Automatic Identification of Privacy and Security Requirements: A Systematic Literature Review

Anonymous Authors

Abstract

The utmost importance of privacy and security requirements in software development calls for adopting methods that enable the identification and proactive mitigation of these issues during the system development. Our survey of 45 primary studies provides an overview of the methods, document types, and datasets employed in tackling this challenge, along with an analysis of approaches demonstrating superior performance based on document types and specific identification problems. Analysis reveals a wide adoption of ML-based systems on diverse datasets, showcasing the effectiveness of leveraging various source of information to identify privacy and security requirements in software development.

Keywords: Software Requirements, Privacy & Security, Automatic Identification, Natural Language Processing, Machine Learning.

1 Introduction

Requirements Engineering (RE) plays a pivotal role in software development by encompassing vital activities that focus on understanding and fulfilling the capabilities and characteristics demanded by a system [46]. Through stages like elicitation, analysis, specification, and validation, RE aims to comprehend customer needs and translate them into precisely defined requirements [47], which include non-functional requirements (NFRs). The latter specify system qualities that extend beyond its core functionality, encompassing various attributes essential for its overall performance and success. Among NFRs, privacy and security have emerged as prominent concerns in software development. Incidents involving unauthorized data exploration, misuse of information, and unauthorized disclosure of personal data have raised awareness regarding privacy risks [48]. Users may be unaware of when and for what purposes their sensitive information is collected, analyzed, or transmitted [49]. To mitigate these concerns, regulations such as General Data Protection Regulation (GDPR) in the European Union have been established, specifically to safeguard individuals' sensitive personal data [50]. Simultaneously, security measures aim to prevent harm caused by

attackers exploiting vulnerabilities within systems [51]. It is essential to address both privacy and security issues as soon as possible during the entire software development, rather than solely focusing on implementation [52, 53]. However, due to the susceptibility to errors and time-consuming nature of RE processes [54], there has been a proliferation of approaches dedicated to automating the identification of privacy and security requirements throughout the entirety of the software development lifecycle.

In this work, we present a systematic literature review (SLR) aiming to identify and analyze the commonly used approaches for automatically identify privacy and security requirements that exploit various sources of information (such as requirements documents, user feedback, vulnerabilities, contracts, and so on) and discuss the factors that affect their performance [55]. In particular, our study reviews relevant papers published between 2000 and 2022 and aims to answer four main research questions (RQs). **RQ1** focuses on analyzing the methods employed to automatically identify privacy and security requirements, seeking insights into the techniques used so far to give an immediate overview to researchers and practitioners in the area. **RQ2** aims to examine the types of documents considered by the selected solutions. **RQ3** focuses on the datasets employed for building and validating the privacy and security detection solutions, providing a comprehensive catalog that facilitates the replication and comparison of research findings, while also providing valuable insights to practitioners regarding the literature's content. Finally, **RQ4** focuses on identifying the most effective methods, presenting a curated list of the top solutions based on the document type and identification problem considered.

A comprehensive search was performed about the existing approaches on four digital libraries indexing the relevant publications of the field to find the answer to each research question. Our search has resulted in finding 5,758 articles, of which 45 articles have been identified as primary study. We compare the selected approaches according to different aspects, including identification problems, classification methods, programming languages, document types, datasets, and performances. Our SLR reveals that the field is experiencing growth, with AI techniques being widely utilized. However, the existing methods face various challenges. Notably, approaches dedicated solely to privacy issue detection employ a diverse range of datasets, highlighting the necessity for a well-established dataset. Additionally, the complexity of document types significantly affects the performance of identification.

To the best of our knowledge, this is the first SLR about automatic identification of privacy and security requirements that helps researchers and practitioners to find the most appropriate methods according to the available source of information. Indeed, only *Netto et al.* [56] present the results of a mapping study aimed to understand the current state of the art regarding privacy and security in RE approaches. The study covers the period from 2000 to 2016 and provides a broader overview of the approaches and techniques used in the field, aiming to understand the overall landscape of research topics, methods, study types, research problems, and future trends or directions related to privacy and security in RE, encompassing both automated and manual methodologies. Two studies focus on the identification and classification of NFRs [57] [58]. In particular, the former aims to understand how the employed ML algorithms work and how they are evaluated, while the latter review the existing

literature to identify approaches, algorithms, and methodologies employed for automatic classification of software requirements, without explicitly targeting privacy and security requirements.

The remainder of this paper is structured as follows. Section 2 details the study design and execution, outlining the methodology employed to ensure the validity of the research. Section 3 presents the study results, encompassing the extracted data from the primary studies included, and offers a comprehensive analysis of these results concerning the research questions. Section 4 lists an exploration of the lessons learned, implications that arise from them and possible future research directions. Section 5 addresses the threats to the validity of our study. Section 6 provides conclusions also summarizing topics for future research.

2 Methodology

Our SLR adheres to the guidelines recommended by Kitchenham [55], as illustrated in Figure 1. The review protocol comprises three main stages: Planning, Conducting, and Reporting Results.

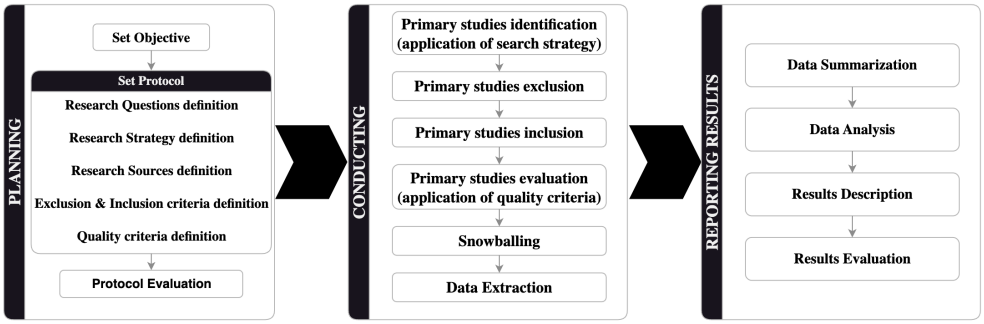


Fig. 1: Stages of the employed SLR protocol.

In the Planning stage, all the criteria for the study are defined. This phase encompasses the definition of the objective with its research questions and the identification of the search strategy that includes the formulation of the search string to be used in the different databases (search sources) chosen to extract the studies of interest. At this stage, we have to define the criteria for the exclusion and inclusion of papers, which are then subjected to quality control through related qualitative analysis. At the end of this phase, the defined protocol is discussed and revised, which turns out to be the final check before moving on to the next stage.

In the Conducting stage, we implement and exploit the definitions from the previous phase. First, there is the application of the search strategy of using search strings in the databases identified as sources. The output of this phase is filtered through the criteria of exclusion, inclusion, and quality. Next is the application of snowballing, which consists of searching through the citations and references of the papers that

make up the starting set, in our case, the papers that meet the criteria of the previous step. Snowballing produces a set of papers that will be added to those identified at the end of the quality assessment, forming the set from which the data of interest in our study will be extracted.

In the Reporting Results stage, we collect and analyze the results, which will identify limitations, open challenges, and possible future developments in the topic.

2.1 Research Questions

The objective of our SLR is to provide a comprehensive overview of studies that focus on the identification of security and privacy requirements from various sources of information. The aim is to conduct an in-depth analysis of these works, with the intent of gaining insights into the most efficient or top-performing methods, the prevalent datasets employed, and the challenges associated with the techniques or datasets used. Additionally, the SLR seeks to identify the key issues and challenges that need to be addressed within the scope of the research topic. To achieve these objectives, we have formulated the following research questions as guidelines for our investigation.

Research Question 1

How is the problem of identifying the privacy and security requirements modeled and which are the methods exploited?

The first research question (RQ1) aims to provide an overview of the approaches and techniques used for the detection of security and privacy requirements and to allow an analysis of alternative solutions not extensively explored in the past. Examining these methods is of utmost importance as it allows for a comprehensive understanding of the field, identifying effective techniques, addressing challenges, and informing both research and practice in RE.

Research Question 2

What document types are exploited to identify privacy and security requirements?

The second research question (RQ2) concerns the documents exploited to build and validate the approaches for detecting privacy and security requirements. This study aims at understanding the documents used in the literature for the above task, identifying the information source types that are still under-examined to include in future studies' approaches.

Research Question 3

Which datasets are employed to identify privacy and security requirements?

Another point of discussion is the set of datasets currently available because the data used may be suitable for specific techniques and methods and may need to be

more representative of the requirements in natural working environments. Analyzing the datasets employed in the identification of privacy and security requirements is essential for assessing their representativeness, evaluating dataset quality, identifying limitations, enabling replication and comparison, and supporting future research directions. It helps in ensuring the reliability and validity of the research findings and contributes to the advancement of knowledge in the field. Therefore, the third research question (RQ3) aims to analyze these possible discrepancies by providing an overview of the datasets available in the literature with their strengths and limitations.

Research Question 4

Which are the most effective methods for identifying privacy and security requirements?

This research question (RQ4) seeks to understand the strengths of the techniques used, highlighting the most efficient ones or those that perform well. The analysis of potential techniques that can be used could highlight how using particular methods in specific contexts or circumstances results in better performance than other techniques that, for example, might be better suited for other types of requirements.

2.2 Search strategy

A search strategy is employed to systematically identify relevant research articles in a SLR. The strategy involves defining keywords and phrases related to the research topic and using them to search selected databases.

The search strategy is crucial for ensuring a comprehensive and systematic identification of articles that address the research questions. Four critical databases were considered for our SLR: ACM Digital Library, IEEE Xplore Digital Library, Springer Link, and ScienceDirect.

The defined search strategy included a combination of keywords and phrases related to three major terms:

- identification
- privacy and security
- requirements engineering.

The employed synonyms for each major term are reported in Table 1, where boolean operators (AND, OR) have been used to combine the search terms appropriately. In particular, the set of terms related to the same major term was grouped with the operator OR. Then, the out-coming set for each major term was combined with the other major terms' set with the AND operator.

The search strategy was designed to be comprehensive, aiming to identify relevant articles that address the research topic from different angles and perspectives. No search terms related to techniques peculiar to NLP, ML, or DL were used to capture all works concerning the identification of privacy and security requirements without concerns for the methods employed.

The search was limited to articles published in English between 2000 and March 2023 to ensure the relevance and currency of the literature and was conducted from late August 2023.

Major Term	Synonyms
Identification	"identification" OR "detection" OR "mining" OR "recovery" OR "recognition" OR "classification" OR "capture" OR "retrieval" OR "analyze"
Privacy & Security	"privacy" OR "data privacy" OR "information privacy" OR "data protection" OR "sensitive data" OR "personal data" OR "confidential data" OR "personal information" OR "data confidentiality" OR "person-related data" OR "security" OR "information security"
Requirements Engineering	"requirement engineering" OR "requirements analysis" OR "requirements solicitation" OR "requirements specification" OR "requirements verification" OR "requirements management" OR "software requirements" OR "requirements elicitation" OR "requirements modeling" OR "requirements documentation" OR "requirement validation"

Table 1: The defined search string.

2.3 Selection criteria

The selection criteria in this SLR encompassed exclusion and inclusion criteria. Exclusion criteria (EC) were applied to screen out articles that did not focus on the relevant topic, were in the form of posters or unpublished documents, or were not freely accessible in their entirety. Such articles were excluded from the final selection to ensure the quality and relevance of the included articles. In particular, the following are the exclusion criteria used in this phase:

EC1 The article does not propose an approach to identify privacy and security requirements automatically.
EC2 The article is a poster or an unpublished document.
EC3 The article is not entirely available for free.

Inclusion criteria (IC) were employed to identify articles that met the specific requirements of the research topic. Articles were included if they pertained to the identification of privacy in the context of RE (IC1) or if they addressed the identification of security in RE (IC2). Subsequently, if at least one the previous criteria is met, we further check if the works provided insights into the types of requirements utilized for identifying privacy, or offered datasets employed in studies evaluating privacy identification methods. Similarly, we replicate the same checks regarding the security aspect. These further inspections are not strictly related to the inclusion of these works in our analysis, since they have been already added through IC1 and IC2. Nevertheless, we decide to leverage those controls to better understand what that work is providing about the identification of privacy and/or security requirements, but also to get a first screening regarding the domain of the works (security, privacy or both). Articles satisfying these inclusion criteria were included in the final selection for further analysis in our SLR, ensuring the comprehensiveness and rigor of the literature review process. For clarity, the following are the criteria for inclusion:

IC1 The article is about the identification of privacy in RE
IC1.1 The article reports the type of requirement used to identify privacy
IC1.2 The article offers a dataset employed in studies evaluating methods to identify privacy
IC2 The article is about the identification of security in RE
IC2.1 The article reports the type of requirement used to identify security
IC2.2 The article offers a dataset employed in studies evaluating methods to identify security

2.4 Quality assessment

Each study passed a quality evaluation to establish its legitimacy and usefulness. It may be considered an additional criterion for selecting papers for our SLR. We used several quality questions to weigh the received candidate studies after applying inclusion/exclusion criteria. For each question, a paper can have a value of 0, 0.5, or 1. Low-quality studies (those with weights less than 5) were removed. We utilized the following questions to apply our quality evaluation criteria:

Q1 Are the RQs or objectives clearly defined?
Q2 Are the findings clearly reported in the paper?
Q3 Have the study's limitations been explicitly analyzed?
Q4 Have all methods been fully defined?
Q5 Is it clear what projects were used to validate the method?
Q6 Is it clear how the method's performance has been measured?
Q7 Have the technique been applied to a case study, or have they been justified by appropriate examples?
Q8 Have the techniques been compared with other methods?

2.5 Snowballing

To ensure capturing of all papers that fall within the searched scope, we applied the technique of snowballing, which involves using the reference list of an article (backward snowballing) or the citations of the article (forward snowballing) to identify other articles. For the application, one must start with a few existing articles on or around the topic of interest; such papers constitute the "initial set" and, in the context of the present SLR, were the output works of the quality assessment phase. Starting with the initial set, each iteration of snowballing consists of backward and forward snowballing to collect related papers. The papers are then filtered through the exclusion and inclusion criteria defined earlier. Papers that pass these criteria become the initial set on which backward and forward snowballing is performed again. The process ends when no new papers are identified after applying the inclusion and exclusion criteria. Note that to the previous exclusion criteria, we have added other criteria that we were not able to set as in the case of the 4 databases used before, because for this step we used Semantic Scholar as search engine database. These criteria are listed below:

EC4 Already checked in previous steps.
EC5 Published out of the interested period.

EC6 The article is not in English.

Please note that EC6 has been included in the snowballing process, unlike the Search strategy, as we did not apply a language-based filter while conducting backward and forward phases through Semantic Scholar.

2.6 Data Extraction

In the data extraction phase, we collected some information of interest from the papers filtered through quality assessments. The information gathered is detailed in Table 2, from which we can see that there are 4 main factors of interest: *Publication*, *Methodology*, *Validation*, and *Contribution*. The Publication factor is for a demographic analysis of the main venues for this type of work and the evolution of interest in this topic over the years. The Methodology factor aims to summarize the techniques and technologies used, particularly regarding the type of task for detecting privacy and security requirements. Through the Validation factor, we want to classify the datasets used and the performance of the models provided so far. Finally, the Contribution factor helps us to understand the open challenges and the limitations highlighted by the analyzed works.

Dimension	Attributes	Description
Publication	Publication Types	The kind of publication (journal, article).
	Publication Venues	The venue of the publication.
	Publication Years	The year of publication.
Methodology	Task	The task addressed (binary classification, clustering, ...).
	Target	The target of the task (NFRs, security/non-security, ...).
	Technique	The technique used to address the task.
	Technology	The technology used to implement the technique.
Validation	Dataset	The dataset used for the evaluation.
	Document Type	The kind of information in the dataset.
	Evaluation criteria	Criteria used to evaluate the effectiveness.
	Experimental Results	Results obtained by applying the technique proposed.
Contribution	Limitations	The limitations of the approach.
	Findings	The insights obtained from the study.
	Open challenges	The challenges and future directions of the study.

Table 2: Information gathered from selected papers.

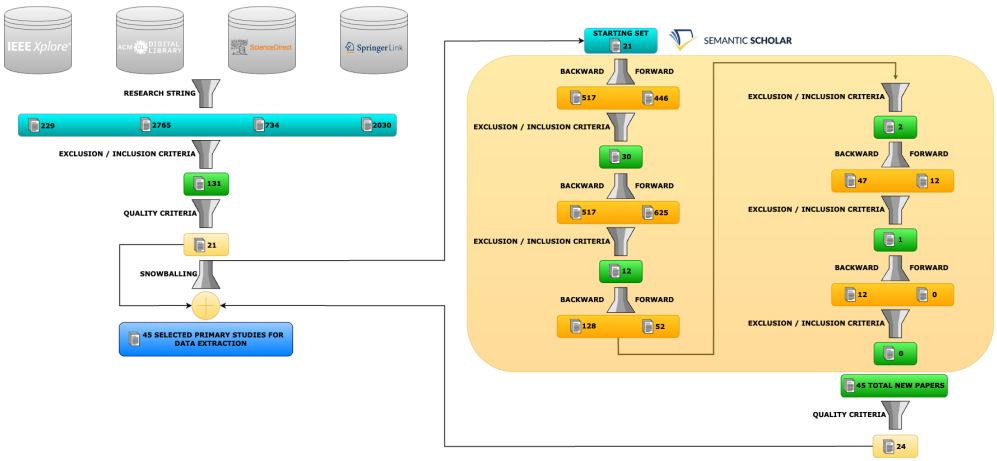


Fig. 2: Steps of the employed SLR protocol and achieved results.

3 Results and Discussion

A graphical summary of the entire process is presented in Figure 2, where we report results for each stage and step. We can observe that the initial search for relevant articles using search strings across the four databases resulted in a total of 5758 articles. The selection process for these articles began with an initial screening using titles and abstracts, applying exclusion criteria to eliminate irrelevant ones. Subsequently, the articles meeting or surpassing these criteria underwent a thorough full-text review. Inclusion decisions were based on their significance in privacy and security identification within requirements engineering, as well as the types of requirements and datasets used in evaluating privacy and security identification methods. From the initial pool of 131 articles, only 21 were considered to meet the quality standards for inclusion and were chosen for data extraction.

The 21 selected articles formed the initial set for the snowballing phase, with subsequent iterations specified in the previous section. After the first iteration, 30 articles were identified and became the starting set for the next iteration. After the second iteration, 12 new papers were identified, which, at the third iteration, led to the discovery of 2 additional articles. The fourth iteration identified only 1 new paper, and no new papers were identified in the fifth iteration. After the completion of all five iterations, a total of 45 articles (30 + 12 + 2 + 1) were identified and subjected to quality control using the pre-defined criteria.

Of the 45 newly identified articles, 24 were selected for data extraction and added to the initial snowballing set of 21 articles, resulting in a total of 45 articles to be analyzed and synthesized to gain insights into the current state of automatically identifying privacy and security requirements.

In the following subsection, we provide a comprehensive overview of the demographic characteristics of the papers. We analyze them based on their objectives, specifically whether they identify security and/or privacy requirements. Additionally, we examine the quality and frequency of publications over the years, as well as

the venues where these works are published. Subsequently, our analysis is dedicated to addressing the research questions outlined in Subsection 2.1, aiming to provide satisfactory answers to these inquiries.

Authors	Paper Ref.	IC1 (12)	IC1.1 (10)	IC1.2 (3)	IC2 (40)	IC2.1 (37)	IC2.2 (11)
A. Mahmoud et al.	[1]	✓	✓		✓	✓	
A. Sainani et al.	[2]	✓			✓		
A. K. Massey et al.	[3]	✓			✓		
D. C. Nguyen et al.	[4]	✓	✓		✓	✓	
M. Riaz et al.	[5]	✓	✓	✓	✓	✓	✓
B. Li et al.	[6]	✓	✓	✓	✓	✓	✓
C. Tao et al.	[7]	✓	✓		✓	✓	
F. Casillo et al.	[8]	✓	✓				
Z. S. Li et al.	[9]	✓	✓	✓			
F. Ebrahimi et al.	[10]	✓	✓				
S. Zimmeck et al.	[11]	✓	✓				
S. McIlroy et al.	[12]	✓	✓				
R. Jindal et al.	[13]				✓	✓	✓
V. Varenov et al.	[14]				✓	✓	✓
D. A. López-Hernández et al.	[15]				✓	✓	✓
C. Baker et al.	[16]				✓	✓	✓
J. Slankas et al.	[17]				✓	✓	✓
A. Kobilica et al.	[18]				✓	✓	✓
Md. Abdur Rahman et al.	[19]				✓	✓	✓
M. Mohamad et al.	[20]				✓	✓	✓
S. Gärtner et al.	[21]				✓	✓	
J. Slankas et al.	[22]				✓	✓	
S. Imtiaz et al.	[23]				✓	✓	
K. Schneider et al.	[24]				✓	✓	
Tong Li et al.	[25]				✓	✓	
X. Xiao et al.	[26]				✓	✓	
S. Imtiaz et al.	[27]				✓	✓	
W. Wang et al.	[28]				✓		
J. Cleland-Huang et al.	[29]				✓	✓	
L. Toth et al.	[30]				✓	✓	
E. Knauss et al.	[31]				✓	✓	
A. Rashwan et al.	[32]				✓	✓	
T. Hey et al.	[33]				✓	✓	
Z. Kurtanović et al.	[34]				✓	✓	
X. Luo et al.	[35]				✓	✓	
O. AlDhafer et al.	[36]				✓	✓	
C. Li et al.	[37]				✓	✓	
M. Younas et al.	[38]				✓	✓	
I. Khurshid et al.	[39]				✓	✓	
M. Younas et al.	[40]				✓	✓	
R. Chatterjee et al.	[41]				✓	✓	✓
S. Amasaki et al.	[42]				✓	✓	
K. Kaur et al.	[43]				✓	✓	
G. Li et al.	[44]				✓	✓	
N. Munaiah et al.	[45]				✓	✓	

Table 3: List categorizing the primary studies based on the problem they address, i.e., privacy and security identification in orange, privacy identification in red, security identification in blue.

3.1 Descriptive Overview

This section provides a summary of the primary studies selected during the Conducting phase. The selected studies, along with the inclusion criteria defined in Subsection 2.3, are listed in Table 3. The table includes details such as authors, article identification, and the inclusion criteria, with check marks denoting whether each article met the criteria. The inclusion criteria are categorized into two groups: privacy and security. The privacy-related criteria consist of IC1 and its related sub-criteria, IC1.1 and IC1.2. Similarly, the security-related criteria include IC2 and its related sub-criteria, IC2.1 and IC2.2. The number of occurrences for each criterion is reported in Table 3 within parentheses for reference.

Among the 45 selected papers, 40 focus on security identification, and the majority of these (37 papers) employ AI techniques for this purpose. In contrast, the number of papers addressing privacy identification is considerably lower, with only 12 papers meeting the criteria (IC1), and the majority of them (10 papers) utilizing AI-based techniques. The results reveal a significant disparity between the number of papers related to security and privacy. Out of the 40 security papers, only 7 of them also delve into privacy identification. This suggests that privacy is often treated as a distinct aspect from security, and it is not always integrated or considered as part of the security concerns. Furthermore, the findings indicate that only 5 papers exclusively focus on privacy requirement identification, highlighting a relatively limited emphasis on this specific aspect within the research landscape.

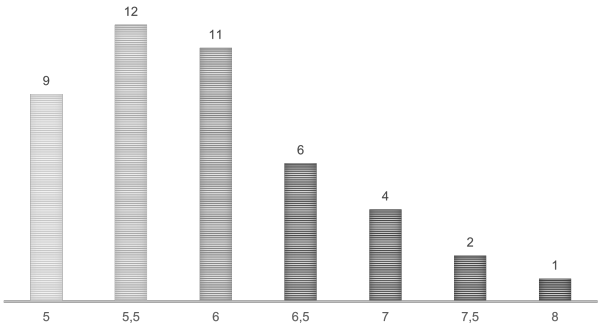


Fig. 3: Quality of the primary study design and results presentation.

The analysis uncovers compelling insights regarding the current state of research on security and privacy requirements detection. Figure 3 illustrates that only a limited number of papers achieved a score of 7 or higher in terms of the quality of their primary study design and results presentation. More precisely, four papers obtained a score of 7, two papers scored 7.5, and one paper received a score of 8. This observation emphasizes that the field still has considerable room for improvement in producing high-quality research.

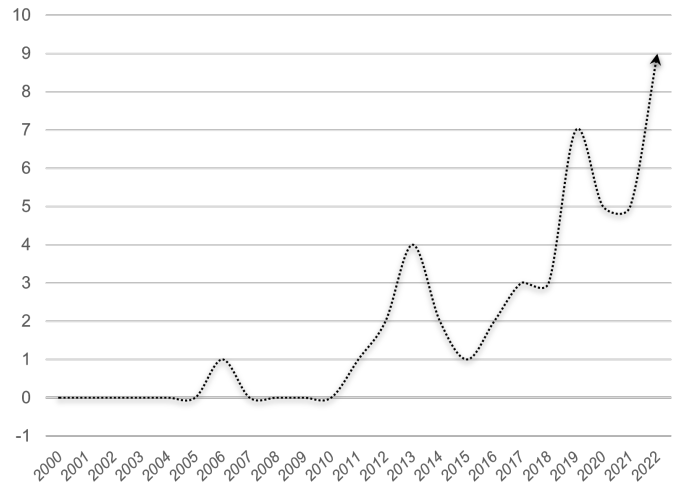
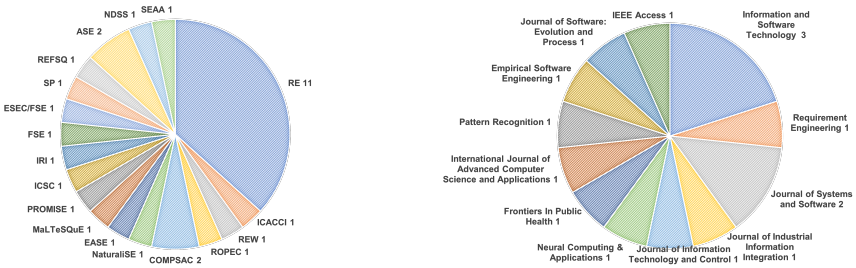


Fig. 4: Papers over the years.

Turning our attention to publication trends, Figure 4 depicts the distribution of publication years for the selected primary studies, showcasing a notable surge in the number of papers published on the topic of interest over the last five years. Among the 45 papers retrieved through the applied search strategy, a significant majority of 32 papers were published after 2017. Notably, the year 2022 saw the highest number of publications, with 9 papers. This observation underscores the increasing interest among researchers in this area and emphasizes the importance of up-to-date and comprehensive reviews to guide future research endeavors.



(a) Conference papers

(b) Journal papers

Fig. 5: Primary studies categorized by conference and journal venues.

Furthermore, our analysis of the publication venues for these papers unveils intriguing patterns, as illustrated in Figure 5. It is noteworthy that the majority of the selected primary studies were presented in the form of conference papers or workshops,

accounting for a total of 30 papers. Additionally, 15 papers were published in journals. Among the conference papers, 3 were featured in symposiums, 3 in workshops, and the remaining 24 were presented at general conferences. Particularly noteworthy is the International Requirement Engineering Conference, which emerged as the most prevalent conference, with a total of 11 papers published there. Regarding the 15 journal papers, 3 were published in Information and Software Technology, 2 in the Journal of Systems and Software, and the rest appeared in various journals with varying rankings. This information provides valuable insights into the popularity of different publication formats in the field and aids in identifying potential gaps in the existing literature.

3.2 RQ1 - How is the problem of identifying the privacy and security requirements modeled and which are the methods exploited?

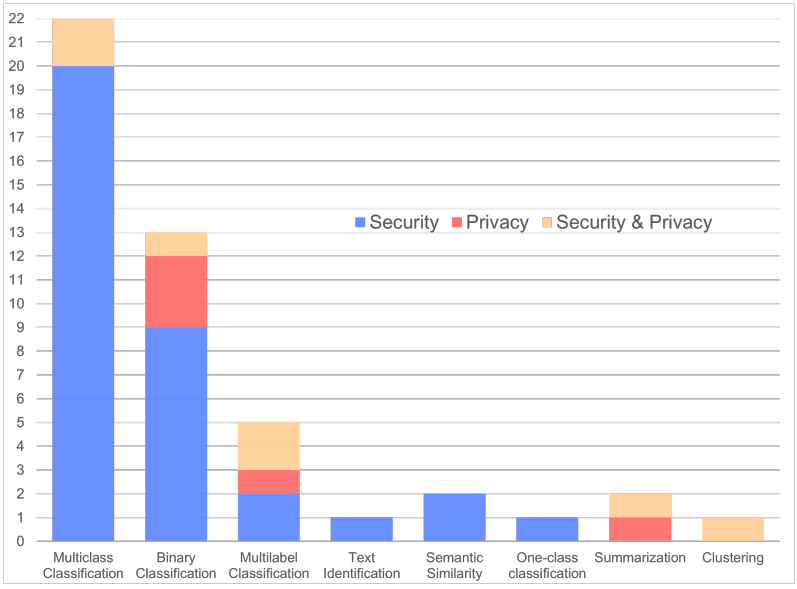


Fig. 6: Number of papers based on the type of task and problem addressed.

During the paper review process, it was observed that various approaches were employed to model the problem and identify relevant requirements, as shown in Figure 6. The chart presents a breakdown of the papers based on their tasks and objectives, revealing that the most commonly used method was multi-class classification. Specifically, 20 papers focused on identifying security requirements among other NFRs [13–17, 19, 27, 29, 30, 32–37, 39, 41–44], while 2 papers covered both security and privacy [1, 6]. The second most popular method utilized was binary classification, with 3 papers focused on identifying privacy requirements [8, 9, 11], 9 papers

concentrated on security requirements [18, 20–22, 24, 25, 28, 31, 37], and 1 paper covered both security and privacy [4]. The remaining methods have lower representation and included multi-label classification [2, 5, 12, 23, 36], summarization [7, 10], semantic similarity [38, 40], and other unique approaches, such as text identification [26], one-class classification [45], and clustering [3].

Concerning the selection of techniques for addressing the challenge of privacy and security requirement identification, certain methods appear to be more prevalent than others, as depicted in Figure 7.

SVM (Support Vector Machine) and TF-IDF (Term Frequency-Inverse Document Frequency) are frequently employed techniques in the context of identifying security and privacy requirements. Additionally, BERT (Bidirectional Encoder Representations from Transformers), a masked-language model introduced by Google researchers in 2018, is also used for identifying privacy- and security-related requirements. Moreover, several other shallow machine learning techniques, such as logistic regression (LR), Naive Bayes (NB), and k-nearest neighbors (KNN), complement the aforementioned approaches.

Generally, researchers have employed various techniques and models when it comes to classifying requirements. Notably, SVM, TF-IDF, and NB have emerged as the most commonly used techniques across 17 papers [2, 4, 8, 11, 12, 18, 20–23, 30, 32, 34, 37, 39, 42, 45], 15 papers [7, 10–13, 15, 17, 20, 22, 30, 37, 39, 40, 42, 45] and 15 papers [2, 5, 8, 12, 17, 18, 20–23, 25, 30, 37, 39, 42], respectively. LR was also a popular choice, appearing in 10 papers [7, 8, 11, 18, 23, 25, 28, 30, 39, 42], while KNN was employed in 11 papers [5, 6, 8, 17, 18, 21–23, 30, 37, 39].

Several other techniques and models were utilized in multiple instances. BERT, for example, appeared in 7 papers [2, 6, 14, 33, 35, 41, 44], while Word2Vec was employed in 6 papers [19, 20, 38, 40, 42, 43]. Decision trees were also a popular choice, featuring in 4 papers [8, 13, 15, 18, 23, 30], and random forests were employed in 6 papers [2, 8, 20, 23, 39, 42]. Additionally, deep learning models such as LSTM appeared in 6 papers [2, 6, 18, 19, 41, 43], demonstrating their recurrent use across the papers.

Less commonly utilized techniques and models encompass various neural network architectures, such as Artificial Neural Network (ANN) [6, 15, 16, 23], Convolutional Neural Network (CNN) [6, 8, 16, 18], and Gated Recurrent Unit (GRU) [19, 36, 41]. Word embeddings, like the Universal Sentence Encoder [9] and GloVe [10, 43], were also employed. Additionally, semantic similarity measures, including Normalized Google Distance (NGD) [1, 40], Lowest Common Subsumer (LCS) [21], and Text Semantic Similarity (TSS) [1]), were used in certain instances. Ensemble learning techniques [18, 23, 39] and combined approaches [5, 22] were also present, albeit less frequently, across the selected papers. Based on the comprehensive analysis presented above, we can now offer a conclusive answer to the first research question.

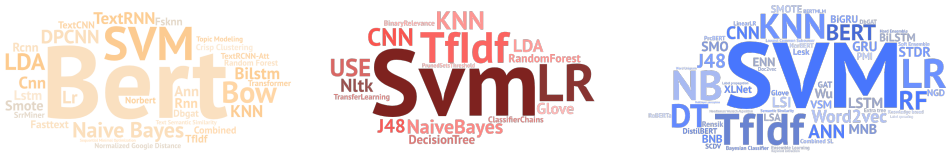


Fig. 7: Most used techniques divided by target: security & privacy (yellow), privacy (red) and security (blue).

Answer to RQ1

Multiclass classification emerges as the predominant approach for modeling the problem of identifying security and privacy requirements. However, binary classification has also been proposed in several instances within the literature.
The prevalence of TF-IDF, SVM, and NB as frequently used techniques underscores their wide popularity and effectiveness in text classification tasks. Moreover, the increasing mentions of BERT and Word2Vec in the papers signal their rising importance and relevance in the field.

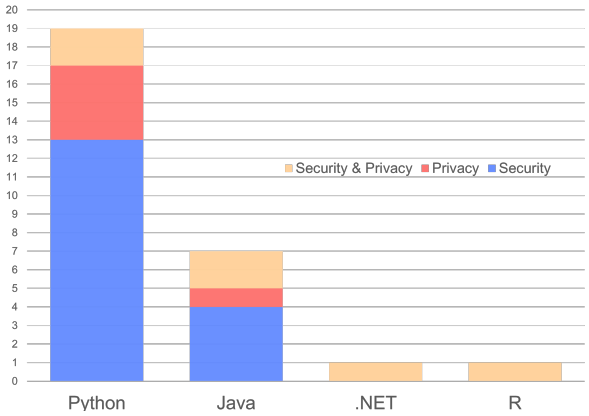


Fig. 8: Most used programming languages.

As further analysis, we also summarize the specific technologies and programming languages adopted to implement the approaches of the primary studies[1].

Regarding the exploited technologies, as summarized in Figure 8, the analysis reveals that Python is the most prevalent programming language, being employed in 19 [2, 7–11, 14–16, 30, 33–35, 38, 41–45] out of the 45 research papers, indicating

[1]It is important to note that we report statistics about those papers that explicitly provide implementation details.

its widespread popularity and versatility in academic research. Java closely follows with 7 occurrences [1, 5, 12, 13, 17, 25, 37], showcasing its enduring presence in the field. Additionally, .NET and R were each used in 1 paper, [1] and [3] respectively, demonstrating their limited but still relevant applications in specific contexts.

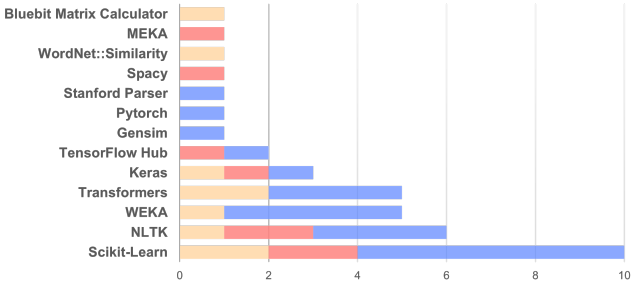


Fig. 9: The most utilized frameworks for the identification of security and privacy are represented in orange, privacy-specific frameworks in red, and security-specific frameworks in blue.

Figure 9 provides an overview of the frequency of usage for different frameworks encountered during the review analysis. Among the Python libraries and frameworks, Scikit-Learn emerges as the most commonly utilized, featuring in 10 papers [2, 7, 8, 11, 30, 34, 42–45], emphasizing its importance in machine learning and data analysis studies. Other prominent libraries include NLTK, which appears in 6 papers [7, 8, 10, 16, 34, 43], indicating its relevance in NLP research, and Transformers, which is mentioned in 5 papers [2, 6, 14, 33, 35], highlighting its significance in natural language understanding tasks. Keras, TensorFlow Hub, and Pytorch are widely adopted, appearing in 3 [2, 8, 43], 2 [9, 16], and 1 paper [15], respectively. Gensim [42], Spacy[8], and Stanford Parser [34] are each employed in 1 paper, underscoring their specific roles in linguistic analysis and semantic similarity computation.

Regarding Java tools, WEKA emerges as a popular choice, appearing in 5 research papers [5, 13, 17, 25, 37], demonstrating its significance in data mining and machine learning tasks. Bluebit Matrix Calculator [1], MEKA [12], and WordNet::Similarity [1] are each utilized in 1 paper, showcasing their unique contributions in specific research contexts.

3.3 RQ2 - What document types are exploited to identify privacy and security requirements?

This research question centers around identifying the most commonly used types of information to detect security or privacy requirements. The following list presents insights into the frequency of occurrences, as depicted in Figure 10, for each document type across the 45 analyzed papers, and also highlights their distribution between security and privacy identification.

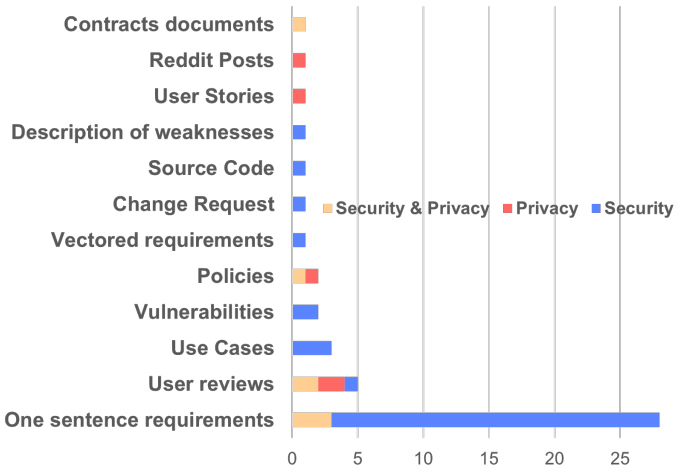


Fig. 10: Employed document types.

- *One sentence requirements* consist of requirements expressed in a concise, single-sentence format. This was the most frequently used, appearing 28 times across the analyzed papers. Among these occurrences, 3 were related to both security and privacy requirements [1, 5, 6], while the remaining 25 were specific to security requirements [13, 14, 16–20, 24, 25, 28–36, 38–44].
- *User stories* are a type of requirement document that describes the desired functionality from the perspective of end users. User stories were mentioned once in the analyzed papers, and they were specifically associated with privacy requirements [8], indicating their limited usage for security concerns.
- *Reddit posts* refer to posts or discussions on the online platform Reddit and were mentioned once, and they were specifically related to privacy requirements [9], suggesting that community discussions on platforms like Reddit can offer valuable insights into privacy concerns.
- *Contract documents* are legal agreements between parties involved in software development or usage and were mentioned once, specifically associated with both security and privacy requirements [2]. This indicates that contractual agreements play a role in addressing security and privacy concerns and can be used as a source for identifying such requirements.
- *Policies* outline guidelines, rules, or regulations that govern software development or usage. Policies were mentioned twice in the analyzed papers, with one occurrence related to both security and privacy requirements [3] and the other related to only privacy requirements [11]. This suggests that policies are considered relevant for both security and privacy concerns in the context of requirement identification.
- *Vectored requirements* previously processed were mentioned once, and it was associated with security requirements [15]. The nature and characteristics of these requirements are not discussed, but their presence in the analyzed papers indicates that preprocessed requirements may play a role in addressing security concerns.

- *Use Cases*, which describe interactions or scenarios involving system users and the software, are mentioned three times in the papers, and all occurrences were related to security requirements [21, 22, 26]. This highlights the frequent utilization of use cases for identifying security concerns during the requirements analysis process.
- *Change requests* are documents that capture proposed modifications or updates to software requirements and were mentioned once, and they were specifically associated with security requirements [23], suggesting that change management processes provide opportunities for addressing security concerns.
- *Vulnerabilities* refer to weaknesses or flaws in software systems that can be exploited by attackers and were mentioned twice in the analyzed papers, and both occurrences were related to security requirements [23, 27]. This underscores the importance of considering vulnerabilities in the identification of security requirements.
- *Source code* was mentioned once in the analyzed papers, specifically in the context of security requirements [27]. This suggests that analyzing the source code of software systems, it becomes feasible to derive models for recognizing security requirements.
- *Description of weaknesses* explicitly describes weaknesses or deficiencies in software systems. Descriptions of weaknesses were mentioned once, and they were specifically associated with security requirements [45], indicating that identifying security requirements can be informed by explicitly describing weaknesses.
- *User reviews*, which refer to feedback or opinions provided by software users, were mentioned five times in the analyzed papers. Among these occurrences, two were related to both security and privacy requirements [4, 7], two specifically addressed privacy requirements [10, 12], and one focused solely on security requirements [37]. These findings underscore the significance of user feedback in identifying both security and privacy requirements.

Thus, taking into account the above analysis, we can provide the following answer to the second research question.

Answer to RQ2

One-sentence requirements stand out as the primary source of information employed to identify privacy and security requirements. However, the analysis emphasizes the diversity of sources employed to build and validate models for identifying security and privacy aspects, such as software requirements, user feedback, source code, and policies.

3.4 RQ3 - Which datasets are employed to identify privacy and security requirements?

In the following we present the main characteristics of the datasets employed by the collected papers. The datasets are listed in Table 4 and categorized according to the specific problems addressed – namely, the recognition of security and privacy (indicated in orange), privacy (indicated in red), and security (indicated in blue) requirements.

Paper Ref.	ID	Details	Type	# of instance	Pub.
[1]	DPS-1	SmartTrip mobile app SafeDrink software system BlueWallet web service	One sentence	170 214 184	
[2]	DPS-2	20 SE contracts	Contracts	18,614	
[3]	DPS-3	Policy documents [59, 60], Google Top 1000 websites and Fortune 500	Policies	2,061	
[4]	DPS-4	Android applications reviews	User reviews	4,547,493	
[5]	DPS-5	6 NL artifacts	One sentence	10,963	✓
[6]	DPS-6	Software NFRs	One sentence	6,222	✓
[7]	DPS-7	Reviews of 17 Android apps	User reviews	64,789	✓
[8]	DP-1	22 dataset of requirements	User Stories	1,680	✓
[9]	DP-2	Posts of 66 communities	Reddit Posts	4,488,467	✓
[10]	DP-3	Investing App Reviews Food Delivery App Reviews Mental Health App Reviews	User reviews	696,073 1,708,831 204,374	✓
[11]	DP-4	Collection of privacy policies	Privacy policies	115	✓
[12]	DP-5	Reviews of 20 Apple apps Reviews of 4 Google apps	User reviews	226,797 3,480	
[13, 16, 17, 19, 25, 29, 30, 32–36, 38, 40–44]	DS-1	PROMISE dataset	One sentence	684	✓
[16]	DS-2	RE Conference's 2017 Dataset for challenge event.	One sentence	481	
[15]	DS-3	Perez-Verdejo dataset	Pre-processed	630	✓
[17, 38]	DS-4	CCHIT Ambulatory Requirements	One sentence	306	
[14, 18, 24, 25, 31, 45]	DS-5	SecReq, requirements of 3 industrial specifications	One sentence	510	✓
[20]	DS-6	Requirements from 15 projects	One sentence	3,880	
[17, 21, 22, 26]	DS-7	iTrust Use cases	Use Cases	40	
[23]	DS-8	Change requests and vulnerabilities, from Firefox version 4.0.1 to 65	Change Reques Vulnerabilities	23,885 687	
[26]	DS-9	Use cases from a module in IBMApp	Use Cases	25	
[27]	DS-10	Requirements, source code and vulnerabilities from Tomcat	Requirements Source Code Vulnerabilities	270 1,822 86	
[28]	DS-11	Axis2 web service Drools management system GeoServer geographic system	One sentence	5,751 326 8,172	✓
[37]	DS-12	KeePass password manager Mumble communication softw. WinMerge file mgmt tool	User reviews	1,000 1,000 1,000	
[39]	DS-13	PROMISE_exp, an expanded version of PROMISE (DS1)	One sentence	969	✓
[41]	DS-14	ISD-1 from Insurance domain ISD-2 from Financial Services ISD-3 from Healthcare domain	One sentence	723 686 713	
[35]	DS-15	NFR-Review NFR-SO	User reviews	1,278 17,434	✓
[45]	DS-16	Files from CWE website	Description of weaknesses	720	✓

Table 4: Catalog of used datasets.

Datasets used to identify both privacy and security requirements:

- DPS-1: contains one-sentence functional requirements extracted from the software requirements specifications of 3 experimental software systems spanning various application domains [1]. The first system, SmartTrip, is an Android mobile application that offers users real-time routing and pricing information for planning road trips across the USA. The second system, SafeDrink, aims to assist users in managing their drinking habits. The third system, BlueWallet, is a subscription-based web service that provides users with budget planning and personal finance management capabilities.
- DPS-2: consists of 20 expired SE contracts derived from 9 distinct application domains, including healthcare, automotive, finance, banking, pharmaceuticals, telecom, technology, clothing-retail, and supermarket [2]. These contracts were prepared by legal experts representing the vendor organization and multinational customers engaged in global business activities. Each contract document varies in size, ranging from 100 to 500 pages. From these contracts, a total of 18,614 sentences were extracted, with sentence lengths varying from 20 to 250 words.
- DPS-3: comprises policy documents collected from multiple sources, including previous RE works [59, 60], Google Top 1000 websites, and Fortune 500 [3]. The collection involves meticulous examination of financial privacy policies, analysis of HIPAA's impact on healthcare privacy notices, and extraction of policies from high-traffic websites and influential Fortune 500 companies.
- DPS-4: includes a comprehensive collection of 4,547,493 reviews for Android applications sourced from the Google Play Store [4], offering a rich repository of user viewpoints on various Android apps. To ensure the dataset quality and relevance, the focus was placed on reviews originating from apps with a minimum of 50,000,000 downloads. Additionally, the dataset exclusively includes English-language reviews, encompassing both the review text and rating score, along with any developer responses provided.
- DPS-5: is obtained by manually selecting six easily accessible natural language requirement artifacts from the field of electronic healthcare [5]. Despite the healthcare domain's typically strict regulations and standardization, the authors intentionally opted for diverse document types, including non-conventionally formatted feature requests, sourced from the USA and Canada. This choice offers a more comprehensive geographical scope within the dataset.
- DPS-6: consists of 6,222 requirement description sentences categorized into 32 NFRs [6]. The dataset was obtained through a multi-step process. Initially, the research group referred to Wikipedia to gain a comprehensive understanding of NFRs, selecting common categories based on their expertise. Relevant NFR description sentences were then filtered using search engines to form the initial dataset. Four students collaborated over a 20-day period to review, refine, and ensure balance in the dataset, aiming for roughly 200 sentences per NFR category.
- DPS-7: consists of reviews from 17 Android apps sourced from the Google Play Store, representing a diverse range of 15 different categories [7]. The researchers aimed to assess their approach against reviews with a wide vocabulary and discussing various security issues by selecting apps from different categories. To ensure

a comprehensive evaluation, they included 12 popular apps with numeric star ratings above 3.5 and over 1000 reviews. Additionally, they considered 5 unpopular apps with lower numeric star ratings (below 2.5) and fewer reviews, as these apps might pose more serious security concerns.

Datasets used to identify privacy requirements:

- DP-1: is a diverse set of 22 datasets, each consisting of over 50 requirements expressed as user stories [8], obtained through online sources and acquired from software companies with authorized permission for disclosure. The primary purpose of these datasets was to conduct experiments on ambiguity detection using the REVV-Light tool[2] [61].
- DP-2 is a comprehensive collection of Reddit posts from 66 communities centered around popular software products like WhatsApp, Telegram, Airbnb, and Instagram. The dataset consists of a staggering 6,052,258 posts and was acquired from Pushshift using its user-friendly API, designed for efficient data analysis [9]. To ensure data quality, the posts were filtered based on their creation date, retaining those posted between the subreddit's inception and December 2021. Duplicate posts and those deleted or removed by authors or moderators were excluded, resulting in a substantial dataset of 4,488,467 posts, ready for analysis and study.
- DP-3: comprising user reviews obtained from various categories of mobile applications (Investing, Food Delivery, and Mental Health Apps) and specifically focusing on gathering user reviews as a form of requirement [10]. User reviews serve as a valuable source of feedback and insights regarding privacy concerns from the standpoint of app users. Including 696,073 reviews for investing apps, 1,708,831 reviews for food delivery apps, and 204,374 reviews for mental health apps, this dataset provides a rich source of information for understanding user perspectives and privacy-related issues in the respective app categories.
- DP-4: consisting of a collection of privacy policies sourced from different entities, highlighting the organizations' commitments and practices related to data handling (named OPP-115 corpus) [11]. Annotated by 10 law students, the dataset comprises 115 privacy policies and includes 2,831 annotations specifically focused on the practices examined in this research.
- DP-5: comprising user reviews of 20 apps from the Apple App Store and 4 apps from the Google Play Store [12]. The apps were selected based on their wide range of categories and a significant number of user reviews, by employing a simple web crawler. A total of 226,797 one-star and two-star user reviews were downloaded from the Apple App Store, while 3,480 one-star and two-star user reviews were obtained from the Google Play Store. The difference in the number of user reviews between the two app stores is attributed to the distinct data collection APIs provided by Apple and Google during the study period.

Datasets used to identify security requirements:

- DS-1 (PROMISE): despite its widespread usage and relevance in the field of security requirements identification [13, 16, 17, 19, 25, 29, 30, 32–36, 38, 40–44],

[2] https://github.com/RELabUU/revv-light

contains a total of 625 requirement sentences. Among these sentences, 255 are classified as functional requirements, while the remaining 370 are categorized as non-functional requirements. The non-functional requirements are further labeled with eleven distinct categories, including availability, legal, look and feel, maintainability, operational, performance, scalability, security, usability, fault tolerance, and portability. However, it should be noted that the dataset is imbalanced, as most of the requirements fall under the functional category.

- DS-2 [16]: is a collection of data specifically curated for the data challenge event held during the 2017 International Requirements Engineering Conference, designed as a realistic and comprehensive dataset for researchers and practitioners to analyze, evaluate, and develop solutions for requirements engineering tasks and challenges.
- DS-3 (Perez-Verdejo) [15]: consists of 630 vectored requirements preprocessed through several steps. The initial steps involved the removal of special characters and conversion of text to lowercase. Punctuation marks, including periods, commas, semicolons, and others, were subsequently excluded. Lemmatization was applied to derive word base forms by replacing suffixes, followed by the elimination of stop words. Finally, vectorization was performed using the TF-IDF technique.
- DS-4 (CCHIT Ambulatory Requirements) [17, 38]: comprising 306 natural language requirements, i.e., 228 functional requirements (FR) and 78 non-functional requirements (NFR) along with other types. Notably, within the CCHIT dataset, a single requirement statement can encompass multiple NFR types.
- DS-5 (SecReq) [14, 18, 24, 25, 31, 45]: consisting of one-sentence requirements, from three industrial specifications, this dataset has been widely used in papers focusing on security requirements. The first specification, Common Electronic Purse (ePurse), includes 177 labeled entries. The second specification, Customer Premises Network (CPN), contains 124 labeled entries. Lastly, the Global Platform Specification (GPS) comprises 210 labeled entries. Each requirement can have one of two labels or classes: sec (security) and non-sec (non-security).
- DS-6 [20]: includes one-sentence requirements obtained from a collection of projects with diverse characteristics. The considered documents originate from various sources, including industrial and commercial projects, domain-specific guidelines, research-driven creations, and student-generated materials. Additionally, the documents encompass different types of specifications, such as System Requirement Specifications (SRS), Requests for Proposals (RFP), non-technical requirement documents (NTR), and Backlog Items (BL) for specific products. The requirements contained within these documents are written at various levels, including the domain, user, and system levels. Consequently, there is significant variation in the level of detail and length (word count) of the requirements.
- DS-7 (iTrust): consists of 40 use cases with non-functional requirements, constraints, and a glossary. There is no precise information on this dataset as each work uses certain parts of the documentation, often indicating different numbers and characteristics [17, 21, 22, 26].
- DS-8 (Firefox) [23]: encompasses a diverse range of data, including change requests and vulnerabilities, collected from Firefox versions 4.0.1 to 65. This dataset provides valuable insights into the integration of security requirements and the identification

of vulnerabilities within the Firefox browser. It consists of 23,885 random requirements and 687 vulnerabilities, all of which are carefully linked to their respective source code files.

- DS-9 (IBMApp) [26]: consisting of 25 use cases extracted from a financial module within a proprietary IBM enterprise application.
- DS-10 (Tomcat) [27]: facilitates a thorough exploration of security dimensions inherent to the Tomcat server software. As Tomcat is widely adopted in the industry, this dataset contains valuable requirements, source code, and vulnerabilities related to the software. It leverages the detailed security report of Tomcat and the Common Vulnerability Database (CVE) to provide accurate linking between vulnerabilities and their corresponding code files. This dataset serves as a valuable resource for conducting in-depth security-related research and analysis on Tomcat.
- DS-11 [28]: includes the security requirements of diverse software systems spanning multiple domains, namely, Axis2, Drools, and GeoServer. Axis2, a web services engine, has received continuous support from the Apache Software Foundation since August 2004. Drools, developed by Red Hat, operates as a business rule management system. Lastly, GeoServer empowers users to collaborate on editing and sharing geospatial information within a geographic system. By integrating the requirements of these systems, the dataset enables a comprehensive exploration of security requirements across various domains.
- DS-12 [37]: is formed from categorized user queries that were gathered from the user requests forum within an open-source project encompassing three distinct software applications: KeePass, Mumble, and WinMerge. Specifically, KeePass functions as a portable password manager categorized under Business & Enterprise/Office/Business, Mumble serves as Communication/Internet Phone software tailored for low-latency, high-quality voice communication during online gaming, and WinMerge operates as a System Administration/Storage/File Management utility, facilitating folder and file comparison and merging.
- DS-13 (PROMISE_exp) [39]: is the expanded version of the PROMISE dataset including 444 functional requirements and 525 non-functional requirements. It should be noted that the distribution of the 11 types of non-functional requirements among the sentences is unbalanced.
- DS-14 (ISD) [41]: obtained by manually extracting and classifying three sets of NFR statements from 10, 9, and 11 Software Requirements Specification (SRS) documents, respectively. The three sets (named ISD-1, ISD-2, and ISD-3) refers to the Insurance, Banking and Financial Services, and Healthcare domains, respectively.
- DS-15 [35]: consists of two sets named NFR-Review and NFR-SO. NFR-Review comprises 1278 user review sentences from two popular apps, i.e, iBooks and WhatsApp, categorized into five NFR categories: security, portability, performance, reliability, and usability. NFR-SO contains NFR obtained by crawling the technical forum StackOverflow, tagged with one of the seven NFR categories: availability, performance, maintainability, portability, scalability, security, and fault-tolerance.
- DS-16 (CWE) [45]: obtained by first acquiring XML formatted data files from the Common Weakness Enumeration (CWE) website, and then parsing the XML files to extract the summary and extended descriptions of 720 weaknesses.

Thus, from the above analysis, we can provide the following answer to the third research question.

Answer to RQ3

Each study focusing on privacy aspects utilizes a unique dataset, highlighting the diverse range of data sources explored by researchers. This observation remains valid for works that encompass both security and privacy, demonstrating the broadened perspective taken by some scholars. On the other hand, when it comes to security-focused investigations, there is a tendency to rely on widely used datasets such as DS-1 (PROMISE), DS-5 (SecReq), and DS-7 (Trust). Nevertheless, even within the realm of security, multiple options are available, ensuring a plethora of choices for researchers and practitioners in this domain.

3.5 RQ4 - Which are the most effective methods for identifying privacy and security requirements?

From Figure 10, resuming the employed document types for privacy and security detection, it can be inferred that there are different sources that researchers and practitioners can exploit when it comes to analyzing certain types of requirements, such as one sentence requirements, user reviews, and use cases. Table 5 summarizes the top-performing approaches identified in our SLR starting from the document type they focus on, categorized by the problem addressed (i.e., the identification of security and privacy in orange, privacy in red, and security in blue). Below, we provide a concise description of each approach, emphasizing their distinctive characteristics compared to the work identifying security and/or privacy requirements from the same document type.

One sentence requirements

Li et al. [6] present the most effective approach, known as NFRNet, for the multiclass classification of both security and privacy from one sentence requirements. It is a neural network composed of two main parts. The first part is an improved BERT word embedding transformer used for representing the requirements, while the second part is a BiLSTM network that captures the contextual information of the requirements. NFRNet is used to classify requirements into 32 NFR categories. The evaluation carried out leveraging software NFRs in DPS-6 reveals impressive results with precision of 91%, recall of 92.6%, and F1-score of 91.5%.

In the domain of security one sentence requirements, a notable work introduces PRCBERT, a prompt learning approach leveraging BERT pre-trained language models. In particular, *Luo et al.* [35] utilize Python and Transformers to implement PRCBERT. By utilizing the sentence requirements from DS-1 and DS-14 datasets for evaluation, their approach attains an impressive F1-score of 98%.

ID	Document Type	Task	Approach	Dataset	Evaluation Metrics	Result
[6]	One sentence	Multiclass Classification	NFRNet	DPS-6	Precision Recall F1-Score	91% 92.6% 91.5%
[35]	One sentence	Multiclass Classification	PRCBERT	DS-1 DS-15	F1-Score	98%
[21]	Use Cases	Binary Classification	Knowledge based	DS-7	Accuracy FP rate FN rate	98% 0% 14%
[22]	Use Cases	Binary Classification	Combined SL	DS-7	Precision Recall F1-Score	87% 91% 89%
[4]	User reviews	Binary Classification	SVM	DPS-4	AUC	93%
[7]	User reviews	Summarization	SRR-MIner	DPS-7	F1-Score MCC	89% 87%
[10]	User reviews	Summarization	LDA	DP-3	Precision	~75%
[37]	User reviews	Binary Classification Multiclass Classification	SVM	DS-12	Accuracy Precision Recall F1-Score	69,3% 81.1% 62.8% 70.8%
[23]	Change Request Vulnerabilities	Multilabel Classification	Soft Ensemble	DS-8	Precision Recall	13.6% 64.1%
[27]	Requirements Source Code Vulnerabilities	Multiclass Classification	LSI	DS-10	Precision Recall F2-Score	50% 70% 18%

Table 5: Most effective approaches, categorized by document type and problem addressed.

Use Cases

One of the most promising work on security problem identification, starting from use cases, proposes a knowledge-based approach that involves extracting security-relevant knowledge from the heuristic findings achieved by using the security assessment approach [21]. In particular, *S. Gärtner et al.* exploit this knowledge-based approach by proposing an extraction method that guides the requirements engineer to refine terms and properties and add new ones to the security knowledge. Through an assessment on the iTrust dataset (DS-7), the authors demonstrate that their knowledge-driven approach, incorporating 55 distinct use cases, attains an accuracy of 98%, with a false positive rate of 0% and a false negative rate up to 14%.

Always on security requirements identification, *Slankas et al.* [22] take advantage of the Stanford Type Dependency Representation (STDR), employing a combination of supervised learning and a majority voting approach involving KNN, NB, and SVM classifiers. Validated on the iTrust dataset (DS-7), the approach achieves good precision, recall, and F1-score, with the combined supervised learning approach achieving 87%, 91%, and 89%, respectively.

Therefore, when examining the studies that utilize use cases as document types for security requirement identification, we face challenges in comparing the effectiveness

of the previous two approaches. This limitation arises due to the differing metrics employed to evaluate them, rendering a direct comparison unfeasible and hindering our ability to determine the most effective approach.

User Reviews

For security and privacy requirements identification from user reviews, one of the best approaches is proposed by *Nguyen et al.* [4]. It includes using SMOTE (Synthetic Minority Over-sampling Technique) for handling imbalanced data, Bag-of-Words (BOW) for requirements representation, and SVM as a classification algorithm. The approach is evaluated on the DPS-4 dataset obtaining an area under the receiver operating characteristic curve (AUC) of 93%. Another promising approach is in the realm of user review summarization, where *C. Tao et al.* [7] have explored techniques like Bag-of-Words (BOW), TF-IDF, LR, and SRR-Miner. During the evaluation on DPS-7, various metrics, including F1-Score and MCC, were used to assess performance. Particularly noteworthy is the SRR-Miner technique, which showcased promising results by achieving an impressive 89% F1-Score and 87% MCC in effectively addressing security-related reviews. As with the works on identifying security problems from use cases, we are unable to select the best-performing approach because of the different metrics adopted for evaluation.

On the privacy concerns detection from user reviews, *F. Ebrahimi et al.* [10] extract meaningful privacy-related insights with a combination of techniques, including Hybrid TF-IDF, GloVe embeddings, and Latent Dirichlet Allocation (LDA). These methods have been applied to analyze diverse app reviews, ranging from Investing Apps and Food Delivery Services to Mental Health Apps (DP-3). The achieved precision, which stands at approximately 75%, demonstrates the efficacy of this approach in distilling valuable information from the intricate landscape of user-generated content.

In the realm of security identification, the study by *C. Li et al.* [37] explores the utility of Binary Classification and Multiclass Classification techniques to assess various NFR aspects, including security. Leveraging methodologies such as Word Unigram, TF-IDF, SVM, NB, and KNN, the authors employ the Java-based WEKA framework for their investigation. Their analysis spans DS-12 datasets, including user reviews from applications such as KeePass, Mumble, and WinMerge. Notably, the SVM technique exhibits accuracy rates of 69.3%, 81.1%, and 62.8%, alongside a commendable F1-Score of 70.8%, indicating that their approach in detecting security software requirements is quite good.

Vulnerabilities

In the context of identifying security requirements, *S. Imtiaz et al.* [23] focus on multilabel classification using techniques such as SMOTE, Edited Nearest Neighbor (ENN), SVM, LR, Naive Bayes (NB), Random Forest (RF), KNN, ANN, Decision Tree (DT), Soft Ensemble, and Hard Ensemble. The study utilizes the DS8 dataset (Firefox) and evaluates the performance based on Precision and Recall metrics. Among the evaluated techniques, the Soft Ensemble approach stands out with a precision of 13.6% and a recall of 64.1% in addressing the identified problem of security requirements.

Another work by *S. Imtiaz et al.* [27] employs a multiclass classification approach focusing on NFRs based on various document types, including requirements, source code, and vulnerabilities. Two prominent techniques, namely the Vector Space Model (VSM) and Latent Semantic Indexing (LSI), are harnessed for feature representation and dimensionality reduction. The experimentation employs the DS-10 dataset (Tomcat) and evaluates model performance using metrics such as precision, recall, and F2-Score. Particularly, the LSI technique exhibits notable outcomes with precision ranging from 0.3% to 50%, recall ranging from 0.6% to 70%, and an F2-Score varying between 2% and 18%.

Despite using the same metrics, we report both studies on vulnerabilities because having reported a range of values in [27], we cannot decide which of the two approaches is the most effective. It appears that addressing this problem through these document-based approaches is quite challenging, as both methods have shown low performance. This may be due to the fact that vulnerabilities are not typically documented during the requirements gathering process, but are instead identified through information extracted from software that is already operational.

Answer to RQ4

The analysis do not allow to conclude that one approach performs better than another in all the cases. Indeed, the achieved results clearly depend on the type of document employed for the identification, the particular task, and the available dataset. However, we can note that when one-sentence requirements are available, multi-class classification tends to yield superior performance.. As a final observation, we can highlight that for more complex document types (i.e., user reviews or use cases) the problem should be addressed as a binary classification.

4 Discussion & Implications

In the following we summarize for each request question addressed in our study some lesson learned and implications for researchers and practitioners.

Research Question 1

How is the problem of identifying the privacy and security requirements modeled and which are the methods exploited?

Our SLR has revealed several key insights and implications regarding the way the problem of identification of privacy and security requirements is modeled. The diverse modeling approaches used in this field illustrate the complexity and multi-faceted nature of the problem. Multi-class and binary classification are predominant, but innovative methods such as multi-label classification, summarization, semantic similarity, text identification, one-class classification, and clustering are also employed

(e.g., [1, 2]). This variety suggests that a single universal approach is insufficient, highlighting the need for a tailored solution based on specific project requirements.

Python has emerged as the dominant programming language in this domain, largely because of its extensive libraries and frameworks like Scikit-Learn, NLTK, and Transformers, which simplify the implementation of ML and NLP tasks [8]. Java also holds a significant position, particularly with tools like WEKA being popular for data mining and ML applications [7]. This preference illustrates the practical and versatile nature of these languages in both academic and industry researches and their ability to support complex analytical tasks effectively.

Prominent techniques like SVM, TF-IDF, and Naive Bayes remain widely utilized due to their effectiveness and reliability [3, 4]. However, the increasing adoption of advanced methods such as BERT and Word2Vec indicates a shift towards leveraging deeper, more contextualized embeddings for requirement identification tasks (e.g., [5, 6]). This trend underscores the impact of integrating cutting-edge NLP techniques to enhance the accuracy and performance of requirement identification systems.

The implications of these findings are manifold. Researchers are encouraged to explore various modeling approaches to capture the complexity of requirement identification tasks fully. For practitioners, this means having access to a wide array of solutions tailored to different needs, from general non-functional requirements (NFRs) identification to specific privacy and security requirements. This flexibility allows practitioners to choose the most suitable method based on the specific context and requirements of their projects [9–12].

Furthermore, the identified techniques such as TF-IDF, SVM, and NB serve as effective starting points for researchers entering this field. These well-established methods provide a reliable foundation upon which new, more advanced techniques can be built and compared. Embracing these techniques ensures that new research builds on proven methodologies, enhancing both robustness and comparability.

The widespread use of Python highlights its suitability for requirements identification tasks. Researchers and practitioners are suggested to leverage the rich ecosystem of Python libraries to streamline their requirements processes. This adoption not only simplifies implementation but also enhances collaboration and reproducibility within the interested communities. Additionally, the use of Python facilitates the transition from research to practical application, benefiting requirements engineers in industry settings [13, 16].

Finally, understanding the prevalent programming languages, libraries, and tools can significantly enhance collaboration and knowledge sharing among researchers, practitioners, and educators. This leads in designing studies that are compatible with widely used tools and methods, ensuring that research findings are more easily adopted and compared across different studies. It also aids in selecting techniques that align with the current state of the art, thus improving the relevance and impact of new research. Promoting collaboration through standardization on common platforms and methodologies simplifies the sharing of data, tools, and findings, fostering a collaborative environment that accelerates advancements in the automatic identification of privacy and security requirements.

In the end, these insights are crucial for guiding future research directions and fostering a collaborative environment that accelerates advancements in the automatic identification of privacy and security requirements. By integrating these lessons learned and their implications into their work, researchers and practitioners can drive the field forward, enhancing the impact, applicability and shareability of their findings.

Research Question 2

What document types are exploited to identify privacy and security requirements?

Our SLR provides interesting insights regarding the types of documents commonly used to identify privacy and security requirements. The findings underscore the significance of various document types and their respective roles in enhancing the identification process for both security and privacy requirements.

In general, the most commonly used type of source information in the analyzed studies is one-sentence requirements. This prevalence highlights their utility in requirement identification, as evidenced by their frequent appearance across multiple studies [1, 5, 6]. One-sentence requirements are particularly effective due to their concise nature, which simplifies the identification process.

Privacy issues are primarily identified through documents representing users' perspectives, such as user reviews, user stories, and Reddit posts. These sources provide rich insights into privacy concerns from the end-users' viewpoint, thereby facilitating the development and validation of models aimed at privacy requirement identification [8, 9]. This approach underscores the importance of considering user feedback and perspectives in understanding and addressing privacy requirements.

In contrast, models for security requirement identification typically start from the description of the systems, including one-sentence requirements and use cases, or from identifying system imperfections, such as vulnerabilities, descriptions of weaknesses, source code, and change requests [23, 27, 45]. These document types provide critical insights into potential security risks and areas of improvement, emphasizing the need to integrate various sources of information to build robust security models.

As for the implications,researchers and practitioners should consider one-sentence requirements as a primary source for building and validating models to capture security and privacy requirements since this document type is of widespread usage andits effectiveness make it a reliable starting point for new studies and practical applications.

Exploring user stories and online community discussions, such as Reddit posts, can offer valuable insights into privacy concerns. These sources provide a direct line to user perspectives, which is crucial for accurately identifying and addressing privacy requirements. Researchers and engineers should consider integrating these document types into their methodologies to enhance privacy requirement identification.

The presence of use cases, change requests, source code analysis, descriptions of weaknesses, and vulnerabilities underscores their importance in identifying security requirements. Leveraging these document types can significantly enhance the process

of identifying and mitigating security risks. For example, starting from existing documentation, analyzing source code, and addressing known problems from previous projects can provide a comprehensive approach to security requirement identification.

In general, for practitioners, adopting a holistic approach that integrates various document types and analysis methods is essential to capture the complex nature of both security and privacy concerns. From one side, user feedback through reviews, stories, and online community discussions to gain valuable insights into privacy issues. On the other side, utilizing diverse document types such as one-sentence requirements, use cases, change requests, and source code can significantly enhance the identification process for security requirements. Thus, practitioners are encouraged for cross-disciplinary collaboration between security and privacy experts to develop integrated solutions that address both types of requirements comprehensively.

For researchers, exploring integrated techniques that consider the interdependencies between security and privacy requirements is an interesting research direction. Methodological innovations that combine different document types and analysis techniques can improve the accuracy and comprehensiveness of requirement identification. Placing a greater emphasis on user-centered approaches, particularly for privacy requirements, can lead to the development of models that better reflect user needs and enhance trust in software systems, allowing, additionally, the developmente of standardized frameworks and benchmarks for evaluating and identifying both security and privacy requirement, facilitating the comparison of approaches and promote best practices and solutions in the field.

Research Question 3

Which datasets are employed to identify privacy and security requirements?

With our SLR we provided a comprehensive overview of the datasets employed to identify privacy and security requirements, revealing that various datasets are utilized to discern privacy, security, or a combination of both within requirements. This diversity underscores the importance of leveraging different data sources to capture the nuanced aspects of security and privacy concerns in software systems.

One key lesson learned is the frequent use of specific datasets such as PROMISE (DS-1), SecReq (DS-5), and iTrust (DS-7) in the identification of security requirements. These datasets have become benchmarks in the field, enabling researchers to evaluate the effectiveness of new approaches against well-established standards. For instance, the PROMISE dataset is widely used due to its comprehensive nature, encompassing both functional and non-functional requirements across various categories, despite its imbalance favoring functional requirements [13, 16, 17]. Similarly, the SecReq dataset, with its labeled entries from industrial specifications, provides a robust foundation for security requirements analysis [14, 18, 24]. The iTrust dataset, consisting of use cases with non-functional requirements, is another commonly used resource that facilitates the identification of security issues through practical, real-world examples [17, 21, 22].

In contrast, privacy requirement identification lacks commonly used datasets, leading to a reliance on unique datasets tailored to specific studies. For example, the

DP-1 dataset, comprising user stories, is used to explore privacy concerns in different projects scenario [8]. The DP-2 dataset, consisting of Reddit posts, provides insights into privacy issues discussed in online communities, highlighting the value of user-generated content in privacy requirement analysis [9]. Other datasets, such as those containing user reviews or privacy policies, offer rich sources of user feedback and organizational practices related to privacy [10–12].

The use of industry-specific datasets also emerged as a significant trend. Studies focusing on specific industries or projects extract requirements from contracts, use cases, and domain-specific datasets, emphasizing the need to consider industry-specific requirements in security and privacy analysis. For instance, the DPS-2 dataset derived from SE contracts in various domains highlights the importance of legal documents in understanding security and privacy requirements [2]. Similarly, datasets like the DPS-3 and DPS-4, which include user reviews from specific app categories and privacy policies, underscore the relevance of contextual data in privacy analysis [3, 4].

The implications of these findings are multiple. The lack of commonly used datasets for privacy requirements suggests a need for the development and sharing of standardized datasets to enable comparative analyses and benchmarking of the approaches. Researchers and practitioners in the field of security requirements, instead, can benefit from the established datasets like PROMISE and SecReq, which provide a solid foundation for implementing and evaluating new approaches with older ones. Furthermore, the use of industry-specific datasets allows researchers to gain insights into the unique requirements and challenges of specific domains, leading to more contextual and relevant analyses of security and privacy issues.

Future research should aim to bridge the gap between security and privacy requirement identification by defining and validate datasets that address both privacy and security aspects, to enable the design of effective and efficient approaches. This integration is crucial for developing comprehensive models that consider the interdependencies between security and privacy concerns in software systems. Additionally, researchers should explore the creation of publicly available datasets that encompass a broad range of privacy requirements to facilitate more robust and comparative studies.

Research Question 4

Which are the most effective methods for identifying privacy and security requirements?

Our SLR finally explores the most effective methods for identifying privacy and security requirements, emphasizing the importance of selecting appropriate methods based on the type of document, classification task, and datasets used.

Results reveal that multiclass classification-based approaches are extensively adopted for identifying privacy and security requirements, particularly when dealing with one-sentence requirements. For instance, *Li et al.* [6] proposed NFRNet, a neural network-based approach that effectively classifies requirements into 32 NFR categories using a combination of BERT embeddings and BiLSTM networks. This method achieved high precision (91%), recall (92.6%), and F1-Score (91.5%) when evaluated on the DPS-6 dataset.

In the realm of security, *Luo et al.* [35] introduced PRCBERT, a prompt learning approach leveraging BERT language models. By using requirements from DS-1 and DS-15 datasets, this approach attained an impressive F1-Score of 98%, highlighting its efficacy in identifying security requirements from one-sentence documents.

When dealing with more complex document types, such as use cases and user reviews, binary classification has shown promising results. *Gärtner et al.* [21] utilized a knowledge-based approach to extract security-relevant concerns from use cases, achieving an accuracy of 98% on the iTrust dataset (DS-7). *Slankas et al.* [22] employed a combination of supervised learning techniques, achieving good performance with precision (87%), recall (91%), and F1-Score (89%) on the same dataset.

For user reviews, *Nguyen et al.* [4] demonstrated the effectiveness of SVM combined with SMOTE and Bag-of-Words for handling imbalanced data, achieving an AUC of 93% on the DPS-4 dataset. *Tao et al.* [7] explored summarization techniques like SRR-Miner, achieving an F1-Score of 89% and MCC of 87% on the DPS-7 dataset. However, the performance of techniques varies significantly depending on the classification task and dataset used, vene though the document type is the same. For example, while *Ebrahimi et al.* [10] achieved a precision of approximately 75% using LDA for summarizing privacy-related insights from diverse app reviews (DP-3), *Li et al.* [37] found that SVM achieved varying accuracy rates and an F1-Score of 70.8% for security requirements using user reviews from DS-12.

In the context of identifying security requirements from vulnerabilities, *Imtiaz et al.* [23] focused on multilabel classification using a Soft Ensemble approach, achieving a precision of 13.6% and recall of 64.1% on the DS-8 dataset (Firefox). Another study by the same authors [27] used Latent Semantic Indexing (LSI) for multiclass classification on the DS-10 dataset (Tomcat), with varying precision (0.3%-50%), recall (0.6%-70%), and F2-Score (2%-18%).

The identified best-performing approaches provide benchmarks for future research and practice in the field of security and privacy requirement identification. Practitioners can leverage efficient solutions, particularly multiclass classification, when defining system requirements as one sentence during the early stages of development. These approaches, such as NFRNet and PRCBERT, offer high performance and can serve as reference points for evaluating new methods.

For more complex document types like use cases and user reviews, binary classification techniques have shown the highest results, suggesting the consideration of binary classification approaches when dealing with articulated document types to enhance the identification of security and privacy requirements.

The results also indicate that different approaches excel in different classification tasks and requirement types. This opens avenues for exploring hybrid or combined approaches that leverage the strengths of multiple methodologies. Future research could investigate the integration of techniques into hybrid models, from multiclass and binary classification tasks, to enhance the overall accuracy and performance of security and privacy requirement identification.

To facilitate the comparison of different approaches and promote best practices in the field, there is a need to develop standardized evaluation metrics and benchmarks. Future studies should focus on establishing common evaluation frameworks that allow for consistent and reliable assessment of different methods across various datasets.

5 Threats to validity

The SLR conducted in this study follows a well-defined protocol to ensure accuracy and objectivity in the search process. However, it is important to acknowledge potential limitations that may arise. One such limitation is the possibility that the search string used in the study may not encompass all existing synonyms for the term "Privacy and Security in Requirements Engineering", which could result in the omission of relevant studies. To address this concern, we incorporated synonyms for the key constructs to broaden the scope of the search and minimize the risk of missing relevant literature.

Construct validity, which pertains to the generalization of results to the underlying concept or theory, is another aspect to consider [62]. To mitigate the risk of personal bias, the SLR was conducted by a Ph.D. student (i.e., the first author), and two graduate professors with expertise in software engineering validating the process. This approach also helps ensure the internal validity of the study, which is concerned with potentially incorrect conclusions about causal relationships between treatment and outcome.

External validity, on the other hand, focuses on the representativeness of primary studies to the review topic [62]. In the context of a literature review, the extent to which the identified literature is externally valid directly impacts the synthesis of its content. Therefore, to ensure the replicability of the work, we have made all files available in the appendix[3]. In this SLR, grey literature papers were excluded, which may affect the external validity to some extent.

Conclusion validity, which addresses the risk of excluding relevant studies, was carefully considered in the research protocol. The authors designed and validated the protocol to minimize the possibility of overlooking important studies. Furthermore, to enhance the coverage of potentially relevant studies, various synonyms for the research constructs were used, ensuring a comprehensive approach to the automatic search.

It is worth noting that the categorization of threats presented by *Wohlin et al.* [62] was utilized to guide the categorization process in this SLR. This categorization helps provide a structured framework for organizing and analyzing the identified studies.

Overall, by following a rigorous mapping protocol, addressing potential limitations, and incorporating expert validation, this SLR aims to provide a reliable and comprehensive understanding of the automated identification of privacy and security requirements.

6 Conclusions

This paper presents a SLR on the automated identification of privacy and security requirements throughout the entirety of the software development lifecycle. The study aims to gain insights into the employed methods, document types, datasets, and techniques. By addressing four research questions, this SLR provides a comprehensive understanding of the current state of research and offers valuable insights.

[3] https://drive.google.com/drive/folders/1OOm55P-iR-nSSpm0PELCMhE8q9--YzcM?usp=sharing

The identified methods used in various application domains shed light on the modeling and technique choices made by researchers. The analysis of adopted document types highlights the diversity of sources used to build and validate models for recognizing security and privacy requirements. The catalog of employed datasets, organized by application domain, facilitates replication and comparison, and informs practitioners about available data sources. Lastly, showcasing the best-performing approaches organized by task and document type offers a comprehensive snapshot of the most efficient and effective approaches.

Through the SLR, several lessons learned and their implications were identified. The variation in performance across different approaches underscores the importance of benchmarking and evaluating new techniques against established methods. The successful application of Transfer Learning for privacy requirements highlights the potential of leveraging pre-trained models and knowledge from related domains. The significance of dataset and model selection emphasizes the need for careful consideration to ensure accurate identification of security and privacy requirements. These lessons provide valuable guidance for future research and practice in the field.

While this SLR provides comprehensive insights, it is essential to consider certain limitations. The inclusion criteria may have inadvertently excluded relevant studies, and the analysis was restricted to the selected primary studies. Furthermore, the SLR focused on the literature published between 2000 and 2022, potentially overlooking recent developments. These limitations open avenues for future research to address these gaps and further advance the understanding of privacy and security requirements identification.

The findings suggest a great deal of potential for future research in this area and that the field of automatic identification of security and privacy requirements is ripe for further exploration using AI-based techniques. With the growing interest and increasing number of high-quality publications, researchers in this field should seize this exciting opportunity to explore new insights and innovations.

Primary studies

[1] A. Mahmoud and G. Williams, "Detecting, classifying, and tracing non-functional software requirements," *Requirements Engineering*, vol. 21, pp. 357–381, 2016.

[2] A. Sainani, P. R. Anish, V. Joshi, and S. Ghaisas, "Extracting and classifying requirements from software engineering contracts," *2020 IEEE 28th International Requirements Engineering Conference (RE)*, pp. 147–157, 2020.

[3] A. K. Massey, J. Eisenstein, A. I. Antón, and P. P. Swire, "Automated text mining for requirements analysis of policy documents," *2013 21st IEEE International Requirements Engineering Conference (RE)*, pp. 4–13, 2013.

[4] D. C. Nguyen, E. Derr, M. Backes, and S. Bugiel, "Short text, large effect: Measuring the impact of user reviews on android app security & privacy," *2019 IEEE Symposium on Security and Privacy (SP)*, pp. 555–569, 2019.

[5] M. Riaz, J. T. King, J. Slankas, and L. A. Williams, "Hidden in plain sight: Automatically identifying security requirements from natural language artifacts," *2014 IEEE 22nd International Requirements Engineering Conference (RE)*, pp. 183–192, 2014.

[6] B. chuan Li and X. Nong, "Automatically classifying non-functional requirements using deep neural network," *Pattern Recognit.*, vol. 132, p. 108948, 2022.

[7] C. Tao, H. Guo, and Z. Huang, "Identifying security issues for mobile applications based on user review summarization," *Inf. Softw. Technol.*, vol. 122, p. 106290, 2020.

[8] F. Casillo, V. Deufemia, and C. Gravino, "Detecting privacy requirements from user stories with nlp transfer learning models," *Inf. Softw. Technol.*, vol. 146, p. 106853, 2022.

[9] Z. S. Li, M. Sihag, N. N. Arony, J. B. Junior, T. B. Phan, N. A. Ernst, and D. E. H. Damian, "Narratives: the unforeseen influencer of privacy concerns," *2022 IEEE 30th International Requirements Engineering Conference (RE)*, pp. 127–139, 2022.

[10] F. Ebrahimi and A. Mahmoud, "Unsupervised summarization of privacy concerns in mobile application reviews," *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, 2022.

[11] S. Zimmeck, Z. Wang, L. Zou, R. Iyengar, B. Liu, F. Schaub, S. Wilson, N. M. Sadeh, S. M. Bellovin, and J. R. Reidenberg, "Automated analysis of privacy requirements for mobile apps," in *Network and Distributed System Security Symposium*, 2017.

[12] S. McIlroy, N. Ali, H. Khalid, and A. Hassan, "Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews," *Empirical Software Engineering*, vol. 21, pp. 1067–1106, 2015.

[13] R. Jindal, R. Malhotra, and A. Jain, "Automated classification of security requirements," *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2027–2033, 2016.

[14] V. Varenov and A. Gabdrahmanov, "Security requirements classification into groups using nlp transformers," *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*, pp. 444–450, 2021.

[15] D. A. López-Hernández, E. Mezura-Montes, J. O. Ocharán-Hernández, and Á. J. Sánchez-García, "Non-functional requirements classification using artificial neural networks," *2021 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*, vol. 5, pp. 1–6, 2021.

[16] C. Baker, L. Deng, S. Chakraborty, and J. Dehlinger, "Automatic multi-class non-functional software requirements classification using neural networks," *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 2, pp. 610–615, 2019.

[17] J. Slankas and L. A. Williams, "Automated extraction of non-functional requirements in available documentation," *2013 1st International Workshop on Natural Language Analysis in Software Engineering (NaturaLiSE)*, pp. 9–16, 2013.

[18] A. Kobilica, M. Ayub, and J. Hassine, "Automated identification of security requirements: A machine learning approach," *Proceedings of the Evaluation and Assessment in Software Engineering*, 2020.

[19] M. A. Rahman, M. A. Haque, M. N. A. Tawhid, and M. S. Siddik, "Classifying non-functional requirements using rnn variants for quality software development," *Proceedings of the 3rd ACM SIGSOFT International Workshop on Machine Learning Techniques for Software Quality Evaluation*, 2019.

[20] M. Mohamad, J.-P. Steghöfer, A. Åström, and R. Scandariato, "Identifying security-related requirements in regulatory documents based on cross-project classification," *Proceedings of the 18th International Conference on Predictive Models and Data Analytics in Software Engineering*, 2022.

[21] S. Gärtner, T. Ruhroth, J. Bürger, K. Schneider, and J. Jürjens, "Maintaining requirements for long-living software systems by incorporating security knowledge," *2014 IEEE 22nd International Requirements Engineering Conference (RE)*, pp. 103–112, 2014.

[22] J. Slankas and L. A. Williams, "Access control policy extraction from unconstrained natural language text," *2013 International Conference on Social Computing*, pp. 435–440, 2013.

[23] S. M. Imtiaz, M. R. Amin, A. Q. Do, S. Iannucci, and T. Bhowmik, "Predicting vulnerability for requirements," *2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI)*, pp. 160–167, 2021.

[24] K. Schneider, E. Knauss, S. H. Houmb, S. Islam, and J. Jürjens, "Enhancing security requirements engineering by organizational learning," *Requirements Engineering*, vol. 17, pp. 35–56, 2012.

[25] T. Li and Z. Chen, "An ontology-based learning approach for automatically classifying security requirements," *J. Syst. Softw.*, vol. 165, p. 110566, 2020.

[26] X. Xiao, A. M. Paradkar, S. Thummalapenta, and T. Xie, "Automated extraction of security policies from natural-language software documents," in *SIGSOFT FSE*, 2012.

[27] S. M. Imtiaz and T. Bhowmik, "Towards data-driven vulnerability prediction for requirements," *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2018.

[28] W. Wang, K. R. Mahakala, A. Gupta, N. Hussein, and Y. Wang, "A linear classifier based approach for identifying security requirements in open source software development," *J. Ind. Inf. Integr.*, vol. 14, pp. 34–40, 2019.

[29] J. Cleland-Huang, R. Settimi, X. Zou, and P. Solc, "The detection and classification of non-functional requirements with application to early aspects," *14th IEEE International Requirements Engineering Conference (RE'06)*, pp. 39–48, 2006.

[30] L. Tóth and L. Vidács, "Comparative study of the performance of various classifiers in labeling non-functional requirements," *Inf. Technol. Control.*, vol. 48, pp. 432–445, 2019.

[31] E. Knauss, S. H. Houmb, K. Schneider, S. Islam, and J. Jürjens, "Supporting requirements engineers in recognising security issues," in *Requirements Engineering: Foundation for Software Quality*, 2011.

[32] A. Rashwan, O. Ormandjieva, and R. Witte, "Ontology-based classification of non-functional requirements in software specifications: A new corpus and svm-based classifier," *2013 IEEE 37th Annual Computer Software and Applications Conference*, pp. 381–386, 2013.

[33] T. Hey, J. Keim, A. Koziolek, and W. F. Tichy, "Norbert: Transfer learning for requirements classification," *2020 IEEE 28th International Requirements Engineering Conference (RE)*, pp. 169–179, 2020.

[34] Z. Kurtanović and W. Maalej, "Automatically classifying functional and non-functional requirements using supervised machine learning," *2017 IEEE 25th International Requirements Engineering Conference (RE)*, pp. 490–495, 2017.

[35] X. Luo, Y. Xue, Z. Xing, and J. Sun, "Prcbert: Prompt learning for requirement classification using bert-based pretrained language models," *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, 2022.

[36] O. AlDhafer, I. Ahmad, and S. Mahmood, "An end-to-end deep learning system for requirements classification using recurrent neural networks," *Inf. Softw. Technol.*, vol. 147, p. 106877, 2022.

[37] C. Li, L. Huang, J. Ge, B. Luo, and V. Ng, "Automatically classifying user requests in crowdsourcing requirements engineering," *J. Syst. Softw.*, vol. 138, pp. 108–123, 2018.

[38] M. Younas, D. N. A. Jawawi, I. Ghani, and M. A. Shah, "Extraction of non-functional requirement using semantic similarity distance," *Neural Computing and Applications*, vol. 32, pp. 7383–7397, 2019.

[39] I. Khurshid, S. Imtiaz, W. Boulila, Z. Khan, A. Abbasi, A. R. Javed, and Z. Jalil, "Classification of non-functional requirements from iot oriented healthcare requirement document," *Frontiers in Public Health*, vol. 10, 2022.

[40] M. Younas, K. Wakil, D. N. A. Jawawi, M. A. Shah, and A. Mustafa, "An automated approach for identification of non-functional requirements using word2vec model," *International Journal of Advanced Computer Science and Applications*, 2019.

[41] R. Chatterjee, A. Ahmed, P. R. Anish, B. K. Suman, P. Lawhatre, and S. Ghaisas, "A pipeline for automating labeling to prediction in classification of nfrs," *2021 IEEE 29th International Requirements Engineering Conference (RE)*, pp. 323–323, 2021.

[42] S. Amasaki and P. Leelaprute, "The effects of vectorization methods on non-functional requirements classification," *2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pp. 175–182, 2018.

[43] K. Kaur and P. Kaur, "Sabdm: A self-attention based bidirectional-rnn deep model for requirements classification," *Journal of Software: Evolution and Process*, 2022.

[44] G. Li, C. Zheng, M. Li, and H. Wang, "Automatic requirements classification based on graph attention network," *IEEE Access*, vol. 10, pp. 30080–30090, 2022.

[45] N. Munaiah, A. Meneely, and P. K. Murukannaiah, "A domain-independent model for identifying security requirements," *2017 IEEE 25th International Requirements Engineering Conference (RE)*, pp. 506–511, 2017.

References

[46] Wiegers, K.E., Beatty, J.: Software Requirements, 3rd edn. Microsoft Press, Redmond, WA, USA (2013)

[47] Kotonya, G., Sommerville, I.: Requirements Engineering: Processes and Techniques, 1st edn. Wiley Publishing, New York (1998)

[48] Kalloniatis, C.: Incorporating privacy in the design of cloud-based systems: a conceptual meta-model. Information & Computer Security **25**(5), 614–633 (2017)

[49] Omoronyia, I., Cavallaro, L., Salehie, M., Pasquale, L., Nuseibeh, B.: Engineering adaptive privacy: on the role of privacy awareness requirements. In: Proceedings of the 2013 International Conference on Software Engineering, pp. 632–641

(2013). IEEE Press

[50] Ayala-Rivera, V., Pasquale, L.: The grace period has ended: An approach to operationalize gdpr requirements. In: 2018 IEEE 26th International Requirements Engineering Conference (RE), pp. 136–146 (2018). IEEE

[51] Haley, C.B., Moffett, J.D., Laney, R., Nuseibeh, B.: A framework for security requirements engineering. In: Proceedings of the 2006 International Workshop on Software Engineering for Secure Systems, pp. 35–42 (2006). ACM

[52] Kalloniatis, C., Kavakli, E., Gritzalis, S.: Addressing privacy requirements in system design: the pris method. Requirements Engineering **13**(3), 241–255 (2008)

[53] Van Lamsweerde, A.: Elaborating security requirements by construction of intentional anti-models. In: Proceedings of the 26th International Conference on Software Engineering, pp. 148–157 (2004). IEEE Computer Society

[54] Meth, H., Brhel, M., Maedche, A.: The state of the art in automated requirements elicitation. Information and Software Technology **55**(10), 1695–1709 (2013)

[55] Kitchenham, B.A., Charters, S.: Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report (July 2007). https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf

[56] Netto, D., Peixoto, M.M., Silva, C.: Privacy and security in requirements engineering: Results from a systematic literature mapping. In: Workshop em Engenharia de Requisitos (2019)

[57] Binkhonain, M., Zhao, L.: A review of machine learning algorithms for identification and classification of non-functional requirements. Expert Syst. Appl. X **1**, 100001 (2019)

[58] López-Hernández, D.A., Ocharán-Hernández, J.O., Mezura-Montes, E., Sánchez-García, Á.J.: Automatic classification of software requirements using artificial neural networks: A systematic literature review. 2021 9th International Conference in Software Engineering Research and Innovation (CONISOFT), 152–160 (2021)

[59] Antón, A.I., Earp, J.B., He, Q., Stufflebeam, W.H., Bolchini, D., Jensen, C.: Financial privacy policies and the need for standardization. IEEE Security & Privacy Magazine **2**, 36–45 (2004)

[60] Antón, A.I., Earp, J.B., Vail, M.W., Jain, N., Gheen, C.M., Frink, J.M.: Hipaa's effect on web site privacy policies. IEEE Security & Privacy **5** (2007)

[61] Dalpiaz, F., van der Schalk, I., Brinkkemper, S., Aydemir, F.B., Lucassen, G.:

Detecting terminological ambiguity in user stories: Tool and experimentation. Information and Software Technology **110**, 3–16 (2019) https://doi.org/10.1016/j.infsof.2018.12.007

[62] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A.: Experimentation in Software Engineering. Springer, New York (2012)