Abstract—Machine learning (ML) has become a core feature for today’s real-world applications, making it a trending topic for the software engineering community. Requirements Engineering (RE) is no stranger to this, and its main conferences have included workshops aiming at discussing RE in the context of ML. However, current research on the intersection between RE and ML mainly focuses on using ML techniques to support RE activities rather than on exploring how RE can improve the development of ML-based systems. This paper concerns a systematic mapping study aiming at characterizing the publication landscape of RE for ML-based systems, outlining research contributions and contemporary gaps for future research. In total, we identified 35 studies that met our inclusion criteria. We found several different types of contributions, in the form of analyses, approaches, checklists and guidelines, quality models, and taxonomies. We discuss gaps by mapping these contributions against the RE topics to which they were contributing and their type of empirical evaluation. We also identified quality characteristics that are particularly relevant for the ML context (e.g., data quality, explainability, fairness, safety, and transparency). Main reported challenges are related to the lack of validated RE techniques, the fragmented and incomplete understanding of NFRs for ML, and difficulties in handling customer expectations. There is a need for future research on the topic to reveal best practices and to propose and investigate approaches that are suitable to be used in practice.

Index Terms—requirements engineering, machine learning, systematic mapping study

I. INTRODUCTION

Machine Learning (ML) is the study of computer algorithms that improve automatically through experience [42]. Its purpose is to build a model based on sample data, known as training data. This kind of systems, unlike traditional software systems, base its behavior on data from the external world instead of explicitly programming hard rules. ML components are often developed by data scientists who typically lack foundations to build reliable software systems [30]. On the other hand, the paradigm shift that makes data, to some extent, replace code, supposes a change in the way of designing, developing and testing this type of systems.

This is challenging from the point of view of Software Engineering (SE). For example, data should be tested and models be validated just as thoroughly as code, but there is, currently, a lack of best practices on how to do so [5]. In that sense, Requirements Engineering (RE) plays an important role in the development of ML-based systems since a machine-learned model is a specification that is built based on training data, that is, a learned description of how the system shall behave. In line with this argument, Kästner [26] stated that ML corresponds to the RE phase of a project rather than the implementation phase. In addition, he also argued that ML should worry about its validation, typically associated with RE, instead of verification, that is, whether the model has learned the right specification. This means that the learned behavior of a ML-based system might be incorrect, even if the learning algorithm is implemented correctly, a situation in which traditional testing techniques are ineffective.

The intersection of RE and ML has been studied in recent years by the RE community [15] and discussed in renown SE conferences such as RE, REFSQ, ESEC/FSE, ICSE, and ESEM. However, current research on this intersection focuses on using ML techniques to support RE activities rather than on exploring how RE can improve the use of these techniques in the entire software development process [53]. ML can benefit from the RE perspective even from studies suggesting that RE is the most difficult activity for the development of ML-based systems [24], [34]. For instance, RE techniques for ML could help in identifying quality metrics beyond accuracy, to allow better dealing with customer expectations, understanding why models do not fit and for whom they do not fit, which data is missing and how the data is analyzed and generalized.

In response to the importance and benefits that RE can offer to the development of ML-based systems, we contribute a first step in synthesizing existing work on RE for ML. In particular, we report on a systematic mapping study with the research objective of outlining the state of the art of RE for ML-based systems. We aim at characterizing RE contributions in terms of RE topics, quality characteristics, challenges, research directions, research type facets and empirical evaluations.

We found that the main RE contributions are in the form of approaches that address different RE activities, quality models, analysis of unique characteristics of RE for ML, taxonomy of problems, checklists and guidelines to support requirements engineers. We identified that requirements elicitation and requirements analysis are the main RE activities addressed by such contributions. We also identified unique quality characteristics (e.g., explainability, fairness, transparency and ethics) and RE challenges to ML (e.g., how to deal with stakeholder expectations, aligning data with the business goals and how to
The remainder of this paper is organized as follows. Section II provides the background and an overview on related work. Section III describes the mapping study protocol and how it was applied. Section IV presents the mapping study results. Section V discusses the results. Section VI describes the threats to the validity of our study. Finally, Section VII presents the concluding remarks.

II. BACKGROUND AND RELATED WORK

RE research is characterized by the involvement of interdisciplinary stakeholders and uncertainty [54]. Hence, RE is highly volatile and inherently complex by nature [17]. On the other hand, the use of ML-based systems has grown considerably in recent years resulting in increasing demands for high quality for such applications. This tacitly involves RE since it plays a critical role for addressing software quality characteristics. RE for ML-based systems faces unique difficulties as it constitutes a paradigm shift compared to conventional software development. As a response, a couple of software process models for ML have emerged [3], [41], [59]. They all have similar stages, starting with requirements and ending with quality assurance. However, the vision of RE is limited and superficial.

Some research has pointed out problems and challenges related to RE for ML [9], [24], [34]. Others have proposed approaches, methods and frameworks to address different concerns [23], [49], [58]. Despite the important contributions in the field so far, the synergy of RE for ML has not been so comprehensively studied [21], [53]. For instance, researchers would benefit from more studies about how RE is addressed in practice in order to propose new contributions. Practitioners, in turn, would benefit from those contributions in order to improve the development of ML-based systems.

To the best of our knowledge, we are aware of only two mapping studies that somehow relate to RE for ML. Schuh et al. [51] conducted a systematic literature review aiming at identifying design patterns, data model requirements, and technology potentials for ML systems in manufacturing companies. The authors found data characteristics such as quantity, quality and dimensionality. On the other hand, Borg et al. [10] conducted a review of verification and validation for ML in the automotive industry. The authors identified challenges such as transparency and requirements specification. However, we consider that the scope of these studies is significantly different from ours since it only addresses RE partially and limits its scope to one specific industrial sector.

Nevertheless, in order to broaden our vision, we also covered secondary studies concerning SE for ML that shed some light on RE as part of our related work. Kumeno et al. [32] and Lorenzoni et al. [38] surveyed the literature in order to outline the SE challenges that emerge during the development of ML systems. Regarding RE, the authors found that software requirements activities for ML applications involve activities such as data and feasibility analysis, requirements elicitation, requirement specification, and validation and performance evaluation of ML-models. We know that pointing out these insights is important, however, we consider the coverage of these studies insufficient from the RE perspective since there are other interesting aspects to review in the literature, such as proposed scientific contributions and their evaluation.

Nascimento et al. [46] conducted a systematic literature review in order to investigate how SE has been applied in the development of ML systems and identified challenges and practices that are applicable. Somehow related to RE, they identified which practices data scientists use to improve the quality of project data such as cross-validation, checking data distribution, and checking implicit constraints.

In summary, these studies agree that RE is a challenging task in SE for ML, but do not focus on synthesizing existing work on RE for ML. To address this gap, our systematic mapping study aims at providing an overview of research contributions regarding RE for ML.

III. SYSTEMATIC MAPPING PROTOCOL

Systematic Mapping (SM) studies are designed to provide a wide overview of a research area, to establish if research evidence exists on a topic and provide an indication of the quantity of the evidence [31].

The SM study was performed following the guidelines proposed by Kitchenham and Charters [31] and the SM-specific guidelines by Petersen et al. [48]. After identifying the need for the review (cf. Section II), we define the research questions, search strategy and inclusion/exclusion criteria.

A. Research Objectives and Questions

The main research objective is to outline the state of the art of RE for ML-based systems. The following research questions were derived from the objective in order to further characterize the RE contributions.

RQ1. What RE contributions have emerged to support the software development of ML-based systems? This question aims at providing a general overview of RE contributions (e.g., approaches, quality models, checklists) that have been proposed for ML.

RQ2. What RE topics do the contributions address? The aim of this question is to identify the specific RE topics that were the focus of the contributions (cf. Table III), helping to further understand their purpose.

RQ3. What quality characteristics do the RE contributions consider for ML-based systems? There is a consensus in the SE community that the quality of ML-based systems must go beyond metrics such as accuracy [29]. This question aims at pointing out concerns about quality in ML systems.

RQ4. What are the reported challenges and research directions on the interplay between RE and ML-based systems? This question aims at identifying open challenges. One of the main reasons to conduct a SM is supporting the planning of new research. Thus, this question seeks to indicate the aspects that may be studied by other researchers.

RQ5. What are the research type facets of the contributions? The purpose of this question is to classify the
papers according to their research type facets. We adopt the classification scheme by Wieringa et al. [57].

**RQ6. Which kind of empirical evaluations have been performed to assess the contributions?** The purpose of this question is to identify what types of empirical studies have been conducted, focusing on the research type facets of evaluation and validation research from the previous question. Obtaining this information allows to get a first idea on the scientific rigour of the evidence reported in the field.

While questions RQ1-RQ4 aim at structuring the publication landscape in a conceptual manner, the last two shall provide insights into the nature of the current reported evidence.

**B. Search Strategy**

The mapping study employed a hybrid search strategy [43] that involves conducting a search string-based database search on a specific digital library (Scopus) and then complementing the set of identified papers with iterative backward and forward snowballing (using Google Scholar) following the guidelines by Wohlin [60]. We intentionally refrained from using various specific libraries, given that the chosen hybrid strategies typically achieve an appropriate balance of precision and recall [43]. Between the different hybrid strategies (sequential, parallel, and iterative) [43], we chose the more complete iterative snowballing for maximizing the recall, even though it would imply in analyzing more papers [61]. Iterative backward and forward snowballing concerns applying backward and forward snowballing on each new included paper.

We chose Scopus because it claims to be the largest database of titles and abstracts [31], which would allow us to identify a representative and unbiased seed set [43]. It is, however, backward and forward snowballing via Google Scholar which we used as an effective way to complement the identification of the broader population of studies [43], [60].

We formulated the search string to conduct the initial database search on Scopus using the PICO (Population, Intervention, Comparison, Outcome) criteria [35] strategy. Our study focuses on ML-based systems (population) and aims at identifying RE contributions for such systems (intervention). As our study concerns a mapping study, there was no specific comparison nor the need of limiting the search space regarding outcomes. Therefore, we needed keywords for ML and RE.

The defined search string, to be applied on titles, abstracts and keywords was: “(Software OR Applications OR Systems) AND (Machine Learning) AND (Requirements Engineering)”. In the next step, during the month of February, we applied a second filter (Filter 2), filtering papers by reading the titles, abstracts and selected paper parts (when necessary) while applying the inclusion and exclusion criteria. This step left us with 15 papers representing the result of the search on Scopus. All exclusions and the final set of included papers were peer reviewed by an independent researcher. In case of divergence, a third researcher was involved and a discussion was held to reach consensus.

In the third step, during the month of February, we applied backward and forward snowballing iteratively following the snowballing guidelines in [60]. In total, four backward snowballing (BS) and two forward snowballing (FS) iterations were applied (order: BS1, BS2, FS1, BS3, BS4, FS2) until reaching our final set of papers. The four backward snowballing iterations involved analyzing 624 (259 + 200 + 131 + 34) papers (including duplicates) and allowed identifying twelve additional papers to be included. The two forward snowballing iterations involved analyzing 304 (290 + 14) papers (including duplicates) and allowed identifying seven additional papers to be included. The whole snowballing process was peer reviewed. It is noteworthy that the first forward snowballing iteration retrieved a paper accepted for publication in 2020.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC1</td>
<td>Papers that do not meet the inclusion criteria</td>
</tr>
<tr>
<td>EC2</td>
<td>Papers about the use of ML techniques for improving RE activities</td>
</tr>
<tr>
<td>EC3</td>
<td>Papers not written in English</td>
</tr>
<tr>
<td>EC4</td>
<td>Grey literature, including blogs, white papers, theses, and papers that were not peer reviewed</td>
</tr>
<tr>
<td>EC5</td>
<td>Papers that are only available in the form of abstracts/posters and presentations</td>
</tr>
</tbody>
</table>

**Fig. 1** shows all the steps performed in the paper selection process. The database search results on Scopus, filters, and backward and forward snowballing are detailed below.
but published January 1st 2021 [45]. As this paper was on the borderline of our scoped time frame, but represents a valuable contribution, we decided to also include it in our mapping. This explains the single paper from 2021.

Finally, in the fifth step, we compared our results against the results provided by the related work (cf. Section II). We found only one paper [8] that was not identified by our search strategy, because this paper didn’t cite any of the remaining studies on the topic. Hence, in total, 35 papers were included in the SM study, where 15 papers came from Scopus, 19 papers came from snowballing and one paper came from analyzing related work as shown in Fig. 1. The selected papers are shown in Table III. A spreadsheet with all details on the filtering and snowballing process, documenting each iteration, can be found in our online open science repository.

D. Data Extraction and Classification Scheme

The information extracted from each of the selected papers and the classification schemes describing the different categories are presented in Table II. The complete extracted data is also available in our online open science repository.

TABLE II
DATA EXTRACTION FORM.

<table>
<thead>
<tr>
<th>Information</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Metadata</td>
<td>Includes the paper title and information such as venue, type of venue and year of publication.</td>
</tr>
<tr>
<td>RE contribution (RQ1)</td>
<td>Description of the RE contribution for ML.</td>
</tr>
<tr>
<td>RE topics (RQ2)</td>
<td>RE topics that the contribution addresses. These topics were coded based on typical RE activities (e.g., elicitation, analysis, modeling, specification, validation, verification, and management) or other RE aspects (e.g., data quality requirements, requirements assurance) that were the focus of the contributions.</td>
</tr>
<tr>
<td>Quality characteristics (RQ3)</td>
<td>Characteristic that influences the quality of ML-based systems. Often refers to Non-Functional Requirements (NFRs) (e.g., safety, explainability, performance).</td>
</tr>
<tr>
<td>RE problems (RQ4)</td>
<td>Fact or situation that requires an action by RE researchers.</td>
</tr>
<tr>
<td>Research directions (RQ4)</td>
<td>RE topics that previous studies point out to research.</td>
</tr>
<tr>
<td>Research type facet (RQ5)</td>
<td>Classification of research type facets according to Wieringa et al. [57], including the following categories: evaluation, research, solution proposal, philosophical paper, opinion paper, or experience paper.</td>
</tr>
<tr>
<td>Empirical evaluation (RQ6)</td>
<td>Classification of the empirical strategy, according to Wohlin et al. [62], including the following categories: experiment, case study, survey.</td>
</tr>
</tbody>
</table>

IV. SYSTEMATIC MAPPING RESULTS

This section presents the results of the SM study. First, we provide an overview of the included papers. Overall, we identified 35 papers. Regarding the years of publication, the papers range from 2018 to 2021. Most of the publications (31) are conference and workshop papers and only 4 papers have been published in journals. The venues in which the topic has been addressed comprise premier international SE conferences and journals such as FSE, ICSE, RE, ESEM, REJ, and TSE. This gives an idea of the relevance and interest on this topic on behalf of the SE community in the last years.

A. RQ.1 What RE contributions have emerged to support the development of ML-based systems?

Similar to our previous mapping study in the field of RE [52], we followed open coding guidelines [50] with the aim of characterizing the papers by the type of contribution. We coded the following different main contribution types for the papers: analyses (e.g., analyzing some RE aspects for ML), approaches (e.g., methods, methodologies, processes, and conceptual frameworks), checklists and guidelines (C & G), quality models (QM), and taxonomies (T). Table III shows an overview of these contributions by contribution type.

B. RQ2. Which RE topics do the contributions address?

The majority of the selected papers concern requirements elicitation practices (14 out of 35), where authors consider problems such as defining business goals and problems of understanding. Furthermore, we found five papers about requirements analysis, more specifically addressing customer

1https://doi.org/10.5281/zenodo.4682374
expectations and requirements prioritization. We also identified contributions regarding data related requirements in five papers. Other contributions regard requirements specification, assurance (typically related to quality model based software product requirements evaluation), modeling, verification of documented requirements, and validation. Fig. 2 shows the distribution of the papers by the covered RE topics.

C. RQ3. What quality characteristics do the RE contributions consider for ML-based systems?

During the analysis of the contributions, it was possible to identify several quality requirements aka non-functional requirements (NFRs) that authors consider in their research. Table IV shows the quality characteristics that were considered in the papers with their frequencies. Note that one paper can address one or more quality characteristics.

Table IV: Frequency of Quality Characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Frequency</th>
<th>Characteristic</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security</td>
<td>6</td>
<td>Testability</td>
<td>2</td>
</tr>
<tr>
<td>Explainability</td>
<td>6</td>
<td>Accountability</td>
<td>2</td>
</tr>
<tr>
<td>Privacy</td>
<td>6</td>
<td>Ethics</td>
<td>2</td>
</tr>
<tr>
<td>Data quality</td>
<td>5</td>
<td>Accuracy</td>
<td>2</td>
</tr>
<tr>
<td>Fairness</td>
<td>5</td>
<td>Suitability</td>
<td>1</td>
</tr>
<tr>
<td>Transparency</td>
<td>5</td>
<td>Uncertainty</td>
<td>1</td>
</tr>
<tr>
<td>Reliability</td>
<td>4</td>
<td>Autonomy</td>
<td>1</td>
</tr>
<tr>
<td>Safety</td>
<td>4</td>
<td>Robustness</td>
<td>1</td>
</tr>
<tr>
<td>Performance</td>
<td>3</td>
<td>Modularity</td>
<td>1</td>
</tr>
<tr>
<td>Maintainability</td>
<td>2</td>
<td>Scalability</td>
<td>1</td>
</tr>
<tr>
<td>Legal requirements</td>
<td>2</td>
<td>Usability</td>
<td>1</td>
</tr>
</tbody>
</table>

D. RQ4. What are the reported challenges and research directions on the interplay between RE and ML-based systems?

Some papers explicitly report challenges from the point of view of RE when developing ML-based systems. We grouped them in order to provide a better understanding and then outline the challenges. A brief overview is summarized below.

**Lack of validated techniques.** Developing ML-components mainly relies on applying techniques to achieve an objective. However, there seems to be a lack of validated techniques for some important aspects of RE for ML. For instance, several studies (e.g., [P5],[P8],[P12],[P13],[P14]) state that ML researchers and users currently lack an ML-specific way to express and specify requirements for ML, including targets and trade-offs, and the influence of domain context. Other studies, such as [P1],[P3],[P5], mention that measuring quality beyond traditional metrics, such as accuracy and precision, may be complicated since identifying quality attributes is often difficult. The authors of [P13] also outline that identifying business metrics is not trivial since customers want to have policies to improve their business, but do not understand what metrics and data are required to do so. This represents a challenge for requirements engineers of ML-based systems.

In addition, in [P28],[P35] the authors raise issues on how to properly cover and validate requirements for ML systems and how to deal with testing and verification activities.

**Knowledge regarding NFRs.** In [P14] the authors state that the understanding of NFRs for ML is fragmented and incomplete, including how to define and refine NFRs in ML-specific contexts. Quality attributes such as explainability ([P22],[P23],[P33]), safety ([P3]), security ([P18]), fairness ([P3]), robustness ([P5]), and transparency ([P19]) are pointed out as challenging by researchers.

**Handling customer expectations.** Organizations did not realize that ML models are mainly probabilistic models that commonly have to learn patterns from messy data. This reflects difficulties customers have to understand potential limitations of ML systems. Papers such as [P7],[P8],[P13],[P17] reveal that customers commonly expect to see magic coming out of data.

Furthermore, we wanted to know what research directions are encouraged by the authors. After analyzing the papers, we identified that the authors are mainly asking the community to conduct more empirical studies to uncover more insights on best practices and to propose and investigate approaches that are suitable to be used in practice. The authors also mention other research directions, such as:

- Address transparency, explainability and safety for ML.
- Develop tools to support requirements specification
- Create guidelines related to ML-based system requirements (training data, specification documents, test design, runtime monitoring, and maintenance).
- How to operationalize ethics, security, and privacy.
- How to verify ML requirements and validate ML models.
- Extend the ML quality characteristics.
- Develop standard quality models for ML-based systems.
- Survey ML literature and/or ML experts on NFRs.
- Understand how ML systems integrate with typical software from a quality perspective.
- Provide guidance to non-technical stakeholders about what is possible and what is not.

E. RQ5. What are the research type facets of the approaches?

Fig. 3 shows the distribution of the research type facets of the papers per year. It is possible to observe that most of the papers (16 out of 35) concern evaluation research papers. In the next question we address the types of empirical evaluations that were conducted. Opinion papers, with ten studies, significantly contribute to the account. We also identified that solution proposals are still scarce in this field. This contrasts the identified lack of techniques and the absence of tools supporting RE activities for ML-based systems, further motivating research directions pointed out by the authors.

![Fig. 2. Distribution of papers per RE topics.](image-url)
When analyzing the empirical evaluations conducted within the studies (Fig. 4), it was possible to identify 16 papers out of 35 that have performed empirical evaluations (twelve case studies, three surveys and one experiment). Note that five papers provided a proof of concept, i.e., a realization of a certain method or idea in order to demonstrate its feasibility. This is not considered as an empirical evaluation by Wohlin et al. [62], therefore they were not classified as evaluation research. In fact, 14 studies did not contain any type of empirical evaluation or even a proof of concept. Most of these concern opinion and experience papers. The most applied empirical evaluation strategy in the analyzed studies was case study (12 papers) in contrast with survey (three papers) and experiment (one paper).

V. DISCUSSION

Several different research contributions have been proposed recently with regard to RE for ML (cf. Table III). These contributions comprise analyses, approaches, checklists, guidelines, quality models, and taxonomies. Fig. 5 presents a bubble plot mapping the identified contribution types against the covered RE topics and the type of empirical evaluation that has been conducted. It is evident that there are still relevant gaps to cover.

The main identified challenges concern the lack of validated techniques to address the identified gaps, a lack of knowledge regarding specific NFRs that are particularly relevant in this context (e.g., explainability, fairness, safety, transparency), and difficulties in handling customer expectations. We listed several examples of research directions to handle these challenges and other opportunities reported by the authors.

With respect to the kind of research and empirical evidence, we observed that there is still a limited amount of solution proposals and that 14 out of 35 (40%) papers did not contain any type of evaluation (not even a proof of concept). This could be related to the fact that it is a recent topic, which is still to acquire maturity and to move towards more rigorously assessed empirical evidence. Nevertheless, we have relatively little evidence about the feasibility of the presented approaches, which represents a problem for practitioners and an opportunity for researchers.

An intriguing fact is that almost all papers, except [P25],[P26], have the conviction that their proposed contributions are rationally applicable for all kind of ML approaches (e.g., supervised and unsupervised learning, reinforcement learning) and models (e.g., artificial neural networks, decision trees, support vector machines). This gives an idea that within the ML field there are no particularities that may affect the generalization of the studies. Based on our experiences developing ML-based systems [27],[28], we believe that this might not hold and should be further investigated.

Another interesting fact is that only two papers [P13],[P21] investigate what processes are used in practice. In [P13] the authors identified that problem understanding in the observed RE for ML context involved studying available data and meeting with customers. For requirements elicitation and specification they found that practitioners typically conduct brainstorming sessions and define business metrics. In [P21] the authors identified that practitioners typically use ad-hoc methods to address requirements elicitation and NFRs assurance.

The importance of RE to design and successfully deliver a ML-based system is clear. However, this work shows that this area faces several problems that need to be addressed both in industry and research. For instance, today it is not clear what tools and methods are used in practice to address requirements for ML-based systems. It is not clear if traditional requirements engineering tools and techniques work for ML, and even worse, it is not known how requirements should be handled in the ML context. We strongly believe that these problems represent interesting research opportunities.

VI. THREATS TO VALIDITY

Internal validity: We used a hybrid search strategy combining a database search on a single database (Scopus) with iterative backward and forward snowballing (using Google Scholar), and precisely documented each step. One could argue that the search string was confined to a small set of
keywords. These keywords were objectively selected using the PICO strategy and are directly related to our research goal. It is important to remember that the database search was used to reveal an unbiased and representative seed set, as starting point for iterative forward and backward snowballing, and that this search strategy has been effective for secondary studies [43].

**External validity:** We systematically applied a search strategy that has shown good results regarding recall [43] and validated it by comparing it against related work. Still, there is a possibility of having missed studies. Nevertheless, we were unable to manually find any additional study to be included and are confident that we have an unbiased and representative sample. The claims made in our paper are related to the findings reported in the primary studies. While all of them were peer reviewed, and many of them were published in venues that have a rigorous selection process, we did not assess their quality. Such quality assessment is typically not part of mapping studies, and could be part of a systematic review extension. The complete information concerning the process, the extracted data and coding is available in our online repository and is publicly auditable.

**Reliability:** In order to reduce the bias when selecting relevant studies, it was decided to examine the selected papers in pairs. Hence, two researchers evaluated the selected studies, extracted data and coding in a peer-reviewed manner.

**VII. CONCLUDING REMARKS**

This paper presents the results of a SM study on RE in the context of ML-based systems. We applied a hybrid search strategy, complementing a database search on Scopus with iterative backward and forward snowballing. Our search strategy allowed identifying a total of 35 studies.

We identified several proposed research contributions, some published in premier software engineering conferences and journals. These contributions comprise analyses, approaches, checklists and guidelines, quality models, and taxonomies. We identified research gaps by relating these contributions to RE investigation topics. We also highlighted quality characteristics considered within the papers and reported on challenges and potentially promising research directions.

Hence, the main contributions of this research are twofold: (i) mapping relevant knowledge about the current state of RE for ML, a subject that is not yet widely explored by researchers and confused by practitioners; and (ii) helping to identify points that still require further investigation. As far as we know, this paper is the first systematic mapping study that organizes evidence to provide a comprehensive overview on contributions related to RE for developing ML-based systems.

**REFERENCES**


