Technical Report on Risk Assessment of Safety-critical Socio-technical Systems: A Systematic Literature Review

Soheila Sheikh Bahaei and Barbara Gallina
School of Innovation, Design and Engineering
Mälardalen University
Västerås, Sweden
{soheila.sheikh.bahaei, barbara.gallina}@mdu.se

***Abstract*—One of the most important activities to ensure safety of safety-critical (socio-technical) systems is risk assessment. To facilitate this activity, various techniques have been proposed for e.g., modeling and analyzing the behavior and the interactions of system entities. In addition, standards have been developed to collect best practices for conducting such activity. What is still lacking is a comprehensive and systematic literature review (SLR) characterizing works on risk assessment of safety-critical socio-technical systems based on the evolution of the conceptualization of socio-technical systems including organizational and technological changes such as digitalization/globalisation, inclusion of augmented reality (AR), evolution of safety standards and safety perspectives. Hence, to be able to investigate the current status of the topic, in this paper, we undertake a SLR of primary studies reporting techniques for risk assessment of safety-critical socio-technical systems. More specifically, we identify and review the available risk assessment techniques and we characterize and analyze them based on how they conceptualize technical and socio aspects, their orchestration, organizational and technological changes effects, AR effects, risk assessment process. In addition, we consider their safety perspective, modeling formality, type of analysis, tool support, application domain and supported standards. Finally, we provide our findings and possible future works based on the analysis of the primary studies and their challenges.**

***Index Terms*—Risk assessment, Socio-technical systems, Safety standards, Safety-criticality**

1 Introduction

Risk assessment is an essential part of the activities required for ensuring safety of safety-critical (socio-technical) systems. Based on standard ISO 31000:2018 [1], which is a generic standard in risk management, the steps of risk assessment are risk identification, risk analysis and risk evaluation. Socio-technical systems are systems including technical and socio entities such as human and organization [2]. Safety-critical systems are "systems whose failure could result in loss of life, significant property damage, or damage to the environment" [3]. In order to assess risk of safety-critical socio-technical systems, risk sources related to socio aspects and their interactions should be considered in addition to risk sources related to technical aspects. There are various techniques for modeling the system entities and their interactions and also for analyzing system behavior that can be used for risk assessment of safety-critical socio-technical systems. However, to be effective, these techniques shall evolve in alignment with the evolution of the conceptualization of safety-critical socio-technical systems. Current socio-technical systems include organizational and technological changes which have the potential to introduce new risk sources. Thus, it is essential to strengthen conceptualization of socio-technical systems and embed such conceptualization within modelling and analysis techniques.

Organizational changes such as globalization, digitalization and appearance of organization networks, besides the provided progress, may lead to new kinds of system risks. In [4], organizational changes over the last two to three decades are discussed and, in [5], it is discussed that it is essential to address new types of system risks due to the new organizational changes.

In addition, new technological changes such as using augmented reality (AR) as human-machine interface, besides the provided improvements, may introduce new kinds of risks to the system. Challenges and risks of using AR in safety-critical applications are discussed in [6] and a method for risk analysis of critical AR applications is proposed in [7].

Furthermore, standards, specifically safety standards and more broadly dependability standards, have been developed to collect best practices for conducting risk assessment. In this work we do not focus on a specific domain, nevertheless we recall information about standards from the automotive domain to be used as an example.

There are various papers reviewing literature on the topic of risk assessment of socio-technical systems considering different research questions. For example, in [8], authors conduct a SLR and report about risk assessment methods to find out the extent to which they support systems thinking. Based on this SLR, the majority of methods exclusively focus on human error. Hence, the methods only focus on the

This work is funded by EU H2020 MSC-ITN grant agreement No 764951. Author B. Gallina is also partially supported by Sweden's Knowledge Foundation via SACSys (Safe and Secure Adaptive Collaborative Systems) project.

human entities of the socio-technical systems and they do not consider safety as a system property. In [9], authors provide a scoping literature survey on applications of STAMP (System-Theoretic Accident Model and Processes) [10] for analyzing socio-technical systems and its associated techniques, STPA (System-Theoretic Process Analysis) [11] and CAST (Causal Analysis based on System Theory). In this survey features of these methods, their methodological steps and their enrichment are presented.

What we still miss in the literature is a SLR based on the evolution of the conceptualization of socio-technical systems which may include technological changes such as AR, organizational changes such as digitalization/globalisation and by considering evolution of safety standards. It is crucial to investigate the development of interpretation of risk assessment and socio-technical systems over time for characterizing technical, human and organizational aspects and effects of new technological and organizational changes.

In this paper, we conduct a SLR based on development of current techniques for risk assessment of safety-critical socio-technical systems and we define our specific research questions. We undertake the SLR based on the guidelines proposed by Kitchenham and Charters [12] and we aim at identifying primary studies contributing to risk assessment of safety-critical socio-technical systems considering their evolution, analyzing them and providing our interpretation on evolution of socio-technical systems' conceptualization. The purpose of our SLR is threefold: first, to provide an overview regarding the evolution of research regarding risk assessment of safety-critical socio-technical systems. Second, to provide a summary of current techniques based on the evolution of socio-technical systems' conceptualization. Third, to extract and report about their challenges and provide research directions for future works based on the findings.

The paper is organized as follows. In Section 2, we recall the background and discuss related work. In Section 3, we present the research method. In Section 4, we report about the results of the SLR, which we conducted. In Section 5, we discuss the results and threats to validity. In Section 6, we draw our conclusion and we present potential future research directions based on our findings.

2 Background and Related Work

2.1 Risk Assessment of Safety-Critical Socio-technical Systems - Basic Concepts

Based on standard ISO 31000:2018 [1], *risk* means "effect of uncertainty on objectives" and *effect* is "deviation from the expected". *Risk* is usually expressed in terms of risk sources, potential events, their consequences and their likelihood". Based on this standard, risk assessment contains *risk identification*, *risk analysis* and *risk evaluation*. In *risk identification*, the objective is to find, recognize and describe risks. In *risk analysis* the objective is to understand the nature of the risk, its characteristics and considering uncertainties, risk sources, consequences, likelihood, events, scenarios, controls and their effectiveness. Finally, in *risk evaluation*, the objective is to support decisions by comparing the risk analysis results with the criteria to determine required actions. These steps are also included and refined in the domain-specific safety standards. For example, *ISO 26262* [13], which is the functional safety standard in the automotive domain, provides the set of activities that should be performed during safety lifecycle. In this standard, risks emanated from technical failures are addressed and, to be able to assess risk, ASILs (Automotive Safety Integrity Levels) are determined. ASILs are determined based on severity, exposure and controllability factors. The severity factor is determined based on severity in case of hazard occurrence. The exposure factor is determined based on probability of exposure with respect of operational situations. The controllability factor is determined based on operator controllability. In addition, safety goals are defined to prevent unreasonable risk. *ISO 21448:2022* [14] defined as *SOTIF (Safety Of The Intended Functionality)* addresses risks due to hazards resulting from functional insufficiencies of the intended functionality or its implementation. This standard considers risks emanated from non-technical behaviors, such as operator's incorrect deciding which would lead to system risk. In this standard ASIL is not determined, however severity and controllability are determined and qualitative analysis is used to define safety measures to improve the *SOTIF*.

As explained in Section 1, socio-technical systems are systems including technical and socio entities such as human and organization [2]. Accordingly, the socio related risks and the risks related to socio and technical teaming are as important as technical-related risks to be considered in the risk assessment process. In addition, considering safety standards are extremely important specially in safety-critical systems. Regarding approaches and other practices for performing risk assessment/safety analysis/hazard analysis, it is worth to mention that a discussion about the validity of basic approaches is ongoing since 2015. This discussion has led to the introduction of specific labels, i.e., *Safety I*, *Safety II*, and *Safety III* to categorize different practices. A comparison between these labels or safety perspectives is shown in Table I. *Safety I* is defined by Erik Hollnagel as the "condition where the number of adverse outcomes (e.g., accidents, incidents and near misses) is as low as possible" [15]. Erik Hollnagel believes that what is done in industry to prevent accidents is based on this definition. To overcome the current limitations caused by increasing the complexity and demands of new systems, he proposes *Safety II* defined as the "condition where the number of acceptable outcomes is as high as possible. It is the ability to succeed under varying conditions" [15]. On the other hand, Nancy Leveson disagrees about the existence of *Safety I* and she believes there is no unique approach used in all industries. She believes *Safety II* is not effective and has been used in the past. Accordingly, she proposes *Safety III* as the "freedom from unacceptable losses as identified by the system stakeholders. The goal is to eliminate, mitigate, or control hazards, which are the states that can lead to these losses" [16]. In summary, based on [17], in *Safety I* there is special focus on malfunctions or failures of specific components such

as technical, human and organizational components leading to system accidents or losses and the aim is to identify and manage hazards and their consequences. In *Safety II*, there is special focus on human role and the aim is to ensure as many things as possible go right. In *Safety III*, there is special focus on interactions and the aim is to control hazards leading to unacceptable losses by enforcing safety-related constraints. Based on [16], *Safety I* is not *reactive* as described in [15] and the reason is that everyone learns from accidents and use them for improving safety and controlling system in the future. Thus, it contradicts with the definition of *reactive*, which means *acting in response to a situation rather than controlling it*. In [16], what is actually done today is called *safety engineering today* and it is discussed that what is done in *safety engineering today* is quite different from *safety I*, *safety II* and *safety III*. In *safety engineering today*, the purpose is to identify the linear chain of events and there is special focus on root cause of an accident, while in *safety III*, linear causality is not assumed and there is no root cause. It also discusses about *safety II* and explains that it is linear because of the existence of causality as a chain (sequence) of events while each event is defined by a necessary and sufficient relationship with a preceding event. In addition, it explains that *safety II* mostly concentrates on human, while the system design seems to be ignored. In contrast, *safety III* is based on System Theory and considers human as part of system containing technical and other aspects. It also emphasizes on interactions between components that would act as causes of hazards.

2.2 Related Work

A review of advances on the foundation of risk assessment and risk management is performed in [18]. Based on this review risk assessment and risk management as a scientific field is not more than 30-40 years old, however, the concept has been available since more than 2400 years. In this study, it is explained that risk field is divided into two groups. The first group is populated by studies on using "the risk assessment and risk management to study and treat the risk of specific activities" and the second group is populated by studies on "generic risk research and development related to concepts, theories, frameworks, approaches, principles, methods and models to understand, assess, characterize, communicate and (in a wide sense) manage/govern risk". Based on the review provided in this study, it is required to develop more modeling and analyzing techniques to be used for new types of systems such as critical infrastructures and complex systems. In addition, this review points out that risks related to socio aspects are still challenging and need more contributions.

A review of developments of accident investigation methods used for improving hazard identification is provided in [19]. As it is discussed in this study, human imagination and inventiveness are essential to incorporate various possible scenarios in both hazard identification and accident investigation. It is more straightforward to consider accidents in order to identify hazards, since it is not possible to have a complete prediction of what potentially can go wrong. Different accident investigation methods are reviewed and it is discussed that socio-technical systems approaches consider the whole systems containing social factors, however the results are still dependent on experience, knowledge and effort of the analyst.

A review and assessment of safety analysis methods is prepared in [20] to be used for improving occupational safety in industry 4.0. A total of 47 essential methods in occupational health and safety (OHS) are reviewed and based on this study, the previous literature are not able to deal with new system properties introduced by industry 4.0. This paper presents key features of Industry 4.0 as "interconnectivity, autonomous systems, automation in joint human-agent activity and a shift in supervisory control", which introduce new challenges in system safety. It discusses that complexity-thinking methods are beneficial for analysis of new complex systems. However, there is a need for new methods to overcome the challenges.

A systematic literature review is provided in [21] on the state of the practice in validation of model-based safety analysis for socio-technical systems (using PRISMA protocol). The analysis in this study covers articles published in period of ten years (2010-2019) in safety science journal. The results reveal that 63% of the articles which propose a new safety model do not provide validation and there is no increasing or decreasing trend in providing validation during the years. There is also no correlation between validation and other investigated variables such as safety concept, model type/approach, stage of the system lifecycle, country of origin or industrial application domain. In addition, in the remaining 37% of the articles, a variety of views on validation is represented. For example, the identified categories are *benchmark exercise, peer review, reality check, quality assurance, validity text, statistical validation* and *illustration*, while it is discussed in this paper that these are not adequate for validating a model comprehensively. It also discusses that lack of focus on validation and using different terminologies referring to validation are common in various industrial application domains. It is therefore suggested to have increased attention to the meaning of validation in safety analysis context in addition to developing a validation framework clarifying validation function(s).

A systematic literature review is provided in [22] on risk factors for human-robot collaboration from system-wide perspective. It considers papers published in the years 2011 – 2021 and 32 papers are analyzed from which 254 risk factors (RFs) are identified. The RFs are classified to five classes and each class contains at least two sub-classes. The identified classes are: 1) *Human*, 2) *Technology*, 3) *Collaborative workspace*, 4) *Enterprise*, 5) *External*. It is discussed in this paper that the identified classes can be used as the fundamental building blocks of a safety evaluation framework considering socio-technical thinking.

These works consider various perspectives of risk assessment in socio-technical systems and they concentrate on different defined research questions. However, there is no systematic literature review (following a protocol) considering conceptualization of evolution of socio-technical systems in the risk assessment process. Due to the broad effects of

Safety Perspective	Definition	Defined by	Special focus on	Type of assumed cauality
Safety I	condition where the number of adverse outcomes is as low as possible	Erik Hollnagel	malfunctions or failures of specific components	Linear
Safety II	condition where the number of acceptable outcomes is as high as possible	Erik Hollnagel	human role	Linear
Safety III	freedom from unacceptable losses as identified by the system stakeholders	Nancy Leveson	interactions	Non Linear
Safety engineering today	freedom from unacceptable losses as identified by the stakeholders, but may be defined in terms of acceptable risk or ALARP in some fields	Nancy Leveson	root cause of an accident	Linear

TABLE I: Comparison between safety perspectives

organizational and technological changes in the recent socio-technical systems, it is essential to consider the evolution in the modeling and analysis phases of risk assessment process to be able to prevent new risks caused by these new changes. In this study, we define our specific research questions concentrating on conceptualization of evolution of socio-technical systems.

3 Research Method

This section describes our research method, which is based on the guidelines for SLR proposed by Kitchenham and Charters [12]. Based on this guideline, an SLR has three main phases, which are briefly recalled in what follows:

1) **Planning the SLR**: In this stage, a plan should be determined for the SLR. This plan includes the following stages:
   - Identifying the need for an SLR: In this stage, the reasons for the SLR and its scope should be clarified.
   - Specifying goal and research questions: In this stage, goal of the SLR and research questions should be defined.
   - Designing the SLR protocol: In this stage, the SLR protocol should be developed by defining search strategy, study selection criteria, study selection procedure, study quality assessment criteria. Search strategy defines search terms and databases that can be used for searching the primary studies. Study selection criteria determines which primary studies should be included and which ones should be excluded. Study selection procedure describes how to apply the study selection criteria. Finally, study quality assessment criteria provide more detailed inclusion/exclusion criteria.
2) **Conducting the SLR**: In this stage, the SLR should be conducted based on the planning. The tasks in this stage are data collection including research identification, selection of primary studies, quality assessment and data extraction.
3) **Reporting the results of the SLR**: In this stage, mechanisms should be defined in order to illustrate results of the SLR and their analysis.

3.1 Planning the SLR

This subsection describes the execution of the recalled phases.

3.1.1 Identifying the Need for a SLR:
The primary goal in risk assessment activities is to prevent unreasonable risk to have an acceptable level of safety. Especially in safety-critical applications it is of high importance, because risks may lead to human loss or injury or can be harmful for the environment. Since there is an increasing use of AR as human-machine interface, it is greatly important to consider AR-related aspects of the system during the risk assessment. In addition, as mentioned in Section 1, new organizational changes may lead to new risks. Hence, it is essential to address their effects on human performance and on influencing factors on human performance during the risk assessment process. In order to investigate the development of conceptualization of risk assessment in socio-technical systems, a SLR can be of value. There are some techniques proposed to assess risk of safety-critical socio-technical systems containing new technological and organizational changes. However, no SLR has been conducted to characterize these techniques based on the evolution of the conceptualization of socio-technical systems including organizational and technological changes such as digitalization/globalisation, inclusion of augmented reality, and evolution of safety standards. Thus, we identify the need to provide a SLR to enable characterizing the available techniques and to provide an overview regarding the evolution of research in this context.

Scope: Based on the guideline proposed by Cooper [23], we determine our focus, goal, representation perspective, coverage, organizing method, and audience. Our focus is on the research outcomes of the available literature developing conceptualization of safety-critical socio-technical systems for being used in the risk assessment techniques. Our goal is to characterize (describe) available literature in this area based on our defined research questions to be able to provide an overview regarding the evolution of the research in risk assessment of safety-critical socio-technical systems. Our representation perspective is neutral, meaning that we present evidence and argument represented by authors without accumulating and synthesizing our viewpoint in the editorial process. We

aim at implementing exhaustive coverage by defining an inclusive review protocol. We organize the review historically, meaning that we introduce the works in chronological order in which they emerge in the literature. Our audience are specialized scholars, practitioners, AR developers, manufacturers of safety-critical systems and safety and reliability engineering communities.

3.1.2 Specifying Goal and Research Questions:

Goal: The goal in this SLR is to characterize the current state-of-the-art regarding risk assessment of safety-critical socio-technical systems based on the evolution of the conceptualization of socio-technical systems. Assessing the risk in safety-critical socio-technical systems requires characterizing socio aspects in addition to technical aspects and it is also important to consider new risks/dependability threats and their interactions. There are different modeling languages and techniques for modeling and analyzing system behavior which provide different levels of automation. These languages and techniques may be capable to be used in different domains or a specific domain. They would support safety standards or do not provide any standard compliance helpful for safety-critical applications. Various scenarios would be presented to demonstrate modeling and analysis capabilities of the languages and techniques. Thus, it is essential to consider languages and techniques used in the literature to be able to illustrate their development over time and to be able to understand the limitations and challenges.

Research Questions: By considering the goal of the SLR we formulate our specific research questions as follows:

- **RQ1**: How interpretation/conceptualization of risk assessment and socio-technical systems evolved over time? (Are there structured conceptualization (there are concepts and well-formedness rules to relate concepts used for characterization), potential for capturing (there are concepts which provides the potential for characterizing) or no characterization (there is no possibility for characterizing)?)
  - 1.1. How human aspects are characterized?
  - 1.2. How organizational aspects are characterized?
  - 1.3. How technical aspects are characterized?
  - 1.4. How orchestration/concertation of socio and technical aspects is characterized? (How the coordination and interactions between socio and technical aspects are characterized?)
  - 1.5. How effects of organizational changes are characterized?
  - 1.6. How effects of technological changes are characterized?
  - 1.7. How AR effects are characterized?
  - 1.8. How risks and dependability threats are characterized?
  - 1.9. Which steps of the risk assessment process are provided/developed? (risk identification, risk analysis, risk evaluation (based on the provided explanation in Section 2))
  - 1.10. Which safety perspective is supported? (safety I, safety II, safety III or safety engineering today (based on Table I))
- **RQ2**: What are the characteristics of the methods described in the primary studies?
  - 2.1. Which is the level of formality of the modeling used to model system entities and their relationships? (Are there semi-formal (defined concepts, formal syntax, but informal semantics), formal (well defined concepts, formal syntax and formal semantics) or informal languages/notations (defined concepts, but informal syntax and informal semantics)?)
  - 2.2. Is the contribution related to extending concepts, syntax or semantics of modeling languages or none of them?
  - 2.3. Which are the techniques for analyzing system behavior? (Are they qualitative/quantitative/both, linear/non-linear, forward looking (predictive)/backward looking (investigative)?)
  - 2.4. Which is the level of automation? (Is it tool-supported?)
- **RQ3**: What is the potential impact/applicability of the proposed methods?
  - 3.1. What are the application domains? (Is it for specific domain or general application?)
  - 3.2. What are the supported standards, if any? (Is there discussion about any support for standards?)
  - 3.3. What are the types of illustrative scenarios presented? (Are there scenarios presented?)
- **RQ4**: What challenges are identified in the primary studies?

  We define abbreviations for different possible options in relation to research questions to be used for summarizing the extracted information from primary studies, shown in Fig 1.

3.1.3 Designing the SLR Protocol:

In this subsection, we present our plan for the SLR and design our SLR protocol.

Search Strategy: In order to identify possible primary studies, it is required to use specific terms and define search string. We use *PICO (Population, Intervention, Comparison, Outcomes)* criteria based on [24] to define the search elements. *Population* might be a specific role or an application area e.g. safety-critical socio-technical systems. *Intervention* is the methodology/tool/technology/procedure that addresses a specific issue. For example, in our SLR, risk assessment, modeling technique, analysis technique can be considered as *intervention*. *Comparison* is the methodology/tool/technology/procedure with which the intervention is being compared. For example, in our SLR, safety standards can be used for comparing different techniques. *Outcomes* refers to factors of importance to practitioners. For example, in our SLR, modeling and analysis capabilities can be considered as outcomes. Based on *PICO* criteria, factors of importance in our SLR are as follows:

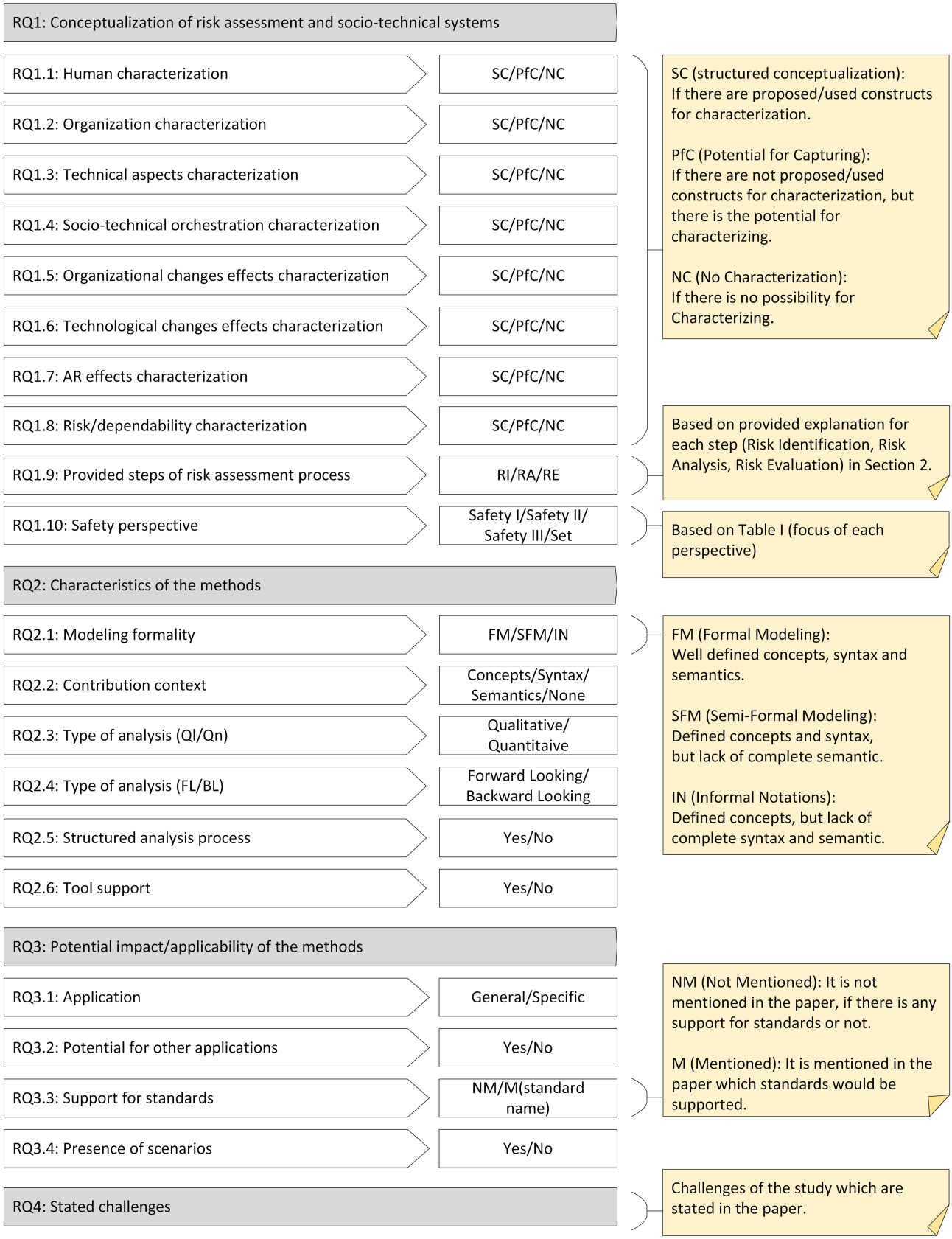

Fig. 1: Defined abbreviations for possible options of extracted information in relation to each research question

- **Population**: safety-critical socio-technical systems
- **Intervention**: risk assessment, modeling technique, analysis technique
- **Comparison**: safety standards
- **Outcomes**: modeling and analysis capabilities

In addition to these factors of importance, we use synonyms of these terms in the literature. Based on literature human-machine systems are synonym for socio-technical systems. Thus, we consider "socio-technical systems" or "human-machine systems" or "safety-critical socio-technical systems". Based on the literature, dependability analysis, safety analysis, hazard analysis and HARA are the concepts related to risk assessment. We use these terms in addition to risk assessment. Finally, we consider "standard" or "technique" or "framework" or "method" in order to include safety standards and techniques providing modeling and analysis capabilities. Thus, the search string would be specified as follows:

- **Search string:** ("socio-technical systems" OR "human-machine systems" OR "safety-critical socio-technical systems") AND ("risk assessment" OR "dependability analysis" OR "safety analysis" OR "hazard analysis" OR "HARA") AND ("standard" OR "technique" OR "framework" OR "method")

Study Selection Criteria: We select the following four databases:

- Science Direct
- Web of Science
- IEEE
- Scopus

Our selection is based on: 1) the database evaluation, which has been reported in [25] and 2) the usage of the evaluation results in systematic literature reviews best practices [26]. In addition, we choose to discard Google Scholar because, based on the evaluation reported in [25], it does not support many aspects required for systematic searches (It fails to deliver replicable results during certain periods. It does not support for boolean search functionality. Its search precision has found to be significantly lower than 1% for systematic searches. Its coverage and recall is not adequate to use it as principal search system in systematic searches.).

We do not limit the search time-frame to have access to all results digitally available related to the topic and to provide the evolution of it over time. We define the inclusion and exclusion criteria as it is shown in Table II.

Study Selection Procedure: In the study selection procedure, we apply the search string to the databases and we identify the results. Then, we filter the results by title screening and we remove duplicated papers, book chapters and related works (related works are considered in the related work section). After that, we remove improper studies by reading the abstracts, considering inclusion and exclusion criteria and preparing a mind map for categorizing the papers to identify the publications relevant to the focus of our SLR. Finally, a preliminary list of primary studies are prepared, which should be checked in the quality assessment phase. In addition, inclusion and exclusion criteria are considered while reading the papers completely. The search and selection procedure is done by first author and quality control is done by the second author. The data extracted from the primary studies is based on the data extraction criteria shown in Table III.

Type	Description
Inclusion1	The primary study is about risk assessment/safety analysis/hazard analysis of safety-critical socio-technical systems.
Inclusion2	The primary study is peer-reviewed article written in English related to risk assessment of safety-critical socio-technical systems.
Inclusion3	The primary study provides contribution in development of conceptualization of risk assessment in safety-critical socio-technical systems.
Exclusion1	The primary study focuses on aspects of safety-critical socio-technical systems, but the aspects are different from risk assessment or safety analysis, e.g., process design, execution, or does not present sufficient details regarding risk assessment of safety-critical socio-technical systems.
Exclusion2	The primary study is about risk assessment of systems other than safety-critical socio-technical systems. A system containing only technical entities is an example.
Exclusion3	The text of the primary study is not accessible.
Exclusion4	The primary study is not clearly related to at least one aspect of the specified research questions.
Exclusion5	The primary study is secondary or tertiary study.
Exclusion6	The primary study belongs to commercial, pure opinions, grey literature with low or moderate credibility, books, tutorials, posters and papers that did not undergo a peer-review process.

TABLE II: Inclusion and exclusion criteria

Quality Assessment Criteria: For qualitative and quantitative assessment of the studies, we develop a checklist based on [27] and [28]. The checklist is shown in Table IV. As it is shown in this table, for each item there are three options that one of them should be selected based on the answer to the related question. For each paper sum of the scores of all answers should be accumulated. The accumulated scores of each paper are helpful to distinguish the studies with higher quality. The list of papers selected after this phase should be completely analyzed and final primary studies are selected from these papers by considering inclusion and exclusion criteria after reading the papers completely.

3.2 Conducting the SLR

In this subsection, we provide the details regarding how we conducted the SLR.

3.2.1 Data Collection:

We applied the SLR protocol described in subsection 3.1.3. In particular, we applied the search string to the four selected databases explained in the study selection criteria without limiting the dates of the publications. The search is performed between January 24 to February 11, 2022. We obtained 1752 results from which 1491, 46, 13 and 202 are obtained from Science Direct, Web of Service, IEEE, and Scopus respectively, as it is shown in Fig. 2. Then, we performed the title screening

Extracted Data	Used for
Study title, year, type of venue	Study overview
Interpretation/conceptualization of socio entities	RQ1.1 and RQ1.2
Interpretation/conceptualization of technical entities	RQ1.3
Interpretation/conceptualization of socio and technical orchestration	RQ1.4
Interpretation/conceptualization of effects of organizational/technological changes	RQ1.5 and RQ1.6
Interpretation/conceptualization of augmented reality	RQ1.7
Interpretation/conceptualization of risk	RQ1.8
Provided steps of risk assessment	RQ1.9
Provided safety perspective	RQ1.10
Modeling formality	RQ2.1
Contribution context	RQ2.2
Analysis technique for analyzing system behavior	RQ2.3
Tool support	RQ2.4
Application domain	RQ3.1
Supported standards	RQ3.2
Presented illustrative scenarios	RQ3.3
Challenges	RQ4

TABLE III: Data extraction criteria

and we removed the papers which were duplicated, book chapters and related work papers. Related work papers are analyzed in the related work section. In the title screening, we considered inclusion and exclusion criteria, which are defined in Subsection 3.1.3. After this step we gained 352 results.

It was not straightforward to include/exclude papers based on their abstract and in order to be able to identify primary studies in relation to the focus of our SLR, we needed to reach an enhanced understanding of the papers. Thus, while we performed the abstract screening, we prepared a mind map to group the papers into categories/subcategories. Then, we identified the relevant studies based on the categorization and the inclusion/exclusion criteria.

We defined five main categories for the papers which are shown in Fig. 3. The first category includes papers which propose a method/framework/technique/model/approach for risk assessment or for contributing to risk assessment of safety-critical socio-technical systems, shown by C1. The second category includes papers which apply one or more methods of risk assessment or contributing to risk assessment of socio-technical systems, shown by C2. The third category includes papers surveying, comparing, evaluating or discussing some methods or viewpoints of risk assessment or contributing to risk assessment of socio-technical systems, shown by C3. The fourth category includes papers providing challenges of using specific methods of risk assessment or contributing to risk assessment of socio-technical systems or challenges of risk

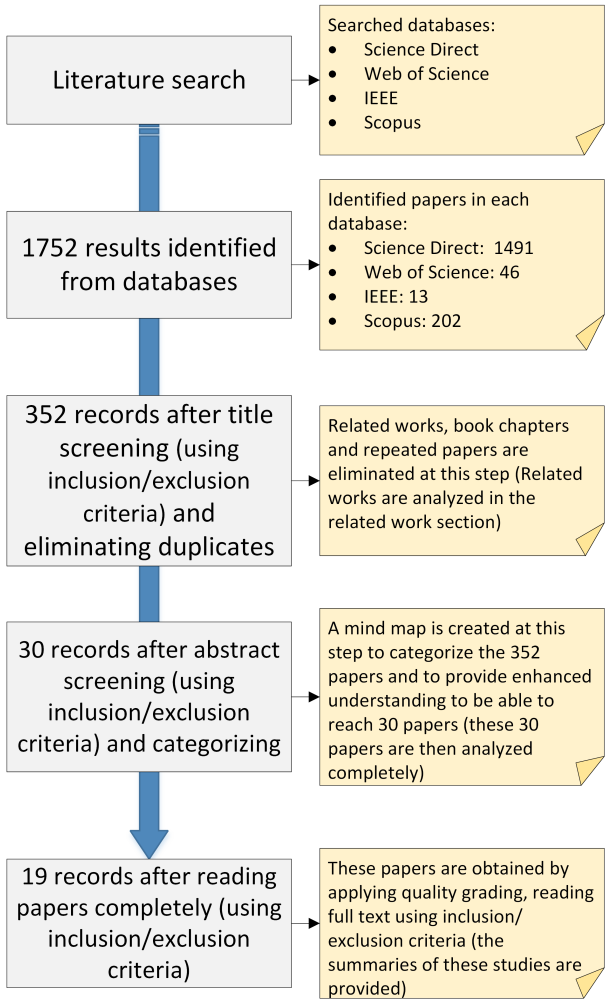

Fig. 2: Process for papers selection

assessment in specific applications, shown by C4. Finally, the last category includes the papers on developing tool for a method for risk assessment or for contributing to risk assessment of socio-technical systems, shown by C5.

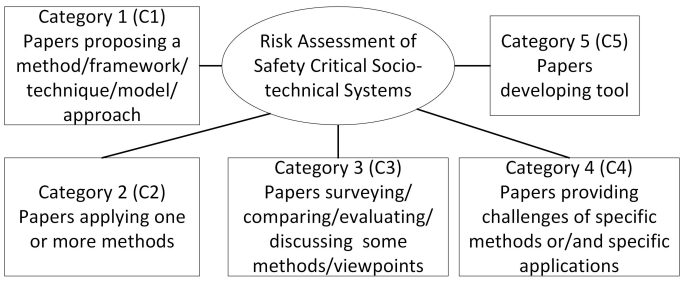

Fig. 3: Proposed high-level categorization for the identified papers

Most of the studies are assigned to C1 (as we expect, because we did a title screening before this step). We divided this category into four subcategories shown in Fig. 4.

Assessment Criteria	Score	Description
QA1: Does the study include a clear statement of the goal?	0	No, the goal is not described.
	0.5	Partially. The goal is described but it is not clear.
	1	Yes, the goal is described well and clear.
QA2: Is there clear statement of findings?	0	No, findings are not discussed.
	0.5	Partially. Findings are discussed, but not completely and clearly.
	1	Yes, the findings are well discussed.
QA3: Is there an adequate description of the context in which the research was carried out?	0	No, context of research is not described.
	0.5	Partially. Context of research is described partially.
	1	Yes, context of research is described well.
QA4: Does the study provides improvement towards risk assessment of safety-critical socio-technical systems?	0	No, no improvement is provided.
	0.5	Partially. The study provides improvements, but it is partially towards risk assessment of safety-critical socio-technical systems.
	1	Yes, improvement towards risk assessment of safety-critical socio-technical systems is provided.
QA5: Are the results in accordance with the goal of the study?	0	No, the results are not in accordance with the goal of the study.
	0.5	Partially. The results are partially in accordance with the goal of the study.
	1	Yes, the results are in accordance with the goal of the study.
QA6: Is the research process documented adequately?	0	No, the research process is not documented.
	0.5	Partially. The research process is documented but not adequately.
	1	Yes, the research process is documented adequately.
QA7: Are the assumptions and limitations explained well?	0	No, assumptions and limitations are not explained.
	0.5	Partially. Assumptions and limitations are explained but not clearly and completely.
	1	Yes, assumptions and limitations are explained well.
QA8: Is the link between data, interpretation and conclusions clearly shown?	0	No, there is no link between data, interpretation and conclusions.
	0.5	Partially. There is link between data, interpretation and conclusion (or partly), but it is not shown clearly.
	1	Yes, the link between data, interpretation and conclusion is shown clearly.

TABLE IV: Quality assessment criteria

The first subcategory includes papers incorporating STAMP (Systems-Theoretic Accident Model and Processes)[10] or STPA (Systems-Theoretic Process Analysis) method [11]. The second subcategory includes papers incorporating FRAM (Functional Resonance Analysis Method) [29] or safety II [30]. The third subcategory includes papers incorporating Probabilistic Risk Assessment (PRA) or Bayesian Networks (BNs). Finally, the forth subcategory includes papers not incorporating any of the mentioned techniques.

Category C1-4 is also divided to five subcategories which are shown in Fig. 4. In the following paragraphs, we explain about STAMP, STPA, FRAM, Safety II, PRA and BNs briefly and we provide an example.

STAMP [10] is an accident model proposed to capture dynamic complexity and non-linear interactions leading to accidents. Based on this model, STPA hazard analysis technique [11] is proposed. In this technique a set of scenarios leading to hazards due to unsafe and unintended interactions among system components is created. More specifically, in this technique hazards are identified and based on the hazards system safety constraints and control structure are defined. Control structure contains system components and paths of control and feedback. In order to provide the analysis, contribution of each control action to hazards is assessed.

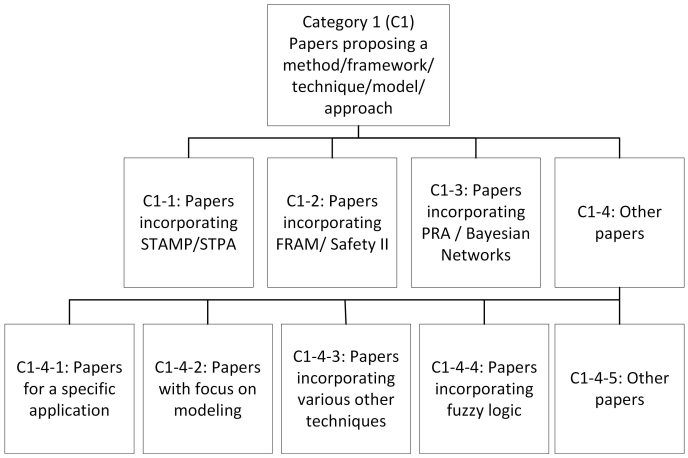

Fig. 4: Subcategories of the Category 1

FRAM [29] is an analysis method proposed to model the functions that are required to succeed. Based on this method, system functions should be identified and described. Potential variability and possible actual variabilities of the functions should be characterized in one or more instances of the model.

Functional resonance should be defined based on dependencies among functions and based on potential for functional variability. Finally, ways to monitor the development of resonance should be identified. Based on Safety II perspective [30], the purpose is to increase the number of acceptable outcomes as high as possible under varying conditions.

PRA-based techniques are techniques using probability for assessing risk. BNs are probabilistic graphical models for representing uncertain knowledge using nodes and edges for modeling random variables and conditional probabilities of the corresponding random variable.

As an example of papers assigned to the first subcategory, in [31], a methodology with the name RiskSOAP is proposed for risk situational awareness provision in road tunnel safety. STPA is used for selecting the elements and their characteristics in the system design specifications. This methodology represents the tunnel status in terms of its self-awareness about its vulnerabilities and threats and it supports designers and engineers to enhance the system based on the risk situational awareness. RiskSOAP is applied to a specific road tunnel in Greece to test the soundness and applicability of the methodology.

Using the paper categorization and by considering inclusion/exclusion criteria, we obtained 30 results to be analyzed completely. We read the full text of 30 papers by considering inclusion/exclusion criteria and we applied the quality criteria. We selected papers with at least 7 score in the quality criteria. As a result 19 papers were selected. The quality grading for the selected papers is presented in Table V using the quality assessment criteria defined in Table IV.

ID	QA1	QA2	QA3	QA4	QA5	QA6	QA7	QA8	Score
[32]	1	1	1	1	1	0.5	0.5	1	7
[33]	1	1	1	1	1	1	1	1	8
[34]	1	1	0.5	1	1	1	0.5	1	7
[35]	1	1	0.5	1	1	1	0.5	1	7
[36]	1	1	1	1	1	1	1	1	8
[37]	1	1	1	1	1	1	1	1	8
[38]	1	1	1	1	1	1	1	1	8
[39]	1	1	1	1	1	1	1	1	8
[40]	1	1	1	1	1	1	1	1	8
[41]	1	1	1	1	1	1	1	1	8
[42]	1	1	1	1	1	1	1	1	8
[43]	1	1	1	1	1	1	0.5	1	7.5
[44]	1	1	1	1	1	1	1	1	8
[45]	1	1	1	1	1	1	0.5	1	7.5
[46]	1	1	1	1	1	1	1	1	8
[47]	1	1	1	1	1	1	1	1	8
[48]	1	1	1	1	1	1	1	1	8
[49]	1	1	1	1	1	1	1	1	8
[50]	1	1	1	1	1	1	1	1	8

TABLE V: Quality grading of the primary studies using Table IV

3.2.2 Data Extraction:

The study overview of the identified 19 primary studies are presented in Table VI. We used Excel spreadsheets for analyzing the identified papers. In the next section, we explain about the results of our SLR in relation to the defined research questions.

4 Results and Analysis

In this section, we present and discuss the results and the analysis of the primary studies. The summary of each of the selected primary studies is conceived in a structured manner and contains the essential information in relation to the research questions. Finally, we provide tables summarizing the findings related to research questions shown in Tables VII - XI.

In [32], the author proposes a methodological framework called Human Error Risk Management for Engineering Systems (HERMES). The framework contains a roadmap (for human factor approaches and methods for specific problems) and a body of possible techniques to deal with essential issues of modern human risk assessment. The first step is to choose a theoretical platform for both retrospective (backward looking) and prospective (forward looking) analysis. In order to do that models for human behavior, systems and for HMI (Human Machine Interface) should be defined. Typical data and parameters of the system are derived by evaluating the socio-technical context using ethnographic study (empirical methods such as simulators, interviews, questionnaires, etc.) and cognitive task analysis (theoretical evaluation of work processes). In retrospective analysis, past events are investigated to identify causes of accidents. The analysis results provide additional insights to be used for prospective study. For a complete prospective study the unwanted consequences and hazards can be evaluated by applying a quantitative risk assessment technique. This framework offers Reference Model of Cognition (RMC) as a human behavior model containing four cognitive functions: *Perception, Interpretation, Planning* and *Execution* (PIPE) and two cognitive processes: *Memory/Knowledge Base* and *Allocation of Resources*. There is also a taxonomy of human erroneous behaviors in relation to the model, which can be used in the framework. The framework also offers Dynamic Logical Analytical Method (DYLAM) method, which enables the evaluation of time dependent behavior of human machine systems. The framework proposed in this paper provides **potential for capturing human, organizational and technological aspects** using the human behavior model and the related taxonomy. In addition, the framework provides the potential for capturing **socio-technical orchestration** by defining the correlation between human and machines. However there are some discussions about the critical issues due to automation, it does not provide means for characterizing **organizational changes effects**, **technological changes effects** and **AR effects** explicitly. In addition, it has the **potential to capture risk** since it is possible to discuss the consequences using the proposed framework. Regarding the support for the risk assessment process, it emerges that this paper supports **risk identification** step (the framework can be used to find, recognize and describe the causal factors and risks). In addition, **risk**

ID	Title	Year	Type
[32]	Human error risk management for engineering systems: a methodology for design, safety assessment, accident investigation and training	2004	Journal
[33]	Human and organisational factors in the operational phase of safety instrumented systems: A new approach	2010	Journal
[34]	Modelling and analysis of socio-technical system of systems	2010	Conference
[35]	MMOSA–a new approach of the human and organizational factor analysis in PSA	2014	Journal
[36]	Modeling a global software development project as a complex socio-technical system to facilitate risk management and improve the project structure	2015	Conference
[37]	Usability of accident and incident reports for evidence-based risk modeling–A case study on ship grounding report	2015	Journal
[38]	Accident modelling of railway safety occurrences: the safety and failure event network (SAFE-Net) method	2015	Journal
[39]	A new framework to model and analyze organizational aspect of safety control structure	2017	Journal
[40]	Incorporating epistemic uncertainty into the safety assurance of socio-technical systems	2017	Journal
[41]	An Accident Causation Analysis and Taxonomy (ACAT) model of complex industrial system from both system safety and control theory perspectives	2017	Journal
[42]	A new organization-oriented technique of human error analysis in digital NPPs: Model and classification framework	2018	Journal
[43]	A hybrid model for human factor analysis in process accidents: FBN-HFACS	2019	Journal
[44]	Functional modeling in safety by means of foundational ontologies	2019	Journal
[45]	Developing a method to improve safety management systems based on accident investigations: The SAfety FRactal ANalysis	2019	Journal
[46]	The development history of accident causation models in the past 100 years: 24Model, a more modern accident causation model	2020	Journal
[47]	Ontology-based computer aid for the automation of HAZOP studies	2020	Journal
[48]	Human functions in safety-developing a framework of goals, human functions and safety relevant activities for railway socio-technical systems	2021	Journal
[49]	A case study for risk assessment in AR-equipped socio-technical systems	2021	Journal
[50]	Model-based safety engineering for autonomous train map	2022	Journal

TABLE VI: Selected primary studies

analysis is supported (using the framework, the results can be provided by considering risk sources, consequences, scenarios and likelihood). Finally, **risk evaluation** is also supported (there can be discussions on the results and required actions to support decisions). Since the focus is on chain of events and the root causes of accidents, it emerges that this framework can be labelled as *safety engineering today* perspective. There are no syntax and semantics proposed and used and a proper modeling language does not emerge, only an **informal notation** is used and **no extending formality contribution** is present (it offers different models and techniques that can be used for the aim of this framework and it does not propose contribution for developing concepts, semantics and syntax for modeling/analyzing system entities). Discussions on causes of accidents using the proposed framework can be described as **qualitative, quantitative and linear** analysis. The perspective in this framework is both **backward and forward looking**, since it provides predictions and possibility of modeling and analyzing accidents which would happen in the future based on the accidents which have happened in the past. This paper does not provide **structured analysis process** and **tool support**. It is proposed for **general application** and **scenarios** from two domains (nuclear power plant and railway) are discussed. It is mentioned in different phases of the framework that standards should be considered (for example it is mentioned that for defining safety measures conformance with safety standards is required), meanwhile it is not discussed if the framework provides **support for standards**. The **challenge** mentioned in this study is lack of readily available data to be used by human factor approaches that can be used in the framework.

In [33], the authors propose an approach for addressing human and organizational factors in the operational phase of safety instrumented systems. A list of eight safety influencing factors are considered based on the literature with slight reformulation. These influencing factors are: *maintenance management, procedures, error-enforcing conditions, housekeeping, goal compatibility, communication, organization* and *training*. The proposed approach contains five main steps. The first step is estimation of proportion of design safety integrity level (SIL) using the system design and based on expert judgment or previous experiences. The second step is determining the weights of influencing factors and calculating the normalized weight factors. The third step is rating the influencing factors. The fourth step is calculating the operational SIL. If the operational SIL is not acceptable, then a fifth step is also considered for taking preventive or corrective

actions to improve safety. The approach proposed in this paper provides **potential for capturing human, organizational, technological aspects** and **socio-technical orchestration** by using the safety influencing factors and their relationships. However, it does not provide means for characterizing **organizational changes effects**, **technological changes effects** and **AR effects** explicitly. In addition, it provides **structured conceptualization to capture risk** since it is possible to discuss the consequences by determining SIL using defined formula based on other defined parameters (such as ratings, weights, etc.). Thus, there is well-formedness rules to relate the proposed concepts in determining SIL which is highly in connection with risk. Regarding the support for the risk assessment process, it emerges that this paper supports **risk identification** step (the approach can be used to find, recognize and describe the causal factors and risks). In addition, **risk analysis** is supported (using the approach, the results can be provided by considering risk sources, consequences and scenarios). Finally, **risk evaluation** is also supported (there can be discussions on the results, SIL and required actions to support decisions). Since the focus is on chain of events and the root causes of accidents, it emerges that this model can be labelled as *safety engineering today* perspective. There are no syntax and semantics proposed and used and a proper modeling language does not emerge, only an **informal notation** is present and a set of **concepts** are proposed that can be used for modeling safety influencing factors and determining SIL. Discussions on causes of accidents using the proposed method can be described as **qualitative, quantitative** (because of SIL calculation) and **linear** analysis. The perspective in this method is both **backward looking** (since parameters can be determined based on previous experiences) and **forward looking** (since it provides predictions that can be used to improve safety for preventing future accidents). This paper provides **structured analysis process** using SIL calculation which is based on proportion of design SIL, weights and rates of influencing factors. However, it does not provide **tool support**. Although the approach is proposed for **specific application** (process industry) and a **scenario** from this domain is discussed as illustrative case study, there is potential to use it for other domains by some modifications. Since the process for improving safety is based on standards IEC 61508 and IEC 61511 and the proposed approach determines SIL, we conclude that the approach provides **support for standards (IEC 61508 and IEC 61511)**. The mentioned **challenges** of this study are 1) determining rates in a way to allow certain influence of a factor (in the case study, rates of all factors are considered equal), 2) difficulty in determining proportion of design SIL and weights of the factors, 3) requiring further research for providing validation, 4) providing some more applications, 5) ensuring consistency over time in the ratings, 6) including effects of system modifications and aging of equipment, 7) incorporating other safety influencing factors.

In [34], authors propose an approach for modeling socio-technical system of systems to help end users identify and analyze the hazards and associated risks. This approach provides notations for representing a system with focus on the defined concepts: *capabilities*, *dependencies* and *vulnerability* in the context of risk management. Then hazards are identified and discussed. This approach proposes the **potential to capture socio and technical aspects and their orchestration** by using the proposed concepts. In addition, the approach proposes the **potential to capture risk** by using the discussions on hazards, probability, severity and consequences (which are provided as **qualitative and linear analysis**, with a **forward looking** perspective). Regarding the support for the risk assessment process, it emerges that this approach supports **risk identification** step (the approach can be used to find, recognize and describe the causal factors and risks) by using the proposed concepts. In addition, **risk analysis** is supported (discussions are provided for considering risk sources, consequences, scenarios and likelihood). Finally, **risk evaluation** is also supported (there are discussions on the results and required actions to support decisions). There are no syntax and semantics proposed and used and a proper modeling language does not emerge, only an **informal notation** is present and the contribution of the framework is in developing **concepts**. There is no **structured analysis process** and the study does not provide **tool support** for providing the analysis results. Although the paper uses a case study from information technology domain, the proposed approach is not specific to a domain (it is proposed for **general application**). There is no discussion for **supported standards** and the mentioned **challenges** of this approach are 1) requiring tools for evaluating quantitative analysis, 2) requiring exploration to mesh with existing safety/dependability assurance processes.

In [35], the authors propose a method called MMOSA (Man-Machine-Organization System Approach) in order to incorporate human and organizational factors in probabilistic safety assessment (PSA). It uses human reliability analysis (HRA) methods such as THERP and SPAR-H and the novelty of the method is considering machine-organization interfaces in human performance evaluation. The method is based on MMOS concepts containing man/machine/organization characteristics and their interfaces. For example, concepts of man-organization interfaces are, *complexity of the action, work environment, procedure, time, communication* and *training*. The proposed method provides an estimation of human error probabilities using basic human error probabilities (BHEP) from HUFAD_E (Human Factor Analysis Database_English) database presented in the paper. The proposed method in this paper provides **potential for capturing human, organizational, technological aspects** and **socio-technical orchestration** by using the MMOS concepts. However, it does not provide means for characterizing **organizational changes effects**, **technological changes effects** and **AR effects** explicitly. In addition, it provides **potential for capturing risk** since it is possible to discuss the consequences and to determine human error probabilities. Regarding the support for the risk assessment process, it emerges that this paper supports **risk identification** step (the approach can be used to find, recognize and describe the causal factors and risks). In addition, **risk**

analysis is supported (using the approach, the results can be provided by considering risk sources, consequences and scenarios). Finally, **risk evaluation** is also supported (there can be discussions on the results, human error probabilities and required actions to support decisions). Since the focus is on chain of events and the root causes of accidents, it emerges that this model can be labelled as *safety engineering today* perspective. There are no syntax and semantics proposed and used and a proper modeling language does not emerge, only an **informal notation** is present and there is **no extending formality contribution** in the paper, instead the contribution is in integrating MMOS concepts in human factors analysis process for modeling and analyzing man-machine-organization factors and their interfaces. Discussions on causes of accidents using the proposed method can be described as **qualitative, quantitative** (because of the probabilities calculations) and **linear** analysis. The perspective in this method is mostly **forward looking** (since it provides prediction about human errors). This paper provides **structured analysis process** using HEP calculations and it also provides **tool support** using MMOS software in Microsoft Visual Basic 6.0 environment. The proposed method can be used for **general application**. However, the focus is mostly on nuclear domain and a **scenario** from this domain is discussed as a case study. In this paper, it is not discussed if the proposed model provides **support for standards** and the mentioned **challenge** of this study is requiring further research for understanding the influence of human and organizational factors on safe operations.

In [36], authors propose a technique for modeling global software development project as a complex socio-technical system to facilitate risk management. This study considers risks caused by geographical, cultural and time distances between the developers in the project and proposes structured conceptualization for socio-technical systems using three main concepts: *functional components*, *output-input arrows* representing the links between the components and *feedback connections* for correcting misinterpretations between components. In addition, socio aspects are considered in the modeling and it contains well-formedness rules to relate the concepts. Using the proposed structured concepts, it proposes **structured conceptualization for capturing human, organizational, technical aspects** and **socio-technical orchestration**. In addition, it proposes concepts for characterizing **organizational changes effects** such as global distances. However, it does not propose concepts for characterizing **technological changes effects** and **AR effects**. It has the **potential to capture risk** since it is possible to discuss the consequences using the proposed modeling. Regarding the support for the risk assessment process, it emerges that this modeling technique supports **risk identification** step by identifying the risk using discussions about causal factors. In addition, **risk analysis** is supported (discussions are provided for considering risk sources, consequences, scenarios and controls). Finally, **risk evaluation** is also supported (there are discussions on the results and required actions to support decisions). Since the modeling technique is non-linear and contains feedback controller, it can be labelled as *safety III* perspective. There are no syntax and semantics proposed and used and a proper modeling language does not emerge, only an **informal notation** is present and the contribution of the paper is a set of **concepts**. Discussions on causes of accidents are described as **qualitative and non-linear analysis** (because there is feedback controller) with a **forward looking** perspective and there is no **structured analysis process**. The proposed framework does not provide **tool support** and it is proposed for **specific domain** (global software development project). However it uses a **scenario** from ICT (Information and Communications Technology), the proposed approach has the potential to be used in other domains. There is no explicit discussion about **support for standards**. The **challenges** in this study are 1) lack of measures to mitigate the risks, 2) not using information from reality such as interviews or analysis of information flows in the development of the methodology.

In [37], authors propose another version of Human Factors Analysis and Classification System (HFACS) and review accident reports based on the new taxonomy. The extended HFACS is called HFACS-Ground by adding factors more related to ship grounding accidents. For example, *infrastructure* is added as a latent failure to cover waterway complexity related issues. This extended taxonomy has five levels: *unsafe acts, preconditions, supervisional influence, organizational influence* and *outside factors*. For each level there are two or three layers. In addition, high level positive functions called Safety Factor (SF) are used for reviewing incident reports. The first reason for using SFs is that incident reports are not as structured as accident reports and it is not practical to use taxonomies such as HFACS (normally only active failures are reported in incident reports, which would be misleading). The second reason is because of the difference of accidents and incidents (incidents are near-miss and they do not result in serious consequences on human life or the environment like accidents). Thus, in incidents it is desirable to detect positive functions which acted as barriers and stopped the incident to become an accident. However, these positive functions are then negated to be used for analyzing the contributing factors to incidents. Pearson correlation coefficient (r) is used to show the statistical dependencies of the factors two-by-two and the significance of the correlations is shown by p-values. The results show the frequency of different levels of failures in the accident reports and if there is weak or strong correlation between different factors. It also discusses that the incident reports are not reliable in their current non-systematic format to be used for evidence-based risk modeling and they can be used as alerts for possible hazards. The extended HFACS taxonomy proposed in this paper provides **potential for capturing human, organizational and technological aspects**. In addition, the proposed taxonomy provides the potential for capturing **socio-technical orchestration** using the relation between human and organization and technological factors. It does not propose means for characterizing **organizational changes effects**, **technological changes effects** and **AR effects** explicitly. In addition, it has the **potential to**

capture risk since it is possible to discuss the consequences using the proposed taxonomy. Regarding the support for the risk assessment process, it emerges that this paper supports **risk identification** step since it uses the taxonomy to identify the causal factors. In addition, **risk analysis** is supported (the results can be provided for considering risk sources, consequences, scenarios, likelihood and controls). Finally, **risk evaluation** is also supported (there can be discussions on the results and required actions to support decisions). Since the focus is on chain of events and the root causes of accidents, it emerges that this model can be labelled as *safety engineering today* perspective. There are no syntax and semantics proposed and used and a proper modeling language does not emerge, only an **informal notation** is present and the contributions of the extension consist of a set of **concepts** for factors related to grounding accidents. Discussions on causes of accidents using the proposed model can be described as **qualitative, quantitative** (considering frequencies and correlations) and **linear** analysis (considering chain of events). The perspective in this model is **backward looking**, since it is modeling and analyzing accidents and incidents that have happened in the past. This paper does not provide **structured analysis process** and **tool support**. It is proposed for a **specific domain** (ship grounding) and **scenarios** from this domain are presented. Since the proposed taxonomy is specified version of HFACS to be used for ship grounding, there is no potential to use the extended version for other domains. There are no discussions about **support for standards** and the mentioned **challenges** in this study are 1) use of limited reports from specific databases, 2) subjectivity in the reports.

In [38], authors propose a model called Safety and Failure Event Network (SAFE-Net) to model the contributing factors of railway safety occurrences. This paper uses Contributing Factors Framework (CFF) for collecting data on contributing factors to railway safety occurrences by using reports submitted to rail safety regular in Queensland for five years (2006-2010). The contributing factors in this framework are categorized to three main groups: *individual/team factors, technical failures* and *local conditions/organizational factors*. 429 safety occurrences are analyzed and contributing factors in each of them are identified. SAFE-Net model is used to model the connections between different contributing factors. In this model all factors that have been attending the same safety occurrence before, are identified and the relations between the factors are listed. Then this information can be entered to a developed human factor tool named SNA (Social Network Analysis) program to calculate centrality (showing factors' importance) measures for each factor and to show the models. The models are networks containing contributing factors as nodes and their relations as links between the nodes. Centrality is also shown by a circle around each factor and the size of the circle shows the extent of the centrality. The model proposed in this paper provides **potential for capturing socio (human and organizational) and technological aspects**. In addition, the proposed model provides the potential for capturing **socio-technical orchestration** using the relation between human and organization and technological factors. It does not propose means for characterizing **organizational changes effects**, **technological changes effects** and **AR effects** explicitly. In addition, it has the **potential to capture risk** since it is possible to discuss the consequences using the proposed model. Regarding the support for the risk assessment process, it emerges that this paper supports **risk identification** step by using the reports and CFF framework. In addition, **risk analysis** is supported (discussions can be provided for considering risk sources, consequences and scenarios). Finally, **risk evaluation** is also supported (there can be discussions on the results and required actions to support decisions). Since the paper discusses about FRAM technique and establishes the work on the new generation of thinking proposed in FRAM, we can label it as *safety II* perspective. In this model an **informal notation** is provided and the contributions of the paper consist of a set of **concepts** for modeling different causal factors and the connections between them based on the previous accident reports. Discussions on causes of accidents using the proposed model can be described as **qualitative, quantitative** (using the amount of centrality) and **non-linear** analysis (the structure of the model is networked). The perspective in this model is **backward looking**, since it is based on the accident reports and it focuses on the accidents that have happened in the past. There is no **structured analysis process** and the proposed model provides **tool support** using SNA program. It is proposed for a **specific domain** (railway) and a **scenario** from this domain is presented. However, there is the potential to use it for other domains. There is no explicit discussion about **support for standards** and the mentioned **challenge** in this study is no criteria for assessment of the significance of introducing this approach.

In [39], authors propose a framework to model and analyze organizational aspects of hierarchical safety control structures. This framework, introduces a specific organizational feedback control loop with a customized process model for adjusting STPA for deficiency analysis of organizational safety control structure. Using the new proposed control structure, hazardous behaviors caused by organizational mechanisms dysfunctionality can be detected. The framework has the **potential for capturing human and organizational aspects** and it has the **potential to capture technical aspects** and **socio-technical orchestration**. It does not propose concepts for characterizing **organizational changes effects**, **technological changes effects** and **AR effects** explicitly. In addition, it has the **potential to capture risk** since it is possible to discuss the consequences using the proposed framework. Regarding the support for the risk assessment process, it emerges that this framework supports **risk identification** step by identifying the risk using discussions about causal factors. In addition, **risk analysis** is supported (discussions are provided for considering risk sources, consequences, scenarios and controls). Finally, **risk evaluation** is also supported (there are discussions on the results and required actions to support decisions). Since the framework is an extension for STPA, it can be labelled as *safety III* perspective. There are no syntax and semantics

proposed and used and a proper modeling language does not emerge, only an **informal notation** is present and the contributions of the model consist of a set of **concepts** for modeling and analyzing organizational aspects of hierarchical safety control structures. Discussions on causes of accidents are described as **qualitative and non-linear analysis** with a **forward looking** perspective and there is no **structured analysis process**. The proposed framework does not provide **tool support**. The proposed approach is for a **general domain** and it uses a **scenario** from aviation maintenance industry. There is no explicit discussion about **support for standards**. The mentioned **challenges** in this study are 1) lack of quantitative analysis, 2) limited scope of case study, 3) lack of assessment of practicality and validity of the framework in macro level, 4) lack of comparison with other widespread methods other than STPA which is done.

In [40], authors propose a model to systematically capture and track known uncertainties. It also proposes a process for integrating the model in the current hazard analysis techniques such as STPA. The proposed model is based on a created reference with a wide range of safety-critical causal relationships from the literature. The reference is a suggested checklist as a guide and direction for possible causal paths that may result in unsafe situation and it is created by conducting an extensive literature review. The reference contains six primary causal factors: *Human, Organization, Technology, Process, Information* and *Environment* (HOT-PIE). Each of them may contain two or three sub categories. The reference is then used for creating the multi-level causal relationship model. Multi-level modeling is used to model both relation between factors and relation between causal factors. It considers that a causal factor may influence another causal factor and it can be modeled using multi-level causal relationship model. Finally, a process is proposed to show how the reference and the model can be used in hazard analysis techniques. The reference and the model proposed in this paper includes concepts for characterizing human, organization and technology and related aspects. Thus, it provides the **potential for capturing human, organizational and technological aspects**. In addition, the proposed model provides the potential for capturing **socio-technical orchestration** using the relation between concepts for socio aspects and concepts for technological aspects. It does not propose concepts for characterizing **organizational changes effects**, **technological changes effects** and **AR effects** explicitly. In addition, it has the **potential to capture risk** since it is possible to analyze the consequences using the proposed reference and model. Regarding the support for the risk assessment process, it emerges that this paper supports **risk identification** step (the model can be used to find, recognize and describe the causal factors and risks). In addition, **risk analysis** is supported (discussions can be provided for considering risk sources, consequences, scenarios and uncertainties). Finally, **risk evaluation** is also supported (there can be discussions on the results to support decisions). Since the paper proposes a process for integrating the reference and the model in the STPA technique, we can label it as *safety III* perspective. There are concepts proposed and used and an **informal notation** is provided and the contribution of the paper consist of a set of **concepts** for modeling causal factors. Discussions on causes of accidents using the proposed conceptualization can be described as **qualitative and non-linear analysis** with a **forward looking** perspective and there is no **structured analysis process**. The proposed reference and model does not provide **tool support** and it is proposed for **general domain**. However, some **scenarios** from ministry of defence are presented. There is explicit discussion about **support for standards** and it is shown that it can support SAE ARP-4761 (an industrial standard for conducting safety assessment process to certify civil aircraft) by the proposed causal paths that are essential in the analysis. The mentioned **challenges** in this study are 1) requiring further study for applicability in larger systems, 2) requiring further study for automating the process, 3) no criteria for assessment of the significance of introducing this approach to existing hazard analysis.

In [41], the authors propose an Accident Causation Analysis and Taxonomy (ACAT) model to provide a comprehensive understanding of accidents and causes statistics. Using this model complex systems can be decomposed based on six factors: *machine, man, management, information, resources* and *environment*. In addition, four control functional abstractions are considered: *actuator, sensor, controller* and *communication*. The combinations of system factors and control functions as a matrix form the proposed model. Using the model the accident causes can be identified and classified. In addition, by calculating the proportions of different types of causes their percentages can be obtained. The proposed model in this paper provides **potential for capturing human, organizational,** and **technological aspects** by using the system factors and provides **potential for capturing socio-technical orchestration** by using the control functions (specially the communication function). However, it does not provide means for characterizing **organizational changes effects**, **technological changes effects** and **AR effects** explicitly. In addition, it provides **potential for capturing risk** since it is possible to discuss the consequences and to determine percentage of different causal factors. Regarding the support for the risk assessment process, it emerges that this paper supports **risk identification** step (the approach can be used to find, recognize and describe the causal factors and risks). In addition, **risk analysis** is supported (using the approach, the results can be provided by considering risk sources, consequences and scenarios). Finally, **risk evaluation** is also supported (there can be discussions on the results, proportions and required actions to support decisions). Since the focus is on chain of events and the root causes of accidents, it emerges that this model can be labelled as *safety engineering today* perspective. There are no syntax and semantics proposed and used and a proper modeling language does not emerge, only an **informal notation** is present and a set of **concepts** are proposed that can be used for modeling causal factors, their proportions and control functions. Discussions on causes of accidents

using the proposed method can be described as **qualitative, quantitative** (because of the percentages calculations) and **linear** analysis. The perspective in this method is mostly **backward looking** (since the focus is on the previous accidents). However, the final aim is to improve the system for preventing future accidents. This paper does not provide **structured analysis process** and **tool support**. The model is proposed for **general application** and **scenarios** from BP Texas refinery case are discussed as the case study. In this paper, it is not discussed if the proposed model provides **support for standards** and the mentioned **challenge** of this study is requiring further research for providing details of the proposed broad concepts.

In [42], authors propose an organization-oriented conceptual model of human error analysis (HEA) in digital Nuclear Power Plants (NPPs). In addition, the classification framework of HEA is developed based on the conceptual model. The proposed model and framework consider new challenges because of the digital technology and its effects on human error and human reliability. The proposed model contains four modules/levels: Performance Shaping Factors (PSFs) (levels of organizational factors, situational factors, error-triggering individual factors), Psychological Error Mechanisms (PEMs), error recovery and human errors. The model shows that performance shaping factors influence on human error and human error influences on error recovery. Safety barrier is also considered as a barrier to prevent human error and to prevent an accident. The classification framework contains classification for human error, organizational factors, situational factors, individual factors, PEMs, Error Recovery Failures (ERFs) and safety barriers. The model and classification framework proposed in this paper provide **potential for capturing human, organizational and technological aspects**. In addition, the proposed taxonomy provides the potential for capturing **socio-technical orchestration** using the relation between socio and technological factors. Furthermore, it provides the potential for capturing **organizational changes effects** and **technological changes effects** by considering digitalization, new computer-based information displays, digital procedures and etc. However, It does not propose means for characterizing **AR effects** explicitly. In addition, it has the **potential to capture risk** since it is possible to discuss the consequences using the proposed model. Regarding the support for the risk assessment process, it emerges that this paper supports **risk identification** step since it uses the model and classification to identify the causal factors. In addition, **risk analysis** is supported (the results can be provided for considering risk sources and consequences). Finally, **risk evaluation** is also supported (there can be discussions on the results and required actions to support decisions). Since the focus is on the chain of events and the root causes of accidents, it emerges that this model can be labelled as *safety engineering today* perspective. There are no syntax and semantics proposed and used and a proper modeling language does not emerge, only an **informal notation** is present. The contributions of the model and categorization consist of a set of **concepts** for modeling human error, organizational factors, situational factors, individual factors, PEMs, ERFs and safety barriers. Discussions on causes of accidents using the proposed model can be described as **qualitative and linear** analysis (considering chain of events). The perspective in this model is **forward looking**, since it provides predictions and it models and analyzes possible accidents which would happen in the future. This paper does not provide **structured analysis process** and **tool support**. It is proposed for a **specific domain** (nuclear power plant). However, **scenarios** from this domain are not presented and are considered as future work. Although the model and categorization are proposed for nuclear power plant, there is potential to use it for other domains by some or little revision. There are no discussions about **support for standards** and the mentioned **challenges** in this study are 1) lack of application, 2) lack of analysis procedure.

In [43], authors propose a hybrid dynamic human factor model by integrating Human Factor Analysis and Classification System (HFACS) [51], fuzzy set theory, and Bayesian network to be used for analyzing accidents. The proposed model is called FBN-HFACs (Fuzzy Bayesian Network-HFACS). The model is used for identifying, characterizing and ranking human and organizational factors causing accidents. First step is scenario development which includes defining scope of the study, gathering data and information and developing the scenario of concern. Then, the next step is qualitative analysis, which is based on HFACS. In this step human factors at all levels are identified and causal model is represented. Finally, the last step is quantitative and inference analysis, which is based on Fuzzy theory, Bayesian Network and expert opinions. HFACS is mostly based on Reason's Swiss cheese model and consist of four levels of failures. These four levels are 1) organizational influences, 2) unsafe supervision, 3) preconditions for unsafe acts and 4) unsafe acts. It defines 19 causal categories and 69 subcategories within these four levels. By using the HFACS concepts characterizing causes of accidents, it has the **potential for capturing human and organizational aspects**. In addition, it has the **potential to capture risk** since it analyzes the consequences. However, the model does not propose concepts to capture **technical aspects**, **socio-technical orchestration**, **organizational changes effects**, **technological changes effects** and **AR effects**. The risks emanated from socio aspects are in focus, because capturing accident causes from the human and organization perspectives are considered. Regarding the support for the risk assessment process, it emerges that this model supports **risk identification** step by using the proposed concepts. **Risk analysis** is also supported (discussions are provided for considering risk sources, consequences, scenarios, likelihood, uncertainties, controls and their effectiveness). Finally, **risk evaluation** is supported (there are discussions on the results and required actions and there is comparison using probability to support decisions). Since the focus is on the chain of events and the root causes of accidents, it emerges that this model can be labelled as *safety engineering today* perspective. There are no syntax and semantics proposed and used and a

proper modeling language does not emerge, only an **informal notation** is present and the contributions of the model consist of providing integration of different approaches and there is **no extending formality contribution**. Discussions on causes of accidents are described as **qualitative, quantitative and linear analysis**. As it is explained in [46], causes of accidents in linear accident causation models such as HFACS are examined in various stages, thus we categorize this model which is based on HFACS, as a linear model. This model has a **backward and forward looking** perspective and there is no **structured analysis process** defining different steps. The proposed model does not provide **tool support** and it is proposed for **general domain**. However, an accident **scenario** from chemical process systems is presented as case study. There is no explicit discussion about **support for standards**, however the methods which are used in this model are usually suggested by different standards. The mentioned **challenges** in this study are 1) requiring further testing, 2) requiring detailed validation.

In [44], authors propose a foundational ontology-based conceptualization for main concepts of FRAM method. The conceptualization uses Unified Foundational Ontology (UFO) to represent the concepts of function and related aspects in FRAM. In addition, it provides semantics to limit variable interpretations of functions in FRAM and it contains well-formedness rules to relate the concepts. By using the proposed concepts and semantics characterizing human, organization and technological functions and related aspects and rules, it provides the **structured conceptualization for capturing human, organizational and technological aspects**. In addition, there is the **potential for capturing socio-technical orchestration** using the relation between concepts of human and organization functions and technological function. It does not propose concepts for characterizing **organizational changes effects**, **technological changes effects** and **AR effects** explicitly. In addition, it has the **potential to capture risk** since it is possible to analyze the consequences using the proposed conceptualization. Regarding the support for the risk assessment process, it emerges that the proposed conceptualization supports **risk identification** step by using the proposed concepts. In addition, **risk analysis** is supported (discussions can be provided for considering risk sources, consequences and scenarios). Finally, **risk evaluation** is also supported (there can be discussions on the results to support decisions). Since the conceptualization has the focus on FRAM, we can label it as *safety II* perspective. There are concepts and semantics proposed and used and **formal modeling** is provided and the contributions of the paper consist of a set of **concepts and semantics**. Discussions on causes of accidents using the proposed conceptualization can be described as **qualitative and linear analysis** with a **forward looking** perspective and there is no **structured analysis process**. The proposed conceptualization does not provide **tool support** and it is proposed for a **general domain**. However, a **scenario** from aviation domain is presented. There is no explicit discussion about **support for standards** and the mentioned **challenges** in this study are 1) lack of quantitative analysis, 2) lack of tool support.

In [45], the authors propose a method called SAfety FRactal ANalysis (SAFRAN) for improving safety management systems based on accident investigations. The method combines three distinct elements: fractal (description of what is required for controlling safety related activities), iterations (an investigation flow for guiding investigators where to continue the investigation) and basic steps. The analysis process in this method contains five main steps: 1) identifying performance variability 2) identifying the expected performance 3) identifying the source of performance variability 4) monitoring the variability 5) learning capability. The method is further developed in [52] by providing a taxonomy to specify human and organizational factors (HOF) required for identifying sources of performance variability. The taxonomy has five main categories: *dynamic situational, dynamic staff, static situational, static staff* and *socio interactional*. Each of these categories contain five factors. The method proposed in this paper provides **potential for capturing human, organizational, technological aspects** and **socio-technical orchestration** in the third step which is identifying the source of performance. However, it does not provide means for characterizing **organizational changes effects**, **technological changes effects** and **AR effects** explicitly. In addition, it has the **potential to capture risk** since it is possible to discuss the consequences using the proposed method. Regarding the support for the risk assessment process, it emerges that this paper supports **risk identification** step (the method can be used to find, recognize and describe the causal factors and risks). In addition, **risk analysis** is supported (using the method, the results can be provided by considering risk sources, consequences and scenarios). Finally, **risk evaluation** is also supported (there can be discussions on the results and required actions to support decisions). Since the non-linear interactions are considered, it emerges that this framework can be labelled as *safety III* perspective. There are no syntax and semantics proposed and used and a proper modeling language does not emerge, only an **informal notation** is present and a set of **concepts** are proposed that can be used for modeling accident causal factors. Discussions on causes of accidents using the proposed method can be described as **qualitative and non-linear** analysis. The perspective in this method is both **backward looking** (since accidents are analyzed) and **forward looking** (since provides predictions and models the system for preventing future accidents). This paper does not provide **structured analysis process** and **tool support**. It is proposed for **general application** and **scenarios** from railway domain are discussed based on available accident reports. It is not discussed if the method provides **support for standards** and the **challenges** of this study are not mentioned. We can consider lack of details and specified techniques in different steps of the method as a challenge.

In [46], authors introduce an accident causation model with the name 24Model. The name 24Model stands for a model of causes of accidents at 2 levels (individual and

organizational levels) and 4 stages (immediate, indirect, radical and root causes). Immediate and indirect causes are assigned to individual level and radical and root causes are assigned to organizational level. The proposed concepts characterizing immediate causes are *safety act* and *safety condition*. The proposed concepts characterizing indirect causes are *safety knowledge*, *safety awareness*, *safety habits*, *psychological status* and *physiological status*. The proposed concept characterizing radical cause is *safety management system*. Finally, the proposed concept characterizing root cause is *safety culture*. By using the proposed concepts characterizing causes of accidents, it has the **potential for capturing human and organizational aspects**. In addition, there is the possibility to analyze causality between the deviations and the causes and it has the **potential to capture risk** since it analyzes the consequences. However, the model does not propose concepts to capture **technical aspects** and **socio-technical orchestration**. In addition, it does not propose concepts for characterizing **organizational changes effects**, **technological changes effects** and **AR effects** explicitly. The risks emanated from socio aspects are in focus, because capturing accident causes from the human and organization perspectives are considered. Regarding the support for the risk assessment process, it emerges that 24Model supports **risk identification** step by using the proposed concepts. In addition, **risk analysis** is supported (discussions are provided for considering risk sources, consequences and scenarios). Finally, **risk evaluation** is also supported (there are discussions on the results and required actions to support decisions). Since the linear chain of events and the root cause are considered, it emerges that 24Model can be labelled as *safety engineering today* perspective. There are no syntax and semantics proposed and used and a proper modeling language does not emerge, only an **informal notation** is present and the contributions of the model consist of a set of **concepts**. Discussions on causes of accidents are described as **qualitative and linear analysis** with a **backward looking** perspective and there is no **structured analysis process**. The proposed model does not provide **tool support** and it is proposed for a **general domain**. However, an accident **scenario** from fire and explosion is presented as case study. There is no explicit discussion about **support for standards** and the mentioned **challenges** in this study are 1) lack of quantitative analysis, 2) lack of identification of the dynamic characteristics of systems, 3) lack of non-linear relationships characterization.

In [47], the authors propose an ontology-based method in order to prepare HAZOP worksheets automatically. In order to provide the conceptualization, they design a knowledge model containing relevant concepts in the form of ontology (concepts and their relationships are identified and modeled). They provide core concepts containing: *deviations, causes, super causes, effects, consequences*, and *safeguards* and complementary concepts containing: *substance, process unit, process* and *circumstances*. In addition, their description, and their relationships are provided as an ontological model. The ontology is then formalized using Web Ontology Language (OWL) and an inference strategy is designed and implemented to generate the HAZOP worksheets automatically from the proposed ontology and a process plant representation using extended concepts such as *causes, chain of consequences* and *safeguards*. The proposed method in this paper provides **structured conceptualization for capturing technological aspects** by using the concepts of the proposed ontology and their relations, while it does not provide **potential for capturing human, organizational, socio-technical orchestration, organizational changes effects, technological changes effects** and **AR effects** explicitly. However, the system is considered as a socio-technical system. In addition, it provides **structured conceptualization for capturing risk** since it is possible to discuss the consequences using the proposed concepts and the rules for their relations and it is possible to determine safeguards. Regarding the support for the risk assessment process, it emerges that this paper supports **risk identification** step (the approach can be used to find, recognize and describe the causal factors and risks). In addition, **risk analysis** is supported (using the approach, the results can be provided by considering risk sources, consequences and scenarios). Finally, **risk evaluation** is also supported (there can be discussions on the results, safeguards and required actions to support decisions). Since the focus is on the chain of events and the root causes of accidents, it emerges that this model can be labelled as *safety engineering today* perspective. There are syntax and semantics used from OWL (Web Ontology Language) modeling language. Thus, **formal modeling** is present and the contribution of the paper includes **concepts and semantics** to conceptualize HAZOP related knowledge. Discussions on causes of accidents using the proposed method can be described as **qualitative** and **linear** analysis. The perspective in this method is mostly **forward looking** (since it provides prediction about possible hazards). This paper provides **structured analysis process** by providing automated extended HAZOP worksheets and it also provides **tool support** using implemented python program. However, it still does not provide automatic risk assessment and presence of human experts is necessary. The method is proposed using knowledge from process and plant safety (PPS) domain (**specific application**) and a **scenario** from this domain is discussed as case study. However, it has the potential to be used for other applications as well. In this paper, it is not discussed if the proposed model provides **support for standards** and the mentioned **challenges** of this study are 1) requiring further research for providing more applications, 2) providing automatic risk assessment, and 3) providing safeguard interpretation.

In [48], authors describe a framework with the name Human Functions in Safety (HFiS) to express the role of human in railway safety. The framework contains concepts (for expressing functions, activities and contextual factors) and the relationship between these concepts and potential impact on safety. The proposed concepts of this framework are *system purpose/goal*, *human function goal*, *human functions*, *personal and organizational goals*, *generic context*, *safety*

relevant activities, *potential error/ recovery/ consequence/ mitigation*. Each of the concepts includes detailed descriptive content containing subcategories and examples. 66 human functions performed by frontline staff and associated activities to railways are identified in this framework and their relation with 8 human function goals are determined. The framework provides **structured conceptualization for socio aspects** as part of socio-technical systems using the proposed concepts and the rules for their relations. However, there are no concepts to capture **technical aspects**, **socio-technical orchestration**, **organizational changes effects**, **technological changes effects** and **AR effects** explicitly. On the other hand there is no structured conceptualization for risk assessment. However, there is the **potential for capturing risk** since it analyzes the consequences. Regarding the support for the risk assessment process, it emerges that HFiS supports **risk identification** step by using the proposed concepts. In addition, **risk analysis** is supported (discussions are provided for considering risk sources, consequences and scenarios Finally, **risk evaluation** is also supported (there are discussions on the results and required actions to support decisions). Since the main focus of the model is on the role of human in system safety, it emerges that HFiS supports *safety II*. There are no syntax and semantics proposed and used and a proper modeling language does not emerge, only an **informal notation** is present and the contributions of the model consist of a set of **concepts**. Discussions on errors and consequences are described as **qualitative and linear analysis** with a **forward looking** perspective, nevertheless there is **no structured analysis process** and it does not provide **tool support**. This framework is specifically proposed for railway as an **specific application** by using railway **scenarios**. However, there is the **potential for other applications**, because it proposes a guidance for other safety-related domains. Rail safety and Standards Board are used as source of information, nevertheless there is no discussion for **support for standards**. The mentioned **challenges** of the proposed framework are 1) complexity in terms of the number of functions, 2) requiring availability of data sources for using in other domains, 3) requiring further study for quantitative analysis, 4) lack of identification of the dynamic characteristics of systems, 5) lack of feedback mechanisms characterization, 6) lack of delays characterization and 7) lack of non-linear relationships characterization.

In [49], the authors propose a framework with the name FRAAR for risk assessment of AR-equipped socio-technical systems based on their proposed modeling extensions for a modeling language. This framework provides the possibility for modeling and analyzing technical aspects, various socio aspects, organizational changes effects, technological changes effects, AR-extended human functions and AR-related influencing factors using modeling extensions of SafeConcert modeling language [53]. In addition, Concerto-FLA analysis technique [54] is used to provide the analysis results. There are four main steps in this framework. The first step is modeling involved entities containing technical and socio entities as composite components. The second step is identifying important aspects of each entity and modeling them as sub-components using the modeling extensions such as *organization and regulation AR adoption* modeling element. The third step is modeling the behavior of each sub-component using FPTC syntax. Finally, last step is analyzing system behavior based on Concerto-FLA analysis technique. Details of these steps are described and it is shown that how these steps would support safety standards such as ISO 26262 and SOTIF. The proposed framework in this paper provides **structured conceptualization for capturing technological aspects, socio aspects, organizational changes effects, technological changes effects** and **AR effects** by using the proposed concepts used in extended SafeConcert modeling language. In addition, it provides **structured conceptualization for capturing socio-technical orchestration** by modeling the relations between the socio and technical concepts. Furthermore, it provides **structured conceptualization for capturing risk** since it is possible to determine the consequences and to define safety goals using Concerto-FLA analysis technique. Regarding the support for the risk assessment process, it emerges that this paper supports **risk identification** step (the approach can be used to find, recognize and describe the causal factors and risks). In addition, **risk analysis** is supported (using the approach, the results can be provided by considering risk sources, consequences and scenarios). Finally, **risk evaluation** is also supported (there can be discussions on the results, safety goals and required actions to support decisions). Since the focus is on chain of events and the root causes of accidents, it emerges that this model can be labelled as *safety engineering today* perspective. There are syntax and semantics used from SafeConcert modeling language and Concerto-FLA analysis technique. Thus, **formal modeling** is present and the contribution of the paper includes **concepts** to conceptualize various socio aspects such as organizational/technological changes effects and AR-related aspects. Discussions on causes of accidents using the proposed method can be described as **qualitative** and **linear** analysis. The perspective in this method is mostly **forward looking** (since it provides prediction about possible future accidents). This paper provides **structured analysis process** by using Concerto-FLA analysis technique. It does not provide **tool support**. However, there is the potential for providing it by implementing the proposed extensions. The method is proposed for **general application**. However, the examples and standards are from automotive domain and some **scenarios** from this domain are discussed as case study. In this paper, it is presented how different steps of the framework provide **support for ISO 26262 and SOTIF standards** and the mentioned **challenges** of this study are 1) requiring further research for providing more applications, 2) providing automatic risk assessment by implementing the extensions, and 3) providing scenarios from other domains.

In [50], authors propose a model-based safety framework by considering railway infrastructure information to be used for autonomous train driving. The proposed safety framework is composed of three main parts: 1) safety analysis 2) model extension 3) safety management. In order to analyze safety, it

uses concepts and semantics defined by DAO (Dysfunctional Analysis Ontology) [55]. The DAO concepts are *Failure*, *Exposure*, *Defect & fault*, *Fault emergence failure*, *Hazard* and *Safety measure* and it contains well-formedness rules to relate these concepts. The sources for these concepts are safety engineering standards such as IEC 61508. Based on these concepts, their relation and specific dangerous events safety model is obtained. Then, safety rules/measures and safety analysis are provided based on the safety model. An extended model for the railway infrastructure is proposed based on the safety rules in order to enable automating safety management decisions. Safety management is provided based on GOSMO concepts containing *SafetyMeasure*, *Task*, *StakeholderRole*, *Context*, *Organization*, *Assignment*, *Permission* (it also contains well-formedness rules to relate these concepts). However the framework in this paper is proposed for autonomous train driving, it still provides **structured conceptualization** for **human, organizational and technological aspects and socio-technical orchestration** because of using the GOSMO concepts such as *StakeholderRole*, *Organization*, *Task* and their related semantics and rules to relate these concepts. However, it does not propose means for characterizing **organizational changes effects**, **technological changes effects** and **AR effects** explicitly. In addition, it provides **structured conceptualization to capture risk** since it uses DAO concepts such as *Failure*, *Hazard*, *Safety measure* and their related semantics and rules to relate these concepts. Regarding the support for the risk assessment process, it emerges that this paper supports **risk identification** step in the first part (safety analysis). In addition, **risk analysis** is supported (the results can be provided for considering risk sources, consequences, scenarios and controls). Finally, **risk evaluation** is also supported (there can be discussions on the results and required actions, to support decisions). Since the focus is on chain of events and the root causes of accidents, it emerges that this model can be labelled as *safety engineering today* perspective. In this framework, **formal modeling languages** are used and the contributions of the paper consist of **concepts** for modeling the connections between different causal factors based on the previous accident reports. Discussions on causes of accidents using the proposed model can be described as **qualitative and linear** analysis (considering chain of events). The perspective in this model is **forward looking**, since it provides predictions and focuses on modeling and analyzing concepts for preventing the accidents in the future. This paper provides **structured analysis process** using the DAO concepts. Since the proposed framework provides automated safety management, it can provide **tool support**. It is proposed for a **specific domain** (railway) and a **scenario** from this domain is presented. However, there is the potential to use it for other domains. There are discussions about **support for standards** because of using the concepts gained from safety engineering standards such as IEC 61508. The mentioned **challenge** in this study is lack of formal verification for checking safety rules consistency and the safety justification.

5 Discussion

5.1 Discussion on the Results

In this subsection we discuss about the results of our SLR and the summarized information provided in the tables for the reviewed papers.

As it is shown in Table VII, there are few methods/techniques/models/frameworks providing structured conceptualization for socio-technical systems and risk assessment and in most cases there are potential for capturing which is provided through conceptual modeling. Based on these results there is a need for more work on providing structured conceptualization to be used for characterizing different aspects of a socio-technical system and risk assessment. In addition, it is noticeable that few papers provide the potential for capturing effects of organizational changes, technological changes and augmented reality. It is not surprising since these organizational and technological changes are recent and augmented reality is a rather novel technology. However, because of the extensive applications of AR technology and because of the broad effects of organizational and technological changes, it is essential to consider conceptualizing the related aspects to enable capturing their effects on system safety and risk assessment.

As it is shown in Table VIII, in spite of providing risk identification, analysis and evaluation in all papers, the risk and dependability characterization is not provided in a structured manner and instead there is potential for capturing. Thus, more research is required on providing structured conceptualization for characterizing risk and dependability. It is also observable that Safety II and Safety III perspectives are used in some of the methods/techniques/models/frameworks and this means that considering interactions between socio and technical aspects in addition to human error studies are receiving more attention which shows the progress in this context. However, it is important to use these different perspectives as complementary aspects for improving and developing the conceptualization of risk assessment for socio-technical systems.

As it is shown in Table IX, in most papers the modeling formality is in the level of informal notation and we can conclude that more research is required in the context of proposing syntax and semantics and providing/using semi-formal and formal modeling languages. It also influences on tool support which is not provided in most of the papers. Improving formality leads to improving the possibility for providing tool support and providing increased automation. In addition, based on the results shown on the table we identify that most of the works provide qualitative and linear analysis. It is not surprising since the incorporation of socio aspects in the analysis requires to provide qualitative analysis or a mixture of qualitative and quantitative results. However, it is substantial to consider non-linear interactions and more research is required for improving the analysis by incorporating the non-linear interactions and overcoming the complexities due to the non-linearity. Forward and backward looking are both considered in different works and it is important to

ID	Socio entities characterization	Technical aspects characterization	Socio-technical orchestration characterization	Organizational changes effects characterization	Technological changes effects characterization	AR effects characterization
[32]	PfC	PfC	PfC	NC	NC	NC
[33]	PfC	PfC	PfC	NC	NC	NC
[34]	PfC	PfC	PfC	NC	NC	NC
[35]	PfC	PfC	PfC	NC	NC	NC
[36]	SC	SC	SC	PfC	NC	NC
[37]	PfC	PfC	PfC	NC	NC	NC
[38]	PfC	PfC	PfC	NC	NC	NC
[39]	PfC	PfC	PfC	NC	NC	NC
[40]	PfC	PfC	PfC	NC	NC	NC
[41]	PfC	PfC	PfC	NC	NC	NC
[42]	PfC	PfC	PfC	PfC	PfC	NC
[43]	PfC	NC	NC	NC	NC	NC
[44]	SC	SC	PfC	NC	NC	NC
[45]	PfC	PfC	PfC	NC	NC	NC
[46]	PfC	NC	NC	NC	NC	NC
[47]	NC	SC	NC	NC	NC	NC
[48]	SC	NC	NC	NC	NC	NC
[49]	SC	SC	SC	SC	SC	SC
[50]	SC	SC	SC	NC	NC	NC

PfC: Potential for Capturing. SC: Structured Conceptualization. NC: No Characterization.

TABLE VII: Summary of the reviewed primary studies in relation to the first research question

ID	Risk/dependability characterization	Provided steps of risk assessment process	Safety perspective
[32]	PfC	RI, RA, RE	Set
[33]	SC	RI, RA, RE	Set
[34]	PfC	RI, RA, RE	Set
[35]	PfC	RI, RA, RE	Set
[36]	PfC	RI, RA, RE	Safety III
[37]	PfC	RI, RA, RE	Set
[38]	PfC	RI, RA, RE	Safety II
[39]	PfC	RI, RA, RE	Safety III
[40]	PfC	RI, RA, RE	Safety III
[41]	PfC	RI, RA, RE	Set
[42]	PfC	RI, RA, RE	Set
[43]	PfC	RI, RA, RE	Set
[44]	PfC	RI, RA, RE	Safety II
[45]	PfC	RI, RA, RE	Safety III
[46]	PfC	RI, RA, RE	Set
[47]	SC	RI, RA, RE	Set
[48]	PfC	RI, RA, RE	Safety II
[49]	SC	RI, RA, RE	Set
[50]	SC	RI, RA, RE	Set

PfC: Potential for Capturing. SC: Structured Conceptualization. NC: No Characterization.
RI: Risk Identification. RA: Risk Analysis. RE: Risk Evaluation.
Set: Safety engineering today.

TABLE VIII: Summary of the reviewed primary studies in relation to the first research question (Con.)

ID	Modeling formality	Contribution context	Type of analysis (Ql/Qn)	Type of analysis (Ln/NL)	Type of analysis (FL/BL)	Structured analysis process	Tool support
[32]	IN	NEFC	Ql + Qn	Ln	BL+FL	No	No
[33]	IN	Concepts	Ql + Qn	Ln	BL+FL	Yes	No
[34]	IN	Concepts	Ql	Ln	FL	No	No
[35]	IN	NEFC	Ql + Qn	Ln	FL	Yes	Yes
[36]	IN	Concepts	Ql	NL	FL	No	No
[37]	IN	Concepts	Ql + Qn	Ln	BL	No	No
[38]	IN	Concepts	Ql + Qn	NL	BL	No	Yes
[39]	IN	Concepts	Ql	NL	FL	No	No
[40]	IN	Concepts	Ql	NL	FL	No	No
[41]	IN	Concepts	Ql + Qn	Ln	BL	No	No
[42]	IN	Concepts	Ql	Ln	FL	No	No
[43]	IN	NEFC	Ql + Qn	Ln	BL+FL	Yes	No
[44]	FM	Concepts + Semantics	Ql	Ln	FL	No	No
[45]	IN	Concepts	Ql	NL	BL+FL	No	No
[46]	IN	Concepts	Ql	Ln	BL	No	No
[47]	FM	Concepts + Semantics	Ql	Ln	FL	Yes	Yes
[48]	IN	Concepts	Ql	Ln	FL	No	No
[49]	FM	Concepts	Ql	Ln	FL	Yes	No
[50]	FM	Concepts	Ql	Ln	FL	Yes	Yes

IN: Informal Notation. FM: Formal Modeling. Ql: Qualitative. Qn: Quantitative.
NEFC: No Extending Formality Contribution.
Ln: Linear. NL: Non-linear. FL: Forward Looking. BL: Backward Looking

TABLE IX: Summary of the reviewed primary studies in relation to the second research question

ID	Application	Potential for other applications	Support for standards	Presence of scenarios
[32]	General	Yes	NM	Yes
[33]	Specific	Yes	M (IEC 61508 and IEC 61511)	Yes
[34]	General	Yes	NM	Yes
[35]	General	Yes	NM	Yes
[36]	Specific	Yes	NM	Yes
[37]	Specific	No	NM	Yes
[38]	Specific	Yes	NM	Yes
[39]	General	Yes	NM	Yes
[40]	General	Yes	M (SAE ARP-4761)	Yes
[41]	General	Yes	NM	Yes
[42]	Specific	Yes	NM	No
[43]	General	Yes	NM	Yes
[44]	General	Yes	NM	Yes
[45]	General	Yes	NM	Yes
[46]	General	Yes	NM	Yes
[47]	Specific	Yes	NM	Yes
[48]	Specific	Yes	NM	Yes
[49]	General	Yes	M (ISO 26262 and ISO/PAS 21448-SOTIF)	Yes
[50]	Specific	Yes	M (IEC 61508, etc.)	Yes

NM: Not Mentioned. M: Mentioned

TABLE X: Summary of the reviewed primary studies in relation to the third research question

ID	Stated challenges
[32]	1) Lack of readily available data to be used by human factor approaches that can be used in the framework
[33]	1) Determining rates in a way to allow certain influence of a factor, 2) difficulty in determining proportion of design SIL and weights of the factors, 3) requiring further research for providing validation, 4) providing some more applications, 5) ensuring consistency over time in the ratings, 6) including effects of system modifications and aging of equipment, 7) incorporating other safety influencing factors
[34]	1) Requiring tools for evaluating quantitative analysis, 2) requiring exploration to mesh with existing safety/dependability assurance processes
[35]	1) Requiring further research for understanding the influence of human and organizational factors on safe operations
[36]	1) Lack of measures to mitigate the risks, 2) not using information from reality such as interviews or analysis of information flows in the development of the methodology
[37]	1) Use of limited reports from specific databases, 2) subjectivity in the reports
[38]	1) No criteria for assessment of the significance of introducing this approach
[39]	1) Lack of quantitative analysis, 2) limited scope of case study, 3) lack of assessment of practicality and validity of the framework in macro level, 4) lack of comparison with other widespread methods (other than STPA which is done)
[40]	1) Requiring further study for applicability in larger systems, 2) requiring further study for automating the process, 3) no criteria for assessment of the significance of introducing this approach to existing hazard analysis
[41]	1) Requiring further research for providing details of the proposed broad concepts
[42]	1) Lack of application, 2) lack of analysis procedure
[43]	1) Requiring further testing, 2) requiring detailed validation
[44]	1) Lack of quantitative analysis, 2) lack of tool support
[45]	Not mentioned
[46]	1) Lack of quantitative analysis, 2) lack of identification of the dynamic characteristics of systems, 3) lack of non-linear relationships characterization
[47]	1) Requiring further research for providing more applications, 2) providing automatic risk assessment, and 3) providing safeguard interpretation
[48]	1) Complexity in terms of the number of functions, 2) requiring availability of data sources for using in other domains, 3) requiring further study for quantitative analysis, 4) lack of identification of the dynamic characteristics of systems, 5) lack of feedback mechanisms characterization, 6) lack of delays characterization and 7) lack of non-linear relationships characterization
[49]	1) Requiring further research for providing more applications, 2) providing automatic risk assessment by implementing the extensions, and 3) providing scenarios from other domains
[50]	1) Lack of formal verification for checking safety rules consistency and the safety justification

TABLE XI: Summary of the reviewed primary studies in relation to the fourth research question

consider both of them since we learn from the past to prevent the accidents in the future. It is also identified from the table that there are few works providing structured analysis process and there is a need for more research in this context.

As it is shown in Table X, there are methods/techniques/models/frameworks for both specific and general applications. However, almost all of them have the potential to be used for other applications. Thus, it is important to consider different domains since it is possible to use methods/techniques/models/frameworks from other domains with tiny changes. Based on the table, there are few papers providing discussions on how they support safety standards. However, they may have the potential to support different safety standards. Thus, it is important to provide evidence on how they can support the standards to ease their selection when practitioners need to choose a method/technique/model/framework for complying with standards. It is also shown that there are scenarios presented in almost all papers which shows a positive feature of the works since it is really important to show the capabilities of the contributions on specific scenarios.

As it is shown in Table XI, there are different challenges provided by different studies. Some of the most important challenges are lack of input data to be used in different phases of the studies, lack of defined criteria for validating and measuring significance of the contributions in different levels, lack of characterization means for specific characteristics of systems such as non-linearity, dynamic behavior, existence of delays and feedback mechanisms, lack of formality and tool-support, lack of sufficient applications, lack of various scenarios from different domains, lack of comparisons with other known methods, existence of subjectivity, complexity and inconsistency over time. Although these challenges are not specific for safety-critical socio-technical systems and they are general challenges in the context of safety and risk analysis, still they provide the possible directions for future work and for extending the current works to have improved risk assessment for socio-technical systems. In addition, there is abundant room for further progress in considering effects of new technological and organizational changes effects on system behavior.

5.2 Threats to Validity

In this subsection, we discuss about validity of the results based on the guideline provided in [56] and [12]. Specifically, we discuss about possible threats regarding publication bias, identification of primary studies and data extraction consistency.

5.2.1 Publication Bias Threats:

Publication bias threats refer to the problem that positive results may have more chance than negative results to be published. It can become more of a problem when specific method or technique is sponsored by influential groups in industry. Our work is not sponsored by influential groups for a specific aim and we used the standard search strategy based on [12] and we designed a SLR protocol in Section 3.1.3. The first author provided the protocol and the second author, who is an expert in the area with previous experiences in providing SLR performed a comprehensive review and assessment. We also scanned grey literature (e.g., standards) to be aware of possible evidences which are not published as articles in journals or conferences.

5.2.2 Identification of Primary Studies Threats:

Identification of primary studies threats refer to the problems in identifying the related studies. In order to prevent threats regarding identification of primary studies, we used standard search strategy based on [12]. We provided search string based on our SLR goal and research questions using the PICO criteria [24]. We selected databases based on systematic literature reviews best practices and database evaluation in the literature. We defined inclusion and exclusion criteria and quality assessment criteria for assessing the studies and identifying the final primary studies.

5.2.3 Data Extraction Consistency Threats:

Data extraction consistency threats refer to the problems in data extraction in consistent manner if the process is done by several researchers. In this SLR the data extraction process is completely done by the first author and the second author reviewed and assessed the process. Thus, there were not several researchers involved in the data extraction process. Since the first author is a PhD student, other checking techniques are used. For example, supervisor (the second author) performed random check for primary studies and their results. In addition, we defined data extraction criteria shown in Table III and we defined the abbreviations used for extracted data in Fig. 1. In addition, we checked and updated them iteratively while we performed the data extraction from primary studies. The aim for defining these criteria and abbreviations is to provide consistent extracted data and to decrease subjectivity while analyzing the primary studies. For each primary study we provided summary in structured manner and we filled TablesVII - XI with information in relation to each research question.

6 Conclusions and Future Work

In this paper, we conducted a systematic literature review to characterize works on risk assessment of safety-critical socio-technical systems based on the development of socio-technical systems such as new organizational and technological changes included in the systems, development of safety standards and safety perspectives. To conduct our systematic literature review we followed best practices, i.e., we defined research questions, search strategy and search string. In addition, we defined study selection criteria and we used databases selected by best practices of systematic literature reviews considering recent database evaluation. Furthermore, we defined study selection procedure and quality assessment criteria in order to select the most relevant publications to the focus of our SLR. Finally, we extracted data from the selected primary studies based on the defined research questions and we provided a structured summary for each primary study. The extracted information is also summarized in tables for more efficient comparability. Based on the research questions, we considered the conceptualization of risk assessment and socio-technical systems (we consider characterization of socio aspects, technical aspects, organizational and technological changes effects, AR effects, risk and assessment process). In addition, we considered the provided safety perspective, level of formality, type of analysis, tool support, application domain, support for standards and mentioned challenges in all the selected primary studies. Then, we provided discussion on the results and we discussed the potential for future work based on the analysis of the primary studies.

In the future, we aim at considering the possible future directions extracted from the identified challenges of primary studies in order to develop the current techniques in risk assessment of safety-critical socio-technical systems. The results of our SLR indicate that most of the papers focus on providing potential for capturing socio-technical aspects and risk and dependability aspects. This means that the structured conceptualization is not provided in most cases. Since the structured conceptualization provides the possibility for increasing formality and for providing tool support, it is crucial to have more research investigating on proposing structured conceptualization and in consequence providing higher level of formality and automation. In the future, formality level of conceptual modeling languages shall be improved by providing syntax, semantics and tool support. In addition, based on the results of our SLR, there is a need for providing more application scenarios considering the current contributions in the risk assessment of safety-critical socio-technical systems. One future research direction can be providing more scenarios from different domains in addition to providing discussion on support for related safety standards to illustrate the applicability of current contributions in risk assessment of safety-critical socio-technical systems. In this paper, characterization of effects of AR is considered. However, there are other technologies that may influence on human behavior and system behavior. Another future research direction is identifying the other influential technologies and characterizing their effects. Furthermore, based on the results of our SLR, further research should be undertaken to investigate dynamic characteristics of the systems, non-linear relationships, feedback mechanisms and delay characterization. There is also abundant room for further progress in validating the proposed contributions,

proposing criteria to assess the significance of the proposed approaches and determining measures to mitigate the identified risks.

References

[1] ISO(31000), Risk management – Guidelines (2018).
URL https://www.iso.org/iso-31000-risk-management.html

[2] G. H. Walker, N. A. Stanton, P. M. Salmon, D. P. Jenkins, A review of sociotechnical systems theory: a classic concept for new command and control paradigms, Theoretical issues in ergonomics science 9 (6) (2008) 479–499. doi:10.1080/14639220701635470.

[3] J. C. Knight, Safety critical systems: challenges and directions, in: Proceedings of the 24th International Conference on Software Engineering, 2002, pp. 547–550. doi:10.1145/581339.581406.

[4] J.-C. Le Coze, Post Normal Accident: Revisiting Perrow's Classic, CRC Press, 2020. doi:10.1201/9781003039693.

[5] J.-C. Le Coze, Globalization and high-risk systems, Policy and practice in health and safety 15 (1) (2017) 57–81. doi:10.1080/14773996.2017.1316090.

[6] K. Lebeck, T. Kohno, F. Roesner, How to safely augment reality: Challenges and directions, in: Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications, Association for Computing Machinery, 2016, pp. 45–50. doi:10.1145/2873587.2873595.

[7] R. R. Lutz, Safe-AR: Reducing Risk While Augmenting Reality, in: 2018 IEEE 29th International Symposium on Software Reliability Engineering (ISSRE), IEEE, 2018, pp. 70–75. doi:10.1109/ISSRE.2018.00018.

[8] C. Dallat, P. M. Salmon, N. Goode, Risky systems versus risky people: To what extent do risk assessment methods consider the systems approach to accident causation? A review of the literature, Safety Science 119 (2019) 266–279. doi:10.1016/j.ssci.2017.03.012.

[9] R. Patriarca, M. Chatzimichailidou, N. Karanikas, G. Di Gravio, The past and present of System-Theoretic Accident Model And Processes (STAMP) and its associated techniques: A scoping review, Safety science 146 (2022) 105566. doi:10.1016/j.ssci.2021.105566.

[10] N. Leveson, A new accident model for engineering safer systems, Safety Science 42 (4) (2004) 237–270. doi:10.1016/S0925-7535(03)00047-X.

[11] T. Ishimatsu, N. G. Leveson, J. Thomas, M. Katahira, Y. Miyamoto, H. Nakao, Modeling and hazard analysis using STPA (2010).

[12] B. Kitchenham, S. Charters, Guidelines for performing systematic literature reviews in software engineering, Tech. rep., Keele University and Durham University Joint Report (2007).
URL https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf

[13] International Organization for Standardization (ISO), ISO 26262: Road vehicles — Functional safety (2018).
URL https://www.iso.org/standard/68383.html

[14] ISO 21448, Road vehicles — Safety of the intended functionality (SOTIF) (2022).
URL https://www.iso.org/standard/77490.html

[15] E. Hollnagel, R. L. Wears, J. Braithwaite, From Safety-I to Safety-II: a white paper, The resilient health care net: published simultaneously by the University of Southern Denmark, University of Florida, USA, and Macquarie University, Australia (2015). doi:10.13140/RG.2.1.4051.5282.

[16] N. Leveson, Safety III: A systems approach to safety and resilience (2020).
URL http://sunnyday.mit.edu/safety-3.pdf

[17] T. Aven, A risk science perspective on the discussion concerning Safety I, Safety II and Safety III, Reliability Engineering & System Safety 217 (2022) 108077. doi:10.1016/j.ress.2021.108077.

[18] T. Aven, Risk assessment and risk management: Review of recent advances on their foundation, European Journal of Operational Research 253 (1) (2016) 1–13. doi:10.1016/j.ejor.2015.12.023.

[19] H. J. Pasman, W. J. Rogers, M. S. Mannan, How can we improve process hazard identification? What can accident investigation methods contribute and what other recent developments? A brief historical survey and a sketch of how to advance, Journal of loss prevention in the process industries 55 (2018) 80–106. doi:10.1016/j.jlp.2018.05.018.

[20] A. Adriaensen, W. Decré, L. Pintelon, Can Complexity-Thinking Methods Contribute to Improving Occupational Safety in Industry 4.0? A Review of Safety Analysis Methods and Their Concepts, Safety 5 (4) (2019). doi:10.3390/safety5040065.

[21] R. Sadeghi, F. Goerlandt, The State of the Practice in Validation of Model-Based Safety Analysis in Socio-Technical Systems: An Empirical Study, Safety 7 (4) (2021) 72. doi:10.3390/safety7040072.

[22] N. Berx, W. Decré, I. Morag, P. Chemweno, L. Pintelon, Identification and classification of risk factors for human-robot collaboration from a system-wide perspective, Computers & Industrial Engineering 163 (2022) 107827. doi:10.1016/j.cie.2021.107827.

[23] H. M. Cooper, Organizing knowledge syntheses: A taxonomy of literature reviews, Knowledge in society 1 (1) (1988) 104–126. doi:10.1007/BF03177550.

[24] B. Kitchenham, E. Mendes, G. H. Travassos, A systematic review of cross-vs. within-company cost estimation studies, in: 10th International Conference on Evaluation and Assessment in Software Engineering (EASE) 10, 2006, pp. 1–10. doi:10.14236/ewic/EASE2006.10.

[25] M. Gusenbauer, N. R. Haddaway, Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google

Scholar, PubMed, and 26 other resources, Research synthesis methods 11 (2) (2020) 181–217. doi:10.1002/jrsm.1378.

[26] M. Campoverde-Molina, S. Luján-Mora, L. Valverde, et al., Systematic literature review on software architecture of educational websites, IET Software 15 (4) (2021) 239–259. doi:10.1049/sfw2.12024.

[27] B. Kitchenham, P. Brereton, A systematic review of systematic review process research in software engineering, Information and software technology 55 (12) (2013) 2049–2075. doi:10.1016/j.infsof.2013.07.010.

[28] J. P. Castellanos Ardila, B. Gallina, F. Ul Muram, Compliance checking of software processes: A systematic literature review, Journal of Software: Evolution and Process e2440doi:10.1002/smr.2440.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/smr.2440

[29] H. Erik, FRAM: the functional resonance analysis method: modelling complex socio-technical systems, CRC Press, 2017. doi:10.1201/9781315255071.

[30] E. Hollnagel, Safety–I and safety–II: the past and future of safety management, CRC press, 2018. doi:10.1201/9781315607511.

[31] M. M. Chatzimichailidou, I. M. Dokas, RiskSOAP: Introducing and applying a methodology of risk self-awareness in road tunnel safety, Accident Analysis & Prevention 90 (2016) 118–127. doi:10.1016/j.aap.2016.02.005.

[32] P. C. Cacciabue, Human error risk management for engineering systems: a methodology for design, safety assessment, accident investigation and training, Reliability Engineering & System Safety 83 (2) (2004) 229–240. doi:10.1016/j.ress.2003.09.013.

[33] M. Schönbeck, M. Rausand, J. Rouvroye, Human and organisational factors in the operational phase of safety instrumented systems: A new approach, Safety science 48 (3) (2010) 310–318. doi:10.1016/j.ssci.2009.11.005.

[34] R. Lock, I. Sommerville, Modelling and analysis of socio-technical system of systems, in: 15th International Conference on Engineering of Complex Computer Systems, IEEE, 2010, pp. 224–232. doi:10.1109/ICECCS.2010.40.

[35] M. Farcasiu, I. Prisecaru, MMOSA–a new approach of the human and organizational factor analysis in PSA, Reliability Engineering & System Safety 123 (2014) 91–98. doi:10.1016/j.ress.2013.10.004.

[36] I. Bider, H. Otto, Modeling a global software development project as a complex socio-technical system to facilitate risk management and improve the project structure, in: 10th International Conference on Global Software Engineering, IEEE, 2015, pp. 1–12. doi:10.1109/ICGSE.2015.13.

[37] A. Mazaheri, J. Montewka, J. Nisula, P. Kujala, Usability of accident and incident reports for evidence-based risk modeling–A case study on ship grounding report, Safety science 76 (2015) 202–214. doi:10.1016/j.ssci.2015.02.019.

[38] K. Klockner, Y. Toft, Accident modelling of railway safety occurrences: the safety and failure event network (SAFE-Net) method, Procedia Manufacturing 3 (2015) 1734–1741. doi:10.1016/j.promfg.2015.07.487.

[39] A. Dehghan Nejad, R. Gholamnia, A. Alibabaee, A new framework to model and analyze organizational aspect of safety control structure, International Journal of System Assurance Engineering and Management 8 (2) (2017) 1008–1025. doi:10.1007/s13198-016-0561-9.

[40] C. Leong, T. Kelly, R. Alexander, Incorporating epistemic uncertainty into the safety assurance of socio-technical systems, Electronic Proceedings in Theoretical Computer Science 259 (2017) 56–71. doi:10.4204\%2Feptcs.259.7.

[41] W. Li, L. Zhang, W. Liang, An Accident Causation Analysis and Taxonomy (ACAT) model of complex industrial system from both system safety and control theory perspectives, Safety science 92 (2017) 94–103. doi:10.1016/j.ssci.2016.10.001.

[42] P.-c. Li, L. Zhang, L.-c. Dai, X.-f. Li, Y. Jiang, A new organization-oriented technique of human error analysis in digital NPPs: Model and classification framework, Annals of Nuclear Energy 120 (2018) 48–61. doi:10.1016/j.anucene.2018.05.021.

[43] E. Zarei, M. Yazdi, R. Abbassi, F. Khan, A hybrid model for human factor analysis in process accidents: FBN-HFACS, Journal of loss prevention in the process industries 57 (2019) 142–155. doi:10.1016/j.jlp.2018.11.015.

[44] A. Lališ, R. Patriarca, J. Ahmad, G. Di Gravio, B. Kostov, Functional modeling in safety by means of foundational ontologies, Transportation research procedia 43 (2019) 290–299. doi:10.1016/j.trpro.2019.12.044.

[45] B. Accou, G. Reniers, Developing a method to improve safety management systems based on accident investigations: The SAfety FRactal ANalysis, Safety science 115 (2019) 285–293. doi:10.1016/j.ssci.2019.02.016.

[46] G. Fu, X. Xie, Q. Jia, Z. Li, P. Chen, Y. Ge, The development history of accident causation models in the past 100 years: 24Model, a more modern accident causation model, Process Safety and Environmental Protection 134 (2020) 47–82. doi:10.1016/j.psep.2019.11.027.

[47] J. I. Single, J. Schmidt, J. Denecke, Ontology-based computer aid for the automation of HAZOP studies, Journal of Loss Prevention in the Process Industries 68 (2020) 104321. doi:10.1016/j.jlp.2020.104321.

[48] B. Ryan, D. Golightly, L. Pickup, S. Reinartz, S. Atkinson, N. Dadashi, Human functions in safety-developing a framework of goals, human functions and safety relevant activities for railway socio-technical systems, Safety

science 140 (2021) 105279. doi:10.1016/j.ssci.2021.105279.

[49] S. Sheikh Bahaei, B. Gallina, M. Vidović, A case study for risk assessment in AR-equipped socio-technical systems, Journal of Systems Architecture 119 (2021) 102250. doi:10.1016/j.sysarc.2021.102250.

[50] N. Chouchani, S. Debbech, M. Perin, Model-based safety engineering for autonomous train map, Journal of Systems and Software 183 (2022) 111082. doi:10.1016/j.jss.2021.111082.

[51] S. A. Shappell, D. A. Wiegmann, The human factors analysis and classification system–HFACS, Tech. rep., Federal Aviation Administration, Civil Aeromedical Institute, Retrieved from https://commons.erau.edu/publication/737 (2000).

[52] B. Accou, F. Carpinelli, Systematically investigating human and organisational factors in complex socio-technical systems by using the "SAfety FRactal ANalysis" method, Applied ergonomics 100 (2022) 103662. doi:10.1016/j.apergo.2021.103662.

[53] L. Montecchi, B. Gallina, SafeConcert: A metamodel for a concerted safety modeling of socio-technical systems, in: International Symposium on Model-Based Safety and Assessment, Springer, 2017, pp. 129–144. doi:10.1007/978-3-319-64119-5_9.

[54] B. Gallina, E. Sefer, A. Refsdal, Towards safety risk assessment of socio-technical systems via failure logic analysis, in: International Symposium on Software Reliability Engineering Workshops, IEEE, 2014, pp. 287–292. doi:10.1109/ISSREW.2014.49.

[55] S. Debbech, S. C. Dutilleul, P. Bon, An Ontological Approach to Support Dysfunctional Analysis for Railway Systems Design, J. Univers. Comput. Sci. 26 (5) (2020) 549–582. doi:10.3897/jucs.2020.030.

[56] S. Keele, et al., Guidelines for performing systematic literature reviews in software engineering, Tech. rep., Technical report, ver. 2.3 ebse technical report. ebse (2007).
URL https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf