29th IEEE International Requirements Engineering Conference (RE2021), 20-24 Sept 2021, South Bend, USA

What's up with Requirements Engineering for Artificial Intelligence Systems?

Khlood Ahmad, Muneera Bano, Mohamed Abdelrazek, Chetan Arora
School of Information Technology
Deakin University, Geelong, Australia
Email: ahmadkhl@deakin.edu.au, muneera.bano@deakin.edu.au,
mohamed.abdelrazek@deakin.edu.au, chetan.arora@deakin.edu.au

John Grundy
Faculty of IT
Monash University
Email: john.grundy@monash.edu

***Abstract*—In traditional approaches to building software systems (that do not include an Artificial Intelligent (AI) or Machine Learning (ML) component), Requirements Engineering (RE) activities are well-established and researched. However, building software systems with one or more AI components may depend heavily on data with limited or no insight into the system's workings. Therefore, engineering such systems poses significant new challenges to RE. Our search showed that literature has focused on using AI to manage RE activities, with limited research on RE for AI (RE4AI). Our study's main objective was to investigate current approaches in writing requirements for AI/ML systems, identify available tools and techniques used to model requirements, and find existing challenges and limitations. We performed a Systematic Literature Review (SLR) of current RE4AI methods and identified 27 primary studies. Using these studies, we analysed the key tools and techniques used to specify and model requirements and found several challenges and limitations of existing RE4AI practices. We further provide recommendations for future research, based on our analysis of the primary studies and mapping to industry guidelines in Google PAIR). The SLR findings highlighted that present RE applications were not adaptive to manage most AI/ML systems and emphasised the need to provide new techniques and tools to support RE4AI.**

***Index Terms*—Requirements Engineering, Artificial Intelligence, Machine Learning, Systematic Literature Review**

I. Introduction

The increase in data we generate on the internet and the advancement in processing power have made AI a practical solution to many of the automation challenges we face today. Thus, making AI a more favoured software alternative to many organizations [1]. However, the process of building AI systems differ from traditional approaches [2]. In ML systems, outcomes remain unclear until the model is trained and tested with a specific data set, making it challenging to use the structural approach adopted in traditional methods [3]. Building such systems has created new challenges to the RE community and made it challenging to use current RE methodologies [4], with more studies emphasizing the need for RE to change and adapt to AI/ML systems [5].

Using existing RE practices when dealing with an unexplainable or unpredictable system has introduced new issues and challenges. These challenges include defining requirements and the emergence of new requirements, such as data and ethics. For example, how do we define ethics? [S19]. New tasks and responsibilities for the requirements engineer have also emerged. In current ML systems, data scientists are found responsible for writing high-level requirements. The results are systems that focus on data selection and quality assurance rather than understanding the business domain and stakeholders' needs [S7]. As a result, it has become essential that both data scientist and software engineers should step up their knowledge and understanding of the issues arising from blending AI into most software projects and learn to work together [6].

Kondermann [7] argues that although RE is well researched, it is still not applied to AI, specifically computer vision, with the need to study it further to include data selection techniques. There is a complexity in understanding how to manage and produce requirements for AI/ML systems. To further understand the methods and issues presented in current RE4AI research, we conducted a systematic literature review to identify existing empirical studies and theories. We also reviewed currently used modeling languages, requirements notations and emerging limitations and challenges.

To our knowledge, this is the first SLR conducted on RE for AI systems. There have been suggestions on the need to perform an SLR on non-functional requirements for AI systems [S19]. During our search, we found that most of the research to date has focused on using AI to manage RE, with little research supporting *RE4AI*.

The main research contributions for this SLR include:

- We identified a list of 27 primary studies that focus on RE for AI systems. Out of these 27 studies, 18 are empirical evaluations of using existing RE techniques when building a system with an AI component. The remaining 9 studies are "ideas only" papers that provide a proposal or model and need to be evaluated in future work.
- We identified five modeling languages and requirements notations currently used when writing requirements for AI systems.
- We identified the most popular domains that research RE techniques as well as domains that lack empirical research on RE4AI.
- We extracted existing limitations and challenges from the literature on RE4AI.

- We mapped the results of our SLR against industry guidelines for developing human-centered AI systems by Google (Google PAIR) [8].

Further in this paper, we provide a brief background on RE4AI in Section II. In Section III, we search for available research on RE for AI systems via a SLR process. Then, we extract and analyse our results from the selected primary studies in Section IV. Finally, we present a discussion of key results and summarise emerging theories in Section V, before concluding in Section VI.

II. Background

RE is considered to be the most crucial phase in the software engineering cycle and plays a significant role in every stage of the software lifecycle [9]. In RE, understanding stakeholders requests are important, and requirements act as the communication channel between the system developers and the stakeholders [10]. RE acts as that channel to gather and document stakeholders needs [11]. Therefore, it is essential to establish requirements early on when building software systems to ensure all stakeholders needs and specifications are captured and documented correctly.

Koelsch defined requirements as "a need, desire, or want to be satisfied by a product or service" [12]. If this need or desire is not satisfied, then the product is not usable. An established process for RE is developed to achieve this need and undergoes the following phases: elicitation, analysis, specification and documentation, validation and management [13]. This process makes sure that requirements are extracted, documented and managed correctly, and comply with users' needs. However, the process of software development differs when AI/ML components are involved. Such systems are not always driven by specifications and parts of the system can be driven from data [S2]. Therefore, RE techniques would need to be adjusted to the changes introduced by this new paradigm of AI systems.

In RE requirements are classified to be either functional or non-functional. Functional requirements represent the systems features and business rules to what the system should include. Whereas non-functional requirements include systems qualities and constraints [12]. Non-functional requirements are more difficult to present [14]. However, there are several modeling languages and tools available that can help display properties of non-functional requirements.

Modeling languages display WHY behaviours and functionalities are selected and WHAT capabilities are needed to support these choices. Modeling languages focus on the high-level abstraction aspect of the required system rather than the details of operations, which is helpful during the early stages of building software systems [15], [16]. There are different RE modeling languages available such as Goal-Oriented Requirements Engineering (GORE) and User Requirements Notation (URN) [15].

When creating systems that are integrated with ML components, the ML code is relatively small compared to the actual process. Most of the work focuses on managing data, feature extraction, analyzing, configuring, etc. [2]. The configuration process, selecting and validating data, and other activities need to be specified in the requirements phase. However, RE is not as established in AI systems as the traditional approach for non-AI software. It is shown that AI systems usually lack proper RE techniques [17]. Having a different ML systems process creates a need to adapt to existing methods and techniques. And the black-box nature of AI has made it challenging to use existing RE methods [4]. Unlike the traditional approach to RE, where systems are designed based on precise requirements, AI / ML systems do not have a distinct set of requirements [18].

Kuwajima et al. [19] concluded that the lack of requirements specifications in current ML systems significantly impacts the ML model's quality. And that most ML models lacked requirements specifications and current practices are not well defined or organized. One of the reasons for the difficulties in writing requirements specifications for ML systems is the inconsistencies in inputs and outputs patterns. There are several available tools used for traditional SE practices to manage code and other issues. However, because of the vast difference in ML systems and traditional SE practices, it is hard to use these tools in managing issues resulting from ML systems [2].

Including AI/ML components in building software systems has impacted RE, and new requirements have appeared in the process. Non-functional requirements (NFR) for ML systems have changed to include transparency, trust, privacy, safety, reliability and security [20]. NFR, such as responsible and trustworthy AI, has recently gained significant attention from both researchers and the industry.

In 2018, the European Commission developed a set of ethical guidelines for trustworthy AI [21]. The guidelines emphasised that to create trustworthy AI, the outcome should be lawful, ethical and robust. Lawful meant that developers should comply with all legal regulations. For example, it should abide by the rules and regulations of the European General Data Protection Regulation (GDPR). Ethical, concluded that the system should have respect for humans, prevent harm, be fair and explicable. Finally, robust meant that delivered systems should be safe, secure and reliable. With the introduction of these guidelines, more research is growing towards creating ethical AI software. However, the concept of ethics is not clearly defined and is difficult to apply [22]. There is still a lot of work that needs to be invested in this area.

Bosch et al. [S10] reviewed current software development methods that could be adapted to AI systems. The authors named three approaches to build software and provided a framework as a guide. Current practices involved requirement-driven, data-driven and AI-driven. The authors explained that businesses are moving towards data-driven systems, and decisions are becoming more dependent on data to determine the system's functionalities. Resulting in a demand to modify current RE practices to become more adaptive to data-driven approaches [S10]. Building AI systems is still ongoing, as there is little research on this topic. From a RE perspective, we are facing a new set of challenges [23], such as specifying and defining requirements [5], [19]. So how do we incorporate RE into a system that is not explainable and vastly different from traditional software?

III. Systematic Literature Review

To perform this SLR we followed the three phases outlined by Kitchenham et al. [24] to address the following research questions (RQs):

- **RQ1.** What are the requirements notations and modeling languages used in building current AI systems?
- **RQ2.** What is the application domain for each study?
- **RQ3.** What are the limitations and challenges reported in existing RE4AI research?

The three steps included planning, conducting and reporting the review. The initial planning phase involved writing a protocol and identifying a set of research questions (RQs). The protocol included a plan for a search strategy. To exhaust our exploration of any existing empirical evidence, we identified relevant keywords and search strings to use in our search. The main keywords included "Requirements Engineering" and "Artificial Intelligence". For both of these main search terms, we derived a number of alternative terms (e.g., machine learning) based on existing literature, and connected the search string using boolean operators. And only included papers that have been published after 2010.

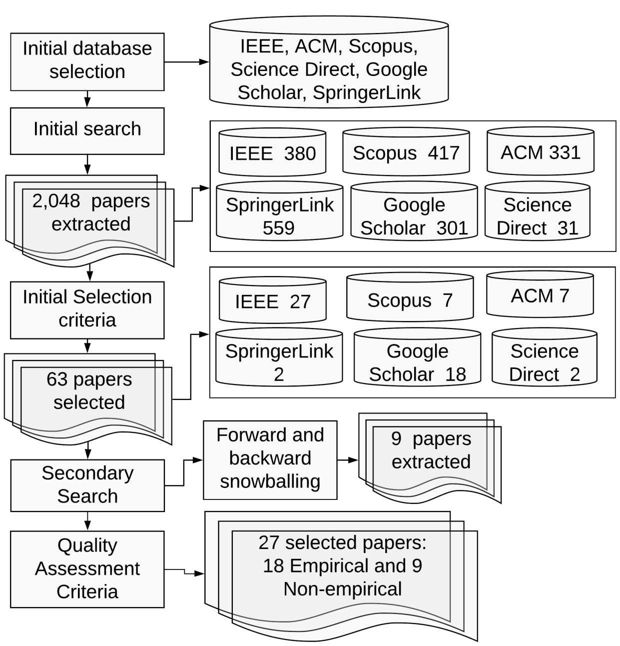

Fig. 1. SLR process of paper extraction

The second stage required identifying existing primary studies that would answer our RQs. We entered a combination of search strings into six different online databases in August 2020 resulting in a total of 2,048 papers. We used the following online databases to extract and identify any research papers related to our study: IEEE Explore, ACM Digital Library, Google Scholar, Science Direct, SpringerLink and Scopus. After conducting the initial search, we performed a selection criteria measure to extract relevant studies. Our inclusion criteria involved selecting primary studies that focused on requirements engineering for AI systems and did not use artificial intelligence to manage requirements engineering and published in English. Our exclusion criteria excluded secondary studies and papers that were not peer-reviewed. Next, we conducted a secondary search on the resulting 63 papers to find any resources we might have missed, which involved collecting relevant research from references, citations and authors profiles. We implemented backward snowballing by examining the references of the selected papers and forward snowballing by checking all the papers that cited our selected studies [25]. References and citations from the resulting studies from the secondary search were also scanned. Next we looked at authors who have published in RE. Our searches involved scanning papers published on authors' google scholar profiles. And finally, we searched for all papers published in RE-related conferences and workshops, those included the RE, conference, AIRE and REFSQ workshops. The same include/exclusion criteria were applied to all the papers extracted from the secondary search.

After completing the initial selection criteria, we found a total of 72 studies, 9 of which we identified during the secondary search. The remaining primary studies underwent a quality assessment check based on Kitchenham's guidelines [24]. The assessment criteria were then used to assess each paper's research methodology. It included a general checklist that assessed all the papers, and a separate checklist for each methodology used. The different types of methodologies we assessed were case studies, surveys and experiments. We used the checklist to identify any biases or validity issues that might have been evident in all primary studies selected from the initial search. Initially, the intention was to include only existing empirical work on the topic; however, due to the limited research available, a decision was made to add non-empirical studies. Non-empirical studies were checked for quality based on the general checklist section. After finalising the quality assessment check, we selected 18 empirical studies and 9 non-empirical studies. Included empirical papers were [S1]–[S18]. And non-empirical included [S19]–[S27].

For data extraction and analysis, first, we entered the selected 27 papers into NVivo[1] for code extraction. We adopted a thematic analysis to extract data with a top-down coding strategy and generated codes and themes based on our research questions. Next, we used an open-coding procedure on each transcript to assign relevant text to its matching code [26]. And finally, we analysed the results from NVivo to answer each research questions and present any emerging theories.

[1]https://www.qsrinternational.com/nvivo

TABLE I
THE NUMBER OF SELECTED PAPERS PUBLISHED IN RE VS NON-RE VENUES BEFORE AND AFTER THE QUALITY ASSESSMENT

Conference / Journal	**Ranking**	**Papers after initial selection**	**Papers after quality assessment**
AIRE Workshop	N/A	3	2
RE Conference	A	7	4
RE Conference Workshop	N/A	5	2
RE Journal	B	3	1
Workshop on REFSQ	B	5	0
Non-RE Venues		46	18
Total		69	27

IV. RESULTS

During the initial stages of our search, the majority of search results focused on using AI to manage RE and limited work was found on RE4AI. For example, the first ten results returned from the initial search strings entered in IEEE Xplore, we found eight of these papers researched ways to manage RE using AI and only two focused on RE4AI. We found the same pattern with most database results using our search string.

We also observed that RE4AI research has gained traction in the last couple of years, as shown in Figure 2. We found that 85% of the published primary studies were during the past four years, with 59% of these results published in the last two years. The increase in publications indicates that more researchers are looking into ways to address RE-related issues and challenges when building AI software.

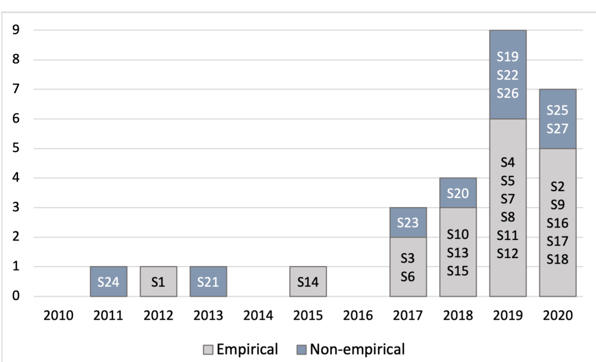

Fig. 2. Increase shown in the number of publications per year

We scanned through the proceedings of top RE related conferences and workshops as well as volumes and issues of RE related journals, and noted the number of selected papers in each conference and compared the results to non-RE venues as shown in Table I. Almost one-third of the published papers were from RE venues and two thirds from non-RE venues. The proportion of selected papers remained consistent after applying the quality assessment criteria as shown in Figure 3.

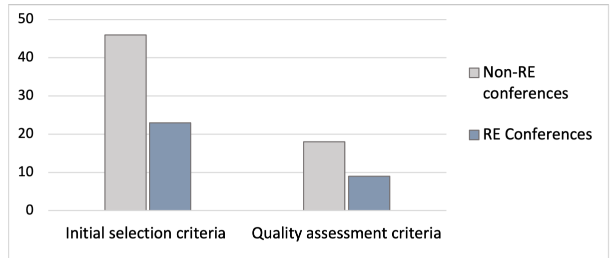

Fig. 3. The number of selected papers that have been published in RE conferences and journals vs non-RE publications

TABLE II
MODELLING LANGUAGES AND REQUIREMENTS NOTATIONS USED IN SELECTED STUDIES

Modelling language	**Study**
Goal-Oriented RE (GORE)	[S1], [S12], [S15], [S16], [S25]
UML / SysML / Use Cases	[S1], [S3], [S6], [S23], [S27]
Signal Temporal Logic (STL)	[S8]
Traffic Sequence Charts (TSC)	[S9]
Conceptual Model (CM)	[S13]

A. *RQ1 What are the Modeling Languages and Requirements Notations used in RE4AI?*

In total, 12 of the studies used modeling languages or requirements notations to present requirements. The most popular modeling notations and languages among the studies were UML and GORE, as shown in Figure 4. The study in [S1] combined two modeling techniques to produce their model, while [S3] created an extension to an existing one.

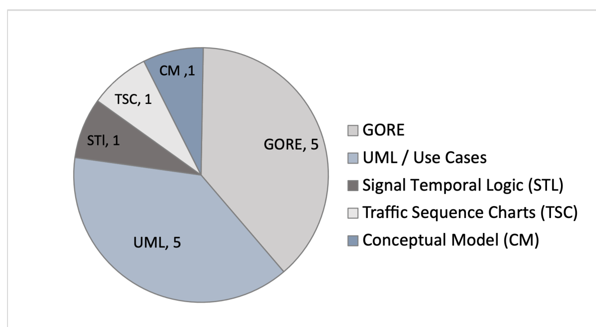

Fig. 4. The distribution of Modeling languages and requirements notations used in selected studies

GORE: Five studies preferred the use of goal-modeling techniques. The authors in [S1] argued that using goals to present requirements for surgical tasks was more favourable as it seemed to be in line with how surgeons think or perceived tasks from a medical perspective. Dimitrakopoulos et al. stated that goal-oriented methods were more suited for capturing business requirements [S12]. Neace et al. [S15] chose to model with GRL (Goal-oriented Requirements Language) because it provided better support for modeling non-functional and quality requirements. They also stated that GORE has become more popular in modeling requirements for autonomous systems. Finally, Lockerbie et al. [S16] favoured the use of soft goals in i* models as it provided a better description of a person's objective, such as quality of life.

Ishikawa et al. [S25] proposed to use GORE-MLOps, a methodology that adapts requirements analysis from GORE methods to ML systems. The study purposed to model uncertainty in RE for AI systems and its impacts on current RE techniques. Table III shows the different studies that used GORE as a modeling language.

TABLE III
THE DIFFERENT GORE METHODS PRESENTED IN SELECTED STUDIES

Gore type	Description	Study
FLAGS (Fuzzy Live Adaptive Goals for Self-adaptive systems) [27]	Used to presents requirements for tasks performed by a smart surgical robot	[S1]
CORE (Capability Oriented Requirements Engineering)	Uses goals to capture the systems current and desired capabilities	[S12]
GRL (Goal-oriented Requirements Language)	Modeled requirements for an autonomous aircraft system to detect radiation levels in disasters	[S15]
i*	Created a model for people living with dementia using i* soft goals	[S16]
GORE-MLOps	Proposes a methodology to models uncertainty in requirements for AI systems	[S25]

UML: Five papers used UML to model requirements, as shown in Table IV. The study in [S3] used SysML to model functional and non-functional requirements to graphically present requirements and the relationships between them. Using SysML allowed them to visually view behaviours between requirements. However, the drawback was that it did not provide enough aid to model non-functional requirements. As a result, the study proposed and tested an extension to SysML that included support to non-functional requirements. Amaral et al. [S27] proposed an ontology of trust to help define requirements for trustworthy AI. The study then implements these trustworthiness requirements in the OntoUML model. The model also assisted in displaying any risks related issues when it came to trust.

The work in [S6] and [S23] adopted use cases to model requirements. In [S6] they extracted requirements from interviews and mapped them into six use case traffic scenarios. They then mapped each scenario into an orientation and navigation system based on computer vision for people who are vision-impaired. On the other hand, Altarturi et al. [S23] proposed a new RE model to accommodate data requirements. The study involved generating actionable use cases for a recommendation system. The new model included the ability to collaborate between the software engineer and data scientist when writing requirements.

Signal Temporal Logic: Signal Temporal Logic (STL) is a specification language that enables real-time reasoning of properties by providing past and future variables [28]. The study in [S8] used STL to specify requirements for

TABLE IV
THE DIFFERENT UML METHODS PRESENTED IN SELECTED STUDIES

UML type	Description	Study
Statechart and sequence diagrams	Uses statecharts and sequence diagrams to model the medical robots procedure, and the interaction between the system and the user	[S1]
SysML	An extension of SysML is used to model functional and non-functional requirements for automotive car systems	[S3]
Use Cases	Use cases are created for six traffic scenarios, and proposing solutions for navigation issues for people with vision impairment's	[S6]
Actionable Use Case	Proposes an actionable use case diagram as a means of collaboration between the data scientist and the software engineer	[S23]
OntoUML	Proposes an ontology using UML to model trustworthy requirement	[S27]

a perception system in an autonomous vehicle. It provided features such as reachability, safety and reactive requirements to include in the specifications. They then mapped these requirements into three testing scenarios using the Sim-ATAV framework and used a virtual environment to generate test cases for autonomous vehicles. Based on the requirements specified using STL, the virtual models gathered data from three sensors: camera images, radar and lidar sensors. The data was then processed using a pre-trained Deep Learning model to identify any critical behaviour that might emerge.

Traffic Sequence Charts: Traffic Sequence Charts (TSC) is a graphical specification language used for traffic scenarios. TSC is based on snapshots, and each snapshot represents a traffic situation. When assembled, snapshot charts consist of history, future and consequence. Snapshots are also linked or combined with operations such as sequences and choice. The work in [S9] used TSC to display requirements for an autonomous vehicle. The main objective of the study was to find any inconsistencies in TSC requirements specifications.

Conceptual Model: The study in [S13] builds a conceptual model for a Smart Process Control System based on requirements gathered from the industry and survey results. Kohl et al, proposes a Softgoal Independency Graph (SIG) to model explainability along with other NFR to minimize conflicts. The author presented a conceptual analysis for explainability as a non-functional requirement and provided a detailed definition of explainability [S22].

B. RQ2 What are the Application Domains to Date of RE4AI?

The application domain varied between the studies as shown in Figure 5. Four papers, listed under general, did not specify a domain when applying their concepts. Horkoff [S19], focused on issues and challenges when applying non-functional requirements to AI. Nakamichi et al. [S17] tried to find ways to improve RE techniques for AI systems. They conducted a study to evaluate quality requirements and deliver customers needs. Bosch et al. [S10] identified the different approaches to building AI systems in general and proposed a framework to combine them. And finally, Ishikawa et al. [S25] presented a model to show uncertainty in AI systems and its impacts on current RE techniques.

Our results showed that the domain with the highest interest in RE4AI was autonomous driving and computer vision. We also observed that all studies in the field of autonomous driving were all based on an empirical evaluation. On the other hand, all studies on ethics, trust and explainability were theoretical papers and proposed methodologies that are not yet evaluated. Our search also reported a large number of secondary studies on explainability and ethical requirements. We conclude, based on the results, that research presented for the autonomous industry is more established. Whereas work on ethics is more theoretical and has just recently gained attention in the research industry.

Our results also showed an increase in studies investigating data requirements. We found that all empirical evaluations that involved investigating data requirements were conducted during the past two years (2019 and 2020), showing that this is an emerging topic that needs further exploration. Several studies have also emphasised the importance of managing data requirements in building AI systems as "data replaces code" [S7].

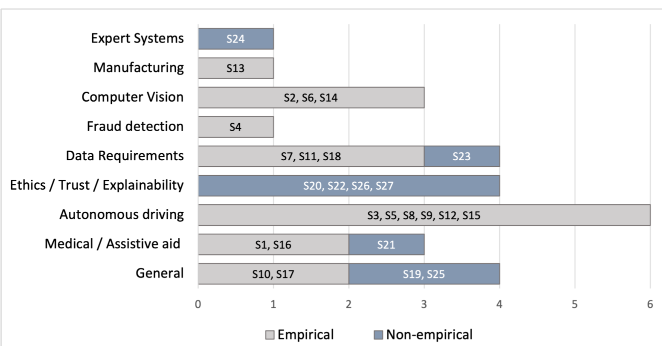

Fig. 5. Number of studies found in each application domain for RE in AI systems

C. *RQ3 What are the Limitations and Outstanding Challenges Reported for RE4AI?*

Lately, more attention has been given to the existing challenges between RE and AI software. Several secondary studies have focused on identifying existing issues in RE for AI systems. Belani et al. [4] created a taxonomy that outlined the list of challenges for each stage of the RE in ML systems, including data, model and system. In contrast, other secondary studies only focused on one aspect of RE, for example Chazette [29] only focused on the challenges associated with transparency requirements. In this SLR we identified all challenges and issues presented in the selected primary studies that have emerged due to the shift in RE4AI. Figure 6 displays the recurrences of each issue in the literature. Issues that appeared more often were linked to data requirements, followed by deciding on the trade-off and dealing with the emergence of new requirements.

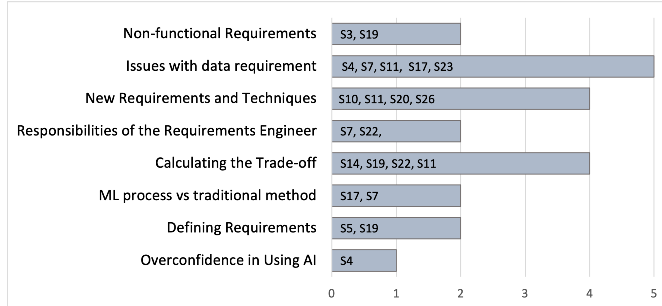

Fig. 6. Number of RE4AI issues appearing in selected primary studies

The Overconfidence in Using AI: There seems to be a common misconception of "AI will solve everything" in most organizations [30]. Sandkuhl [S4] explained that the general public usually overestimates the capabilities of AI solutions. During a series of workshops and meetings with several industrial partners, they found that many organizations would choose to use AI without having the experience and expertise in building systems with AI components. Sandkuhl emphasized that capturing requirements early on in the project is essential so stakeholders clearly understand the capabilities and limitations of AI. And, while AI is not always a feasible solution, companies should establish the need to use AI before proceeding further in the projects.

Defining Requirements: Defining requirements for AI systems can be somewhat challenging. Issues with requirements definition, especially with automated cars, can affect model training and evaluation [5]. Some requirements might be vague or hard to define. For example, how do we define a "Pedestrian" to a self-driving car? Every person might have a different definition for a pedestrian. Rahimi et al. [S5] focused on finding the requirements for "pedestrians" and how a self-driving vehicle would recognize pedestrians? The process involved searching for any feature that represented a pedestrian. Horkoff [S19] explained that our understanding of non-functional requirements is not complete and we need to set out standards to how we define them; for example, how do we define fairness?

Nature of Machine Learning Systems and the Traditional Approach of RE: Traditional non-AI systems have a process for RE techniques that are well researched and established. However, this is not the case in AI/ML, as they are usually built from available data rather than a detailed set of requirements or stakeholders needs [S2], [S5], [S10]. In [S17] the study emphasised the need to implement new quality techniques for ML systems. Vogelsang and Borg [S7] explained that the existing methods used in RE need to change to accommodate the different activities currently used for AI systems. The author defined new types of methods for RE to be included when creating requirements for ML systems. For example, the elicitation phase should identify requirements, such as data and explainability. Hence, there is a need to develop new tools or re-evaluate existing ones to support RE4AI.

Calculating the Trade-off: The issue with deciding on how to calculate the trade-off came up in 4 different studies. Horkoff [S19] explained that one of the challenges with ML systems' non-functional requirements was how to calculate trade-off when choosing an ML algorithm. For instance, do we trade privacy for transparency or fairness for accuracy? And how would we specify or express these choices? How do we decide on what requirements could be traded and at what cost? In [S14], the authors traded a slight cutback to efficiency for a substantial increase in modifiability to the system's design. This cutback provided an easier method to maintain data validity. Although there was a slight reduction in efficiency, it did not affect the system performance and was still within the processing power range. Therefore accuracy and reliability remained unchanged.

Explainability requirements can also conflict with others such as security, cost and precision. Having an AI system that is more explainable might be more expensive to build. In such a case, when would it be worth the expense to have a more explainable system? And how do you calculate the trade-off? Kohl et al. [S22] proposed to use the Softgoal Independency Graph (SIG) to model explainability along with the other NFR to minimize conflicts.

Shin et al. [S11] found that it was important to measure the trade-off between performance/cost when it came to the data's sampling rate. Algorithms such as classification and regression performed poorly when the sampling rate was low. So the higher the sampling rate, the better the quality of data was. The number of houses used was also important when it came to better performance, and including more samples in the dataset provided a more comprehensive range of diversity. However, to what extend can we invest in cost? There should be a limit to how far we can choose between performance vs cost in such cases.

Responsibilities of the Requirements Engineer: Vogelsang and Borg [S7] stated that data scientists are responsible for writing requirements in current ML systems. As ML is integrating into software systems, a new role for data scientists is emerging in the process, forcing software teams to adapt to these changes. These new roles have resulted in a gap between the Software Engineering practices, AI/ML communities and data scientists [6], [18].

On the other hand, Challa et al. [S18] reports that the RE community are not equipped to handle the vast amounts of data needed in building AI systems. In [S23], the authors emphasise the importance of including the data scientist in the process of defining and eliciting requirements, especially requirements related to data extraction. As AI systems are data centred, it is crucial to have a data scientist elicit and identify relevant data for the project. Therefore, there should be some type of communication between the requirements engineer and the data scientist, especially during the early phases of RE.

The Emergence of new Requirements and Techniques: With the emergence of new requirements for AI systems such as data, ethics, trust and transparency, new challenges are born for RE. Bosch et al. [S10] emphasized the need to adapt and complement old practices and techniques with new ones rather than replace the old practices entirely. Some authors noted that studies on RE4AI are not applied to practice, and such research findings are not being used or addressed by other researchers. For example, Shin et al. [S11] identified some data requirements that should be used in AI systems for energy consumption. However, they found that similar projects were not practising the use of such requirements.

The same goes with ethical requirements, Aydemir and Dalpiaz [S20] argued that ethics is usually overlooked and with the change in today's software systems and the introduction of AI, ethical requirements need to adapt to these changes. The authors argue that ethics is widely discussed for AI systems but neglected during the process of building such systems. Kuwajima et al. [S26] noted that most software standards such as ISO/IEC 25000 series did not apply to ML systems and had no support for ethical requirements.

Issues with data requirement: One of the significant issues with data requirement is the expense that comes with data-generation [S11]. Then there is the availability, quality [S13], training and testing of data [S17]. Requirements need to make sure the quality of data is appropriate, whether the data is available, how to test it, and which data to select for training. Altarturi et al. [S23] argue that RE methods focus on requirements that are user-centric and do not give enough attention to data requirements.

The emergence of data requirements has posed new issues for RE. Sandkuhl [S4] found that data needed for an AI project was easily accessible for a number of companies. However, the available data lacked structure and rules that were necessary to implement and train the AI system. The study then listed numerous AI requirements to consider. These requirements included data quality, structure and format. Since data is the primary driver in AI systems, it is essential to set rules and carefully select requirements for data selection and management.

Non-functional Requirements: In traditional approaches to RE, non-functional requirements (NFR) are usually well researched and established. However, our understating of NFR has changed when ML systems are involved, and we require new methods and solutions to evolve and fit NFR into RE4AI. For example, some NFR, such as compatibility and modularity, are not as important in ML systems. In contrast, other overlooked requirements such as fairness and transparency hold more value [S19]. There also appears to be less research on modelling non-functional requirements, and research tends to focus mainly on functional requirements [S3].

D. Threats to Validity

In all phases of our SLR, we considered and attempted to mitigate potential threats to validity, common in SLRs for software engineering [31]. We report on more serious threats of selection and researchers' biases and mitigation strategy in our SLR. To address any selection bias while conducting our search, we included a comprehensive set of keywords and chose 6 databases for our search to broaden our results. From the initial selection criteria, we manually read all the titles and abstracts to filter them further. The first author did the first round of selection and then verified the results in consultation with the remaining authors, to reach the consensus for the final result. Once the final list was established, we performed a more detailed scan of the entire document for the resulting papers.

The second filtration process involved a more comprehensive quality assessment test to only include primary studies that passed a specific grade or was in scope. These studies were selected based on the criteria we set, and for some papers, we were not sure if they fit our criteria. In such situations, we discussed the paper's selection in several meetings among the authors to ensure the selected article's focus was on RE4AI and to reduce researchers' bias. The similar process was followed for data extraction and analysis to reach the consensus on the results that were used to answer the research questions. Our results might have been subject to publication biases. We checked each primary study as part of the quality assessment criteria if there are any reports on any issues and what reliability measures they have performed. We did observe that out of the 18 empirical studies, 8 did not report negative results or address validity issues, however, all the selected studies passed the quality criteria.

V. Discussion

New challenges have emerged in the world of RE with the increase in AI systems. The availability of data and processing power has produced more AI centred software. Therefore, creating new issues and limitations to using existing RE techniques and methodologies. How do we set requirements for a self-driving car? For example, one would be "not to pass a red light". How do you specify details for requirements such as 'the detection of the traffic light or a pedestrian', 'the decision to pass or not' or 'the accuracy and precision of such model'. Also, what happens in a situation when an ambulance approaches and the lights are red? Does the car pass the red light to allow the ambulance to pass by, or shall it move to the side of the road [23]? How do we tackle these new issues when existing techniques are no longer adequate?

Google published a new set of guidelines for developing AI applications with a human-centric approach [8]. In this section we provide an insight into Google PAIR's document and map our results from the SLR to the methods used in Google's document. We also present some proposed research recommendations that require attention in future RE4AI research.

A. Google PAIR vs SLR Results

Google's PAIR (People + AI Research) [8] guidelines involved six chapters to creating AI, as shown in Figure 7. While conducting the SLR, we found that most of the focus from an RE perspective was on data requirements and explainability and limited research on identifying user needs. For that reason

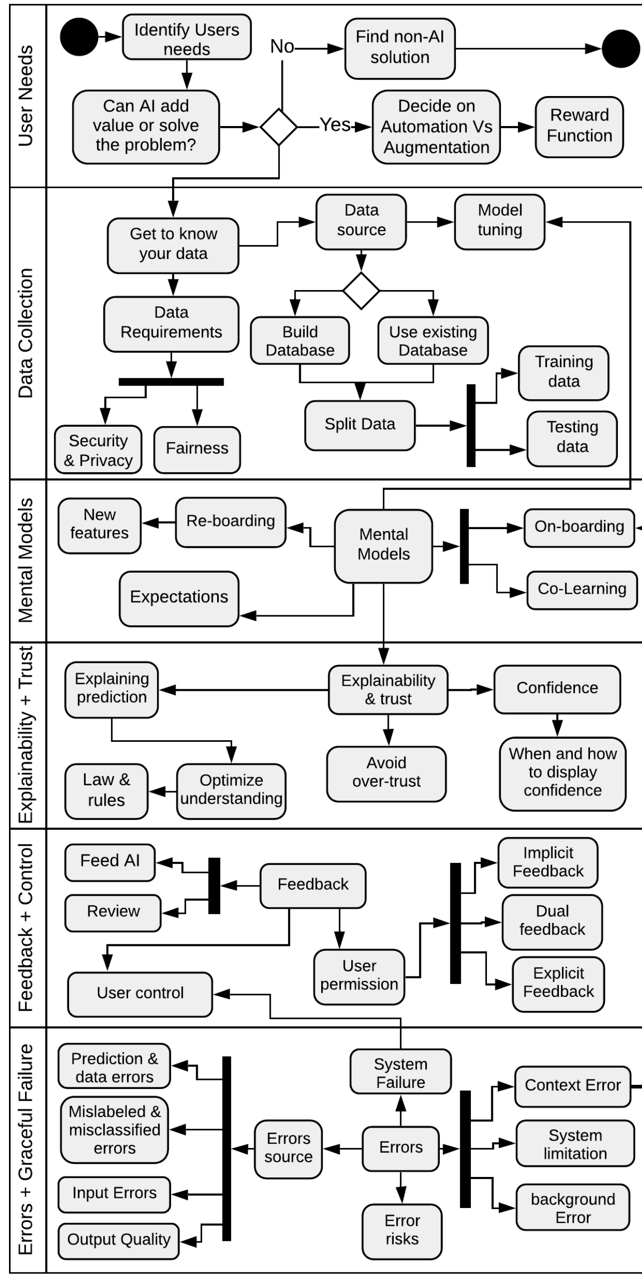

Fig. 7. Googles - People AI (PAIR) breakdown when building AI systems

we only focused on three chapters from Google's document: User needs, data collection and explainability and trust.

User Needs: The guide starts with the importance of users needs and how the first step in building AI powered software systems should involve identifying the problem, and if AI is a beneficial solution. The need for AI should be established before moving on to the next step. Is AI a feasible solution, for example, is there limited information or data. The next step involved deciding on automation vs augmentation and weigh out the reward function (precision vs recall). The studies from the SLR only addressed the first part of the users needs chapter in Google's guidelines, and that was to establish the need for

AI and make sure its a feasible decision [S4].

Data Collection: Google PAIR focused on obtaining, using, and managing data. The guideline pointed to complying with data requirements when collecting data by maintaining privacy and safety measures. And emphasised on avoiding biases by ensuring that the data is inclusive. Protecting information that identified people was crucial when building AI systems, and databases should be sourced responsibly. Additionally, if any issues emerged, they should always map back to the data, features, or labels associated with them.

Whereas, data requirements that appeared in the SLR only covered sampling rate and data quality. Shin et al. concluded that the higher the sampling rate, the better the quality of data was. More samples in the dataset provided a more comprehensive range of diversity when it came to the data [S11]. For data quality, [S18] listed five essential characteristics to having quality data to include accuracy, completeness, consistency, credibility and currentness. However, most of the studies focused on the importance of data requirements for AI systems, since there was limited empirical research available.

Explainable and Trustworthy AI: Google PAIR noted that it is important to provide realistic expectations of the AI model to avoid users over trusting the system. As explanations enforced trust, the guidebook underlined the importance of only explaining what is needed. For example, explain why the system made a given prediction or how specific data was used but leave out the technical detail of how it works. When displaying confidence levels, the guidelines emphasised that it was necessary to determine when and how to display them, as sometimes displaying confidence could lead to miss-trust. The SLR results also indicated that explanations enforced trust [S22].

Although explainable AI systems can help build up trust and strengthen the person's ability to form a more accurate mental model of the product [32]. They are still not applied to AI development as they should. Amershi et al [33]. developed eighteen guidelines for human-centric AI interaction by examining research from over 20 years of human-centric interaction with AI systems. The study involved over 150 AI design recommendations collected from research and industrial sources. The study demonstrated that in current AI systems, most violations made in line with the guidelines were linked to explainability. Explaining programs that provide unpredictable outputs can be a challenging task. However, when given correctly, explanations to predictions can improve peoples choices in decision making when it comes to using AI software [34], [35].

Summary Our results from the SLR showed that there is still a lack of research in the area of RE4AI as shown in our mapping with Google PAIR's document. We found no evident research from an RE perspective on topics such as feedback, mental models, user control and error handing. To validate these results, we need to perform a survey with the industry to find what existing RE techniques are used when building AI systems.

B. Research Recommendations

Based on the results from our literature analysis, we are proposing the following research directions to overcome some of the presenting issues in RE4AI.

Recommendation 1: Is there a need for AI? The first step in building AI-powered software systems should involve identifying if AI is a beneficial choice. Is AI required to provide predictions, personalise or recommendation? Or is there a need for speech and language understanding, image recognition or fraud detection? The need for AI should be established before moving on to the next step [8]. Companies and organisations that choose to use AI as a solution should understand AI's limitations and capabilities. We propose that a checkpoint be maintained to note all required elements needed to create an AI software system. The checkpoint should include what the solution is going to solve, why it is needed, and how it will be used. As well as finding out if the organisation has all the resources needed to build the AI product.

Recommendation 2: How can we extend existing requirements specifications to support AI systems? Most ML systems lack requirements specifications, resulting in poor quality ML systems [19]. This is mainly due to the different methods, techniques, and the emergence of new requirements. For instance, identifying the types of data needed for a project is imperative and should be documented during the RE phase [7]. Not having inclusive data can result in biases when developing AI systems. For example, in a navigation app, a physically healthy person might miss out on potential biases in the app, whereas, a person with physical disabilities might disagree and find the app biased after it provides them with a navigation route with no accessibility features [33].

The same applies to ethical and explainability requirements. How do we explain decisions made by a self-driving car? For example, a car might suddenly brake in front of a bus instead of changing lanes. This decision would be because it weighs between having to injure four people crossing the road vs the one person driving the car getting hurt or possibly killed. Would the driver make the same decision or disagree with the ethical decisions the car company has made on their behalf [20]. What requirements do we need to provide in such cases?

We suggest that to identify and deal with requirements for AI systems is to construct a reference model. The reference model should capture key components and attributes needed when specifying AI system requirements. We propose using research from the industry and literature to plan a matrix that would list all possible requirements for AI systems. The matrix can be used as a guide to map any emerging requirements into the reference model. The model may be broken into separate sub-models to ensure all elements are captured.

Recommendation 3: How do we decide on what modeling language to use? About 60% of the primary studies demonstrated the use of a modeling language to support requirements. Two of the studies proposed the development of a modeling language to support RE4AI. For example,

[S20] suggested developing a modeling language that would capture ethical requirements, and [S19] suggested one to capture NFRs. We suggest using the reference model created from Recommendation 2 to extend or augment an existing modeling language to present AI requirement.

Recommendation 4: How do we bridge the gap between requirements engineers, data scientists and machine learning specialists? Currently, there is a lack of communication and integration between data scientists and software engineers and an apparent gap between these practices [5], [S7]. We propose to create a platform to share and visually present requirements. The platform should allow all sides of the building team to collaborate and share ideas and tools in an environment that could allow aspects of RE and ML to be linked and traced.

Recommendation 5: How do we address issues that are related to calculating the trade-off? Trade-off should be calculated in order to prioritise the importance of requirements. Google PAIR indicated the importance of weighing out the trade-off. For instance, an incorrect prediction in diagnosing a cancer patient would have greater stakes than providing a movie recommendation that the user does not like. When calculating trade-off, what other requirements can make up for the lost cause? For example, the study in [S11] experimented with algorithms to find which one could produce better results with lower costs. They found that specific algorithms performed better than others. So in such cases, the trade-off could be replaced with other measures that can make up for the loss. Another study explained that some algorithms provide more reliable predictions but not easily explained. Whereas others can better explain why a predicted is delivered, but predictions are less in confidence [36]. So how do we decide on which algorithm to choose? In what situations do we prefer to use explainable algorithms vs higher confidence.

Google PAIR pointed out the importance of calculating the trade-off of the reward function. Moreover, to evaluate and weigh the risks of choosing an appropriate reward function that would suit the users needs accordingly. For a notification system in an autonomous car, a false negative would *not* notify a sleeping driver in case of an emergency (precision), which could lead to deadly consequences. While having too many notifications that are false positives (recall) can lead the driver to ignore them [30]. When building AI software, we should always calculate the trade-off between precision and recall. When can we choose precision over recall or vice versa. For example, Dimatteo et al. [30] explained that in the notification system finding alternative human-centric ways to engage the driver, using recall might be a more feasible and safer choice. We propose that the trade-off be displayed along with requirements when modelling with a list of outcomes as to why the decision was made.

Recommendation 6: How do we identify the different software engineering methods and techniques that should replace existing ones? As the process of RE is changing; new techniques should be provided. For example, methods that require human intervention in gathering data are now being replaced by new forms of data collection such as online forms and social media, sensors and feedback from users [37]. Data collected from such sources require new RE techniques to elicit and manage them. We propose to create a taxonomy that will list all the different techniques and methods required when building AI systems.

VI. Conclusion

AI-based techniques have recently become much more embedded into many software systems and are increasingly used by companies to improve performance and reduce costs. However, using existing RE techniques for current AI systems is challenging due to the different nature of the development process between traditional software engineering methods and AI-based systems. Current AI systems also show a lack of integration with existing RE tools and methodologies, with limited research on the topic. The SLR described in this paper identified the different modeling techniques and requirements notations used in current studies. Our results show that most studies favoured UML and GORE to model requirements. We also found that application domains such as automated-driving and computer vision were more popular than others, and research areas such as ethics lacked empirical evaluations. Our findings identified that many issues and challenges exist in current RE4AI techniques. For example, defining requirements, explaining predictions or addressing ethical issues and data requirements. Another major issue presented in the literature was the lack of integration between the software engineers and data scientists. We concluded by providing a comparison between our results obtained from the SLR and Google's PAIR. We also provided a set of key recommendations for further research. With the lack of current practices available, there is a need to introduce and research new methodologies alongside integrating existing RE techniques. The next step should involve documenting any requirements for AI systems, identifying modeling languages and creating a platform for requirements engineers and data scientists to collaborate and share their ideas.

Acknowledgements

Grundy is supported by ARC Laureate Fellowship FL190100035.

Included Primary Studies

[S1] M. Bonfe, F. Boriero, R. Dodi, P. Fiorini, A. Morandi, R. Muradore, L. Pasquale, A. Sanna, and C. Secchi, "Towards automated surgical robotics: A requirements engineering approach," in *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 2012, pp. 56–61.

[S2] B. C. Hu, R. Salay, K. Czarnecki, M. Rahimi, G. Selim, and M. Chechik, "Towards requirements specification for machine-learned perception based on human performance," in *Workshop on Artificial Intelligence for Requirements Engineering (AIRE)*. IEEE, 2020.

[S3] K. Gruber, J. Huemer, A. Zimmermann, and R. Maschotta, "Integrated description of functional and non-functional requirements for automotive systems design using SysML," in *2017 7th IEEE International Conference on System Engineering and Technology (ICSET)*. IEEE, 2017, pp. 27–31.

[S4] K. Sandkuhl, “Putting AI into context-method support for the introduction of artificial intelligence into organizations,” in *2019 IEEE 21st Conference on Business Informatics (CBI)*, vol. 1. IEEE, 2019, pp. 157–164.

[S5] M. Rahimi, J. L. Guo, S. Kokaly, and M. Chechik, “Toward requirements specification for machine-learned components,” in *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*. IEEE, 2019, pp. 241–244.

[S6] J. Jakob, K. Kugele, and J. Tick, “Defining camera-based traffic scenarios and use cases for the visually impaired by means of expert interviews,” in *2017 IEEE 14th International Scientific Conference on Informatics*. IEEE, 2017, pp. 128–133.

[S7] A. Vogelsang and M. Borg, “Requirements engineering for machine learning: Perspectives from data scientists,” *IEEE International Requirements Engineering Conference Workshops*, 2019.

[S8] C. E. Tuncali, G. Fainekos, D. Prokhorov, H. Ito, and J. Kapinski, “Requirements-driven test generation for autonomous vehicles with machine learning components,” *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 2, pp. 265–280, 2019.

[S9] J. S. Becker, “Partial consistency for requirement engineering with traffic sequence charts.” in *Software Engineering (Workshops)*, 2020.

[S10] J. Bosch, H. H. Olsson, and I. Crnkovic, “It takes three to tango: Requirement, outcome/data, and AI driven development.” in *SiBW*, 2018, pp. 177–192.

[S11] C. Shin, S. Rho, H. Lee, and W. Rhee, “Data requirements for applying machine learning to energy disaggregation,” *Energies*, vol. 12, no. 9, p. 1696, 2019.

[S12] G. Dimitrakopoulos, E. Kavakli, P. Loucopoulos, D. Anagnostopoulos, and T. Zographos, “A capability-oriented modelling and simulation approach for autonomous vehicle management,” *Simulation Modelling Practice and Theory*, vol. 91, pp. 28–47, 2019.

[S13] D. Weihrauch, P. A. Schindler, and W. Sihn, “A conceptual model for developing a smart process control system,” *Procedia CIRP*, vol. 67, pp. 386–391, 2018.

[S14] S. Fenn, A. Mendes, and D. M. Budden, “Addressing the non-functional requirements of computer vision systems: a case study,” *Machine Vision and Applications*, vol. 27, no. 1, pp. 77–86, 2016.

[S15] K. Neace, R. Roncace, and P. Fomin, “Goal model analysis of autonomy requirements for unmanned aircraft systems,” *Requirements Engineering*, vol. 23, no. 4, pp. 509–555, 2018.

[S16] J. Lockerbie and N. A. Maiden, “Using a requirements modelling language to co-design intelligent support for people living with dementia.” in *REFSQ Workshops*, 2020.

[S17] K. Nakamichi, K. Ohashi, I. Namba, R. Yamamoto, M. Aoyama, L. Joeckel, J. Siebert, and J. Heidrich, “Requirements-driven method to determine quality characteristics and measurements for machine learning software and its evaluation,” in *2020 IEEE 28th International Requirements Engineering Conference (RE)*. IEEE, 2020, pp. 260–270.

[S18] H. Challa, N. Niu, and R. Johnson, “Faulty requirements made valuable: On the role of data quality in deep learning,” in *2020 IEEE Seventh International Workshop on Artificial Intelligence for Requirements Engineering (AIRE)*. IEEE, 2020, pp. 61–69.

[S19] J. Horkoff, “Non-functional requirements for machine learning: Challenges and new directions,” in *2019 IEEE 27th International Requirements Engineering Conference (RE)*. IEEE, 2019, pp. 386–391.

[S20] F. B. Aydemir and F. Dalpiaz, “A roadmap for ethics-aware software engineering,” in *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)*. IEEE, 2018, pp. 15–21.

[S21] B. Bruno, F. Mastrogiovanni, and A. Sgorbissa, “Functional requirements and design issues for a socially assistive robot for elderly people with mild cognitive impairments,” in *2013 IEEE RO-MAN*. IEEE, 2013, pp. 768–773.

[S22] M. A. Köhl, K. Baum, M. Langer, D. Oster, T. Speith, and D. Bohlender, “Explainability as a non-functional requirement,” in *2019 IEEE 27th International Requirements Engineering Conference (RE)*. IEEE, 2019, pp. 363–368.

[S23] H. H. Altarturi, K.-Y. Ng, M. I. H. Ninggal, A. S. A. Nazri, and A. A. Abd Ghani, “A requirement engineering model for big data software,” in *2017 IEEE Conference on Big Data and Analytics (ICBDA)*. IEEE, 2017, pp. 111–117.

[S24] J. K. Ang, S. B. Leong, C. F. Lee, and U. K. Yusof, “Requirement engineering techniques in developing expert systems,” in *2011 IEEE Symposium on Computers & Informatics*. IEEE, 2011, pp. 640–645.

[S25] F. Ishikawa and Y. Matsuno, “Evidence-driven requirements engineering for uncertainty of machine learning-based systems,” in *2020 IEEE 28th International Requirements Engineering Conference (RE)*. IEEE, 2020, pp. 346–351.

[S26] H. Kuwajima and F. Ishikawa, “Adapting SQuaRE for quality assessment of artificial intelligence systems,” in *2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*. IEEE, 2019, pp. 13–18.

[S27] G. Amaral, R. Guizzardi, G. Guizzardi, and J. Mylopoulos, “Ontology-based modeling and analysis of trustworthiness requirements: Preliminary results,” in *International Conference on Conceptual Modeling*. Springer, 2020, pp. 342–352.

References

[1] L. E. Holmquist, “Intelligence on tap: artificial intelligence as a new design material,” *interactions*, vol. 24, no. 4, pp. 28–33, 2017.

[2] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo, and D. Dennison, “Hidden technical debt in machine learning systems,” in *Advances in neural information processing systems*, 2015, pp. 2503–2511.

[3] A. Arpteg, B. Brinne, L. Crnkovic-Friis, and J. Bosch, “Software engineering challenges of deep learning,” in *2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, 2018, pp. 50–59.

[4] H. Belani, M. Vukovic, and Ž. Car, “Requirements engineering challenges in building AI-based complex systems,” in *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*. IEEE, 2019, pp. 252–255.

[5] L. E. Lwakatare, A. Raj, J. Bosch, H. H. Olsson, and I. Crnkovic, “A taxonomy of software engineering challenges for machine learning systems: An empirical investigation,” in *International Conference on Agile Software Development*. Springer, 2019, pp. 227–243.

[6] S. Amershi, A. Begel, C. Bird, R. DeLine, H. Gall, E. Kamar, N. Nagappan, B. Nushi, and T. Zimmermann, “Software engineering for machine learning: a case study,” in *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice*. IEEE Press, 2019, pp. 291–300.

[7] D. Kondermann, “Ground truth design principles: an overview,” in *Proceedings of the International Workshop on Video and Image Ground Truth in Computer Vision Applications*, 2013, pp. 1–4.

[8] Google Research. (2019) The people + AI guidebook. [online] Available at: https://research.google/teams/brain/pair/ [Accessed 1 April 2020].

[9] J. Dick, E. Hull, and K. Jackson, *Requirements engineering*. Springer, 2017.

[10] L. S. Wheatcraft and M. J. Ryan, “Communicating requirements–effectively!” in *INCOSE International Symposium*, vol. 28, no. 1. Wiley Online Library, 2018, pp. 716–732.

[11] B. Nuseibeh and S. Easterbrook, “Requirements engineering: a roadmap,” in *Proceedings of the Conference on the Future of Software Engineering*, 2000, pp. 35–46.

[12] G. Koelsch, *Requirements writing for system engineering*. Springer, 2016.

[13] I. Inayat, S. S. Salim, S. Marczak, M. Daneva, and S. Shamshirband, “A systematic literature review on agile requirements engineering practices and challenges,” *Computers in human behavior*, vol. 51, pp. 915–929, 2015.

[14] A. Lapouchnian, “Goal-oriented requirements engineering: An overview of the current research,” *University of Toronto*, vol. 32, 2005.

[15] D. Amyot and G. Mussbacher, “User requirements notation: the first ten years, the next ten years,” *JSW*, vol. 6, no. 5, pp. 747–768, 2011.

[16] E. Gonçalves, M. A. de Oliveira, I. Monteiro, J. Castro, and J. Araújo, “Understanding what is important in iStar extension proposals: the viewpoint of researchers,” *Requirements Engineering*, vol. 24, no. 1, pp. 55–84, 2019.

[17] M. Agarwal and S. Goel, “Expert system and it’s requirement engineering process,” in *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)*. IEEE, 2014, pp. 1–4.

[18] F. Khomh, B. Adams, J. Cheng, M. Fokaefs, and G. Antoniol, “Software engineering for machine-learning applications: The road ahead,” *IEEE Software*, vol. 35, no. 5, pp. 81–84, 2018.

[19] H. Kuwajima, H. Yasuoka, and T. Nakae, “Engineering problems in machine learning systems,” *Machine Learning*, pp. 1–24, 2020.

[20] L. M. Cysneiros, M. Raffi, and J. C. S. do Prado Leite, “Software transparency as a key requirement for self-driving cars,” in *2018 IEEE 26th International Requirements Engineering Conference (RE)*. IEEE, 2018, pp. 382–387.
[21] H.-L. E. G. on Artificial Intelligence, “Ethics guidelines for trustworthy AI,” 8 April 2019, document can be accessed at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.
[22] L. M. Cysneiros and J. C. S. do Prado Leite, “Non-functional requirements orienting the development of socially responsible software,” in *Enterprise, Business-Process and Information Systems Modeling*. Springer, 2020, pp. 335–342.
[23] F. Houdek and S. Schmerler, “Automotive future and its impact on requirements engineering.” in *REFSQ Workshops*, 2017.
[24] B. Kitchenham and S. Charters, “Guidelines for performing systematic literature reviews in software engineering,” 2007.
[25] C. Wohlin, “Guidelines for snowballing in systematic literature studies and a replication in software engineering,” in *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, 2014, pp. 1–10.
[26] C. Urquhart, *Grounded theory for qualitative research: A practical guide*. Sage, 2012.
[27] L. Baresi, L. Pasquale, and P. Spoletini, “Fuzzy goals for requirements-driven adaptation,” in *2010 18th IEEE international requirements engineering conference*. IEEE, 2010, pp. 125–134.
[28] E. Bartocci, J. Deshmukh, A. Donzé, G. Fainekos, O. Maler, D. Ničković, and S. Sankaranarayanan, “Specification-based monitoring of cyber-physical systems: a survey on theory, tools and applications,” in *Lectures on Runtime Verification*. Springer, 2018, pp. 135–175.
[29] L. Chazette, “Mitigating challenges in the elicitation and analysis of transparency requirements,” in *2019 IEEE 27th International Requirements Engineering Conference (RE)*. IEEE, 2019, pp. 470–475.
[30] J. DiMatteo, D. M. Berry, and K. Czarnecki, “Requirements for monitoring inattention of the responsible human in an autonomous vehicle: The recall and precision tradeoff.” in *REFSQ Workshops*, 2020.
[31] X. Zhou, Y. Jin, H. Zhang, S. Li, and X. Huang, “A map of threats to validity of systematic literature reviews in software engineering,” in *2016 23rd Asia-Pacific Software Engineering Conference (APSEC)*. IEEE, 2016, pp. 153–160.
[32] M. M. De Graaf and B. F. Malle, “How people explain action (and autonomous intelligent systems should too),” in *2017 AAAI Fall Symposium Series*, 2017.
[33] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen *et al.*, “Guidelines for human-AI interaction,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.
[34] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim, “Designing theory-driven user-centric explainable AI,” in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–15.
[35] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza, “Power to the people: The role of humans in interactive machine learning,” *Ai Magazine*, vol. 35, no. 4, pp. 105–120, 2014.
[36] J. Krause, A. Perer, and K. Ng, “Interacting with predictions: Visual inspection of black-box machine learning models,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 5686–5697.
[37] B. Kostova, S. Gurses, and A. Wegmann, “On the interplay between requirements, engineering, and artificial intelligence.” in *REFSQ Workshops*, 2020.