowledge and perception about security of clinical mobile health apps: A case study with two saudi arabian mhealth providers," *Journal of Systems and Software*, vol. 195, pp. 1–24, 2023.

[S20] O. Olukoya, L. Mackenzie, and I. Omoronyia, "Towards using unstructured user input request for malware detection," *Computers & Security*, vol. 93, pp. 1–18, 2020.

[S21] F. Bemmann, M. Windl, J. Erbe, S. Mayer, and H. Hussmann, "The influence of transparency and control on the willingness of data sharing in adaptive mobile apps," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. MHCI, pp. 1–26, 2022.

[S22] Z. Liu, X. Wang, X. Li, and J. Liu, "Protecting privacy on mobile apps: A principal–agent perspective," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 29, no. 1, pp. 1–32, 2022.

[S23] M. Kowalewski, C. Utz, M. Degeling, T. Schnitzler, F. Herbert, L. Schaewitz, F. M. Farke, S. Becker, and M. Dürmuth, "52 weeks later: Attitudes towards covid-19 apps for different purposes over time," *Proceedings of the ACM on Human-Computer Interaction*, vol. 7, no. CSCW2, pp. 1–45, 2023.

[S24] A. Ekambaranathan, J. Zhao, and M. Van Kleek, "How can we design privacy-friendly apps for children? using a research through design process to understand developers' needs and challenges," *Proceedings of the ACM on Human-Computer Interaction*, vol. 7, no. CSCW2, pp. 1–29, 2023.

[S25] W. Pei, Y. Likhtenshteyn, and C. Yue, "A tale of two communities: Privacy of third party app users in crowdsourcing-the case of receipt transcription," *Proceedings of the ACM on Human-Computer Interaction*, vol. 7, no. CSCW2, pp. 1–43, 2023.

[S26] A. Bourdoucen, L. Nurgalieva, and J. Lindqvist, "Privacy is the price: Player views and technical evaluation of data practices in online games," *Proceedings of the ACM on Human-Computer Interaction*, vol. 7, no. CHI PLAY, pp. 1136–1178, 2023.

[S27] M. Marsch, J. Grossklags, and S. Patil, "Won't you think of others?: Interdependent privacy in smartphone app permissions," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CSCW2, pp. 1–35, 2021.

[S28] X. Zhang, J. Heaps, R. Slavin, J. Niu, T. Breaux, and X. Wang, "Daisy: Dynamic-analysis-induced source discovery for sensitive data," *ACM Transactions on Software Engineering and Methodology*, vol. 32, no. 4, pp. 1–34, 2023.

[S29] Z. Dong, L. Wang, H. Xie, G. Xu, and H. Wang, "Privacy analysis of period tracking mobile apps in the post-roe v. wade era," in *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, 2022, pp. 1–6.

[S30] S. Liu, B. Zhao, R. Guo, G. Meng, F. Zhang, and M. Zhang, "Have you been properly notified? automatic compliance analysis of privacy policy text with gdpr article 13," in *Proceedings of the Web Conference 2021*, 2021, pp. 2154–2164.

[S31] N. Vinayaga-Sureshkanth, R. Wijewickrama, A. Maiti, and M. Jadliwala, "An investigative study on the privacy implications of mobile e-scooter rental apps," in *Proceedings of the 15th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2022, pp. 125–139.

[S32] R. Sun, M. Xue, G. Tyson, S. Wang, S. Camtepe, and S. Nepal, "Not seen, not heard in the digital world! measuring privacy practices in children's apps," in *Proceedings of the ACM Web Conference 2023*, 2023, pp. 2166–2177.

[S33] U. Kishnani, N. Noah, S. Das, and R. Dewri, "Privacy and security evaluation of mobile payment applications through user-generated reviews," in *Proceedings of the 21st Workshop on Privacy in the Electronic Society*, 2022, pp. 159–173.

[S34] V. Schmitt, M. Poikela, and S. Möller, "Android permission manager, visual cues, and their effect on privacy awareness and privacy literacy," in *Proceedings of the 17th International Conference on Availability, Reliability and Security*, 2022, pp. 1–12.

[S35] X. Zhang, X. Wang, R. Slavin, T. Breaux, and J. Niu, "How does misconfiguration of analytic services compromise mobile privacy?" in *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering*, 2020, pp. 1572–1583.

[S36] H. O. Obie, W. Hussain, X. Xia, J. Grundy, L. Li, B. Turhan, J. Whittle, and M. Shahin, "A first look at human values-violation in app reviews," in *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. IEEE, 2021, pp. 29–38.

[S37] F. Xie, Y. Zhang, C. Yan, S. Li, L. Bu, K. Chen, Z. Huang, and G. Bai, "Scrutinizing privacy policy compliance of virtual personal assistant apps," in *Proceedings of the 37th IEEE/ACM international conference on automated software engineering*, 2022, pp. 1–13.

[S38] S. Liao, C. Wilson, L. Cheng, H. Hu, and H. Deng, "Measuring the effectiveness of privacy policies for voice assistant applications," in *Proceedings of the 36th Annual Computer Security Applications Conference*, 2020, pp. 856–869.

[S39] W. Seymour, M. Coté, and J. Such, "Legal obligation and ethical best practice: Towards meaningful verbal consent for voice assistants," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–16.

[S40] S. Liu, F. Zhang, B. Zhao, R. Guo, T. Chen, and M. Zhang, "Appcorp: a corpus for android privacy policy document structure analysis," *Frontiers of Computer Science*, vol. 17, no. 3, pp. 1–10, 2023.

[S41] J. Lou, X. Zhang, Y. Zhang, X. Li, X. Yuan, and N. Zhang, "Devils in your apps: Vulnerabilities and user privacy exposure in mobile notification systems," in *2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 2023, pp. 28–41.

[S42] V. K. Malviya, C. W. Leow, A. Kasthuri, Y. N. Tun, L. K. Shar, and L. Jiang, "Right to know, right to refuse: Towards ui perception-based automated fine-grained permission controls for android apps," in *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, 2022, pp. 1–6.

[S43] S. Zhang, H. Lei, Y. Wang, D. Li, Y. Guo, and X. Chen, "How android apps break the data minimization principle: An empirical study," in *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 2023, pp. 1238–1250.

[S44] T. T. Nguyen, M. Backes, and B. Stock, "Freely given consent? studying consent notice of third-party tracking and its violations of gdpr in android apps," in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, 2022, pp. 2369–2383.

[S45] M. Mehrnezhad, "A cross-platform evaluation of privacy notices and tracking practices," in *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2020, pp. 97–106.

[S46] Z. Tan and W. Song, "Ptpdroid: Detecting violated user privacy disclosures to third-parties of android apps," in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*. IEEE, 2023, pp. 473–485.

[S47] K. Zhao, X. Zhan, L. Yu, S. Zhou, H. Zhou, X. Luo, H. Wang, and Y. Liu, "Demystifying privacy policy of third-party libraries in mobile apps," in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*. IEEE, 2023, pp. 1583–1595.

[S48] T. T. Nguyen, M. Backes, N. Marnau, and B. Stock, "Share first, ask later (or never?) studying violations of {GDPR's} explicit consent in android apps," in *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 2021, pp. 3667–3684.

[S49] S. Koch, B. Altpeter, and M. Johns, "The {OK} is not enough: A large scale study of consent dialogs in smartphone applications," in *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, 2023, pp. 5467–5484.

[S50] V. K. Malviya, Y. N. Tun, C. W. Leow, A. T. Xynyn, L. K. Shar, and L. Jiang, "Fine-grained in-context permission classification for android apps using control-flow graph embedding," in *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 2023, pp. 1225–1237.

[S51] R. Slavin, X. Wang, M. B. Hosseini, J. Hester, R. Krishnan, J. Bhatia, T. D. Breaux, and J. Niu, "Toward a framework for detecting privacy policy violations in android application code," in *Proceedings of the 38th International Conference on Software Engineering*, 2016, pp. 25–36.

[S52] X. Wang, X. Qin, M. B. Hosseini, R. Slavin, T. D. Breaux, and J. Niu, "Guileak: Tracing privacy policy claims on user input data for android applications," in *Proceedings of the 40th International Conference on Software Engineering*, 2018, pp. 37–47.

[S53] A. Xiang, W. Pei, and C. Yue, "Policychecker: Analyzing the gdpr completeness of mobile apps' privacy policies," in *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, 2023, pp. 3373–3387.

[S54] L. Yu, X. Luo, J. Chen, H. Zhou, T. Zhang, H. Chang, and H. K. Leung, "Ppchecker: Towards accessing the trustworthiness of android apps' privacy policies," *IEEE Transactions on Software Engineering*, vol. 47, no. 2, pp. 221–242, 2018.

[S55] S. D. Gupta, "Developing a privacy risk analysis framework for heterogeneous iot network," in *2022 IEEE 30th International Requirements Engineering Conference (RE)*. IEEE, 2022, pp. 207–212.

[S56] M. Robol, T. D. Breaux, E. Paja, and P. Giorgini, "Consent verification under evolving privacy policies," in *2019 IEEE 27th International Requirements Engineering Conference (RE)*. IEEE, 2019, pp. 422–427.

[S57] T. Huang, V. Kaulagi, M. B. Hosseini, and T. Breaux, "Mobile application privacy risk assessments from user-authored scenarios," in *2023 IEEE 31st International Requirements Engineering Conference (RE)*. IEEE, 2023, pp. 17–28.

[S58] C. Braghin, S. Cimato, and A. Della Libera, "Are mhealth apps secure? a case study," in *2018 IEEE 42nd annual computer software and applications conference (COMPSAC)*, vol. 2. IEEE, 2018, pp. 335–340.

[S59] J. Gardner, Y. Feng, K. Reiman, Z. Lin, A. Jain, and N. Sadeh, "Helping mobile application developers create accurate privacy labels," in *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE Computer Society, 2022, pp. 212–230.

[S60] Y. Wang, M. Fan, J. Liu, J. Tao, W. Jin, H. Wang, Q. Xiong, and T. Liu, " Do as You Say: Consistency Detection of Data Practice in Program Code and Privacy Policy in Mini-App ," *IEEE Transactions on Software Engineering*, no. 01, pp. 1–23, 2024.

References

[1] R. Godwin-Jones, "Emerging technologies for language learning," *The Encyclopedia of Applied Linguistics*, 2012.

[2] J. L. Fernández-Alemán, I. C. Señor, P. Ángel Oliver Lozoya, and A. Toval, "Security and privacy in electronic health records: A systematic literature review," *Journal of Biomedical Informatics*, vol. 46, no. 3, pp. 541–562, 2013.

[3] B. Martínez-Pérez, I. De La Torre-Díez, M. López-Coronado, J. Herreros-González *et al.*, "Mobile apps in cardiology," *JMIR mHealth and uHealth*, vol. 1, no. 2, p. e2737, 2013.

[4] M. Milne-Ives, C. Lam, C. De Cock, M. H. Van Velthoven, E. Meinert *et al.*, "Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review," *JMIR mHealth and uHealth*, vol. 8, no. 3, p. e17046, 2020.

[5] A. A. Shaikh, H. Alamoudi, M. Alharthi, and R. Glavee-Geo, "Advances in mobile financial services: a review of the literature and future research directions," *International Journal of Bank Marketing*, vol. 41, no. 1, pp. 1–33, 2022.

[6] D. Torre, S. Abualhaija, M. Sabetzadeh, L. C. Briand, K. Baetens, P. Goes, and S. Forastier, "An ai-assisted approach for checking the completeness of privacy policies against GDPR," in *28th IEEE International Requirements Engineering Conference*, 2020.

[7] E. Union, "General data protection regulation," Accessed Nov. 07, 2021 [Online]. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

[8] P. Klaus and R. Chris, *Requirements Engineering Fundamentals*. Rocky Nook, 2011.

[9] A. I. Antón, J. B. Earp, T. A. Alspaugh, and C. Potts, "The role of policy and stakeholder privacy values in requirements engineering," in *5th IEEE International Symposium on Requirements Engineering (RE 2001), 27-31 August 2001, Toronto, Canada*. IEEE Computer Society, 2001, pp. 138–145.

[10] B. Martínez-Pérez, I. de la Torre-Díez, and M. López-Coronado, "Privacy and security in mobile health apps: A review and recommendations," *Journal of Medical Systems*, vol. 39, no. 1, 2014.

[11] M. E. Morales-Trujillo, G. A. García-Mireles, E. O. Matla-Cruz, and M. Piattini, "A systematic mapping study of privacy by design in software engineering," *CLEI Eletronic Journal (CLEIej)*, vol. 22, no. 1, 2019.

[12] L. H. Iwaya, A. Ahmad, and M. A. Babar, "Security and privacy for mhealth and uhealth systems: A systematic mapping study," *IEEE Access*, vol. 8, pp. 150 081–150 112, 2020.

[13] V. C. Andrade, R. D. Gomes, S. Reinehr, C. O. D. A. Freitas, and A. Malucelli, "Privacy by design and software engineering: a systematic literature review," in *Proceedings of the XXI Brazilian Symposium on Software Quality*, ser. SBQS '22. Association for Computing Machinery, 2023.

[14] O. Amaral, S. Abualhaija, D. Torre, M. Sabetzadeh, and L. C. Briand, "Ai-enabled automation for completeness checking of privacy policies," *IEEE Transactions on Software Engineering*, vol. 48, no. 11, pp. 4647–4674, 2022.

[15] S. Ghanavati, A. Rifaut, E. Dubois, and D. Amyot, "Goal-oriented compliance with multiple regulations," in *Proceedings of 22nd IEEE International Conference on Requirements Engineering*, 2014.

[16] M. I. Azeem and S. Abualhaija, "A multi-solution study on GDPR AI-enabled completeness checking of dpas," *Empir. Softw. Eng.*, vol. 29, no. 4, p. 96, 2024.

[17] O. A. Cejas, M. I. Azeem, S. Abualhaija, and L. C. Briand, "NLP-based automated compliance checking of data processing agreements against gdpr," *IEEE Transactions on Software Engineering*, vol. 49, no. 9, pp. 4282–4303, 2023.

[18] J. Bhatia, M. C. Evans, and T. D. Breaux, "Identifying incompleteness in privacy policy goals using semantic frames," *Requirements Engineering*, vol. 24, no. 3, 2019.

[19] Z. S. Li, C. Werner, N. Ernst, and D. Damian, "Towards privacy compliance: A design science study in a small organization," *Information and Software Technology*, vol. 146, p. 106868, 2022.

[20] I. Omoronyia, L. Cavallaro, M. Salehie, L. Pasquale, and B. Nuseibeh, "Engineering adaptive privacy: on the role of privacy awareness requirements," in *2013 35th International Conference on Software Engineering (ICSE)*. IEEE, 2013, pp. 632–641.

[21] M. Fan, L. Yu, S. Chen, H. Zhou, X. Luo, S. Li, Y. Liu, J. Liu, and T. Liu, "An empirical evaluation of gdpr compliance violations in android mhealth apps," in *2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2020, pp. 253–264.

[22] S. Kununka, N. Mehandjiev, and P. Sampaio, "A comparative study of android and ios mobile applications' data handling practices versus compliance to privacy policy," in *Privacy and Identity Management. The Smart Revolution - 12th IFIP WG 9.2, 9.5, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Ispra, Italy, September 4-8, 2017, Revised Selected Papers*, 2017.

[23] M. Alloghani, T. Baker, D. Al-Jumeily, A. Hussain, J. Mustafina, and A. J. Aljaaf, *A Systematic Review on Security and Privacy Issues in Mobile Devices and Systems*. Springer International Publishing, 2020, pp. 585–608.

[24] F. H. Semantha, S. Azam, K. C. Yeo, and B. Shanmugam, "A systematic literature review on privacy by design in the healthcare sector," *Electronics*, vol. 9, no. 3, 2020.

[25] G. Shrivastava, P. Kumar, D. Gupta, and J. J. P. C. Rodrigues, "Privacy issues of android application permissions: A literature review," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 12, 2020.

[26] The European Parliament and the European Council, "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data," 10 1995, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046.

[27] B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," *Information and Software Technology*, vol. 55, no. 12, pp. 2049–2075, 2013.

[28] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," 01 2007.

[29] E. Vanezi, G. M. Kapitsaki, D. Kouzapas, A. Philippou, and G. A. Papadopoulos, "Diálogop - A language and a graphical tool for formally defining GDPR purposes," in *Proceedings of the 2020 Research Challenges in Information Science - 14th International Conference, RCIS*, vol. 385. Springer, 2020, pp. 569–575.

[30] N. Mousavi Nejad, P. Jabat, R. Nedelchev, S. Scerri, and D. Graux, "Establishing a strong baseline for privacy policy classification," in *ICT Systems Security and Privacy Protection.* Springer International Publishing, 2020, pp. 370–383.

[31] M. Fan, L. Yu, S. Chen, H. Zhou, X. Luo, S. Li, Y. Liu, J. Liu, and T. Liu, "An empirical evaluation of gdpr compliance violations in android mhealth apps," in *2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE).* IEEE Computer Society, 2020, pp. 253–264.

[32] United States Congress, "Health Insurance Portability & Accountability Act. 104th Cong., Public Record 104-191 (1995-1996)," 08 1996.

[33] F. Ebrahimi, M. Tushev, and A. Mahmoud, "Mobile app privacy in software engineering research: A systematic mapping study," *Information and Software Technology*, vol. 133, 2021.

[34] C. Negri-Ribalta, M. Lombard-Platet, and C. Salinesi, "Understanding the gdpr from a requirements engineering perspective—a systematic mapping study on regulatory data protection requirements," *Req. Eng.*, vol. 29, no. 4, pp. 523–549, 2024.

[35] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering*, vol. 17, 06 2008.

[36] "Guidelines for conducting systematic mapping studies in software engineering," *Inf. Softw. Technol.*, vol. 64, no. C, p. 1–18, 2015.

[37] T. Dybå, T. Dingsøyr, and G. K. Hanssen, "Applying systematic reviews to diverse study types: An experience report," in *First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)*, 2007, pp. 225–234.

[38] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Information and Software Technology*, vol. 64, pp. 1–18, 2015.

[39] L. Zhao, W. Alhoshan, A. Ferrari, K. J. Letsholo, M. A. Ajagbe, E.-V. Chioasca, and R. T. Batista-Navarro, "Natural language processing for requirements engineering: A systematic mapping study," *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, pp. 1–41, 2021.

[40] L. Montgomery, D. Fucci, A. Bouraffa, L. Scholz, and W. Maalej, "Empirical research on requirements quality: a systematic mapping study," *Requirements Engineering*, vol. 27, no. 2, pp. 183–209, 2022.

[41] K. Krippendorff, "Estimating the reliability, systematic error and random error of interval data," *Educational and psychological measurement*, vol. 30, no. 1, pp. 61–70, 1970.

[42] A. Aljeraisy, M. Barati, O. Rana, and C. Perera, "Privacy laws and privacy by design schemes for the internet of things: A developer's perspective," *ACM Computing Surveys (Csur)*, vol. 54, no. 5, pp. 1–38, 2021.

[43] D. A. Tamburri, "Design principles for the general data protection regulation (gdpr): A formal concept analysis and its evaluation," *Information Systems*, vol. 91, p. 101469, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306437919305216

[44] European Data Protection Board, "Guidelines 01/2022 on data subject right - right of access," 2022. [Online]. Available: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-012022-data-subject-rights-right-access_en