Mälardalen University Press Dissertations
No. 371

ORGANIZATIONAL CHANGES-AWARE SAFETY-CENTERED RISK ASSESSMENT IN AUGMENTED REALITY-EQUIPPED SOCIO-TECHNICAL SYSTEMS

Soheila Sheikh Bahaei

2023

School of Innovation, Design and Engineering

ISBN 978-91-7485-577-7
ISSN 1651-4238
Printed by E-Print AB, Stockholm, Sweden

Mälardalen University Press Dissertations
No. 371

ORGANIZATIONAL CHANGES-AWARE SAFETY-CENTERED RISK ASSESSMENT IN AUGMENTED REALITY-EQUIPPED SOCIO-TECHNICAL SYSTEMS

Soheila Sheikh Bahaei

Akademisk avhandling

som för avläggande av teknologie doktorsexamen i datavetenskap vid Akademin för innovation, design och teknik kommer att offentligen försvaras onsdagen den 15 mars 2023, 13.30 i Delta och online via Zoom/Teams, Mälardalens högskola, Västerås.

Fakultetsopponent: Professor Philippe Palanque, University of Paul Sabatier, France.

Akademin för innovation, design och teknik

Abstract

In the last two to three decades, organizations have extremely changed. Post normal accident theory argues about implication of new organizational changes such as digitalization which may lead to new kinds of accidents called post normal accidents. In addition, there are technological changes such as usage of augmented reality (AR) as human-machine interface within various types of safety-critical systems. The organizational changes in addition to technological changes and their effects on human may introduce new system risks and should be considered in the risk assessment activities in compliance with related safety standards. Systems including technical entities and socio entities (i.e. humans and organizations) are called socio-technical systems. We consider socio-technical systems containing augmented reality which we call AR-equipped socio-technical systems. In order to adequately assess risk in such systems, it is essential to consider new dependability threats caused by augmented reality and new organizational changes. In the literature, various modeling and analysis techniques exist which are beneficial to be used for risk assessment. Furthermore, in the context of safety-critical systems, it is crucial to consider safety standards and assess the risk in compliance with the related safety standards.

This thesis aims at strengthening risk assessment in AR-equipped socio-technical systems in compliance with safety standards considering post normal accidents by providing a safety-centered risk assessment framework (we call it safety-centered due to the support it provides for safety standards). Our work provides modeling capabilities for modeling dependability threats caused by organizational changes leading to post normal accidents in AR-equipped socio-technical systems. The capabilities are provided through metamodel extensions. The extensions are used for extending analysis techniques to address the requirements of AR-equipped socio-technical systems analysis considering safety standard and post normal accidents. To achieve this, we capture dependability threats leading to post normal accidents via new modeling elements, which we add to SafeConcert, a conceptual metamodel for modeling socio-technical systems, and its AR-related extensions. The extended metamodel is then used to strengthen a risk analysis technique used for socio-technical systems analysis. We propose a dependability analysis process in order to analyze the behavior of AR-equipped socio-technical systems. Based on the modeling and analysis extensions, we propose a safety-centered framework for risk assessment of AR-equipped socio-technical systems and we apply it in two different domains. First, we conduct a case study in the automotive domain in cooperation with our industrial partner and we show how the required activities in the related safety standards are supported by different steps of our framework. Then, we use a digitalized socio-technical factory system in robotic domain containing both organizational and technological changes as a case in a new domain in order to evaluate applicability and effectiveness of our framework for capturing new kinds of accidents due to dependability threats caused by organizational changes and AR. Furthermore, we conduct a systematic literature review to position our contributions and to compare our work with other related works (we had preliminary literature reviews in the initial steps, nevertheless in this step we conduct a systematic literature review for positioning and comparing our work).

ISBN 978-91-7485-577-7
ISSN 1651-4238

Abstract

In the last two to three decades, organizations have extremely changed. Post normal accident theory argues about implication of new organizational changes such as digitalization which may lead to new kinds of accidents called post normal accidents. In addition, there are technological changes such as usage of augmented reality (AR) as human-machine interface within various types of safety-critical systems. The organizational changes in addition to technological changes and their effects on human may introduce new system risks and should be considered in the risk assessment activities in compliance with related safety standards. Systems including technical entities and socio entities (i.e. humans and organizations) are called socio-technical systems. We consider socio-technical systems containing augmented reality which we call AR-equipped socio-technical systems. In order to adequately assess risk in such systems, it is essential to consider new dependability threats caused by augmented reality and new organizational changes. In the literature, various modeling and analysis techniques exist which are beneficial to be used for risk assessment. Furthermore, in the context of safety-critical systems, it is crucial to consider safety standards and assess the risk in compliance with the related safety standards.

This thesis aims at strengthening risk assessment in AR-equipped socio-technical systems in compliance with safety standards considering post normal accidents by providing a safety-centered risk assessment framework (we call it safety-centered due to the support it provides for safety standards). Our work provides modeling capabilities for modeling dependability threats caused by organizational changes leading to post normal accidents in AR-equipped socio-technical systems. The capabilities are provided through metamodel extensions. The extensions are used for extending analysis techniques to address the requirements of AR-equipped socio-technical systems analysis considering safety standards and post normal accidents. To achieve this, we capture

dependability threats leading to post normal accidents via new modeling elements, which we add to SafeConcert, a conceptual metamodel for modeling socio-technical systems, and its AR-related extensions. The extended metamodel is then used to strengthen a risk analysis technique used for socio-technical systems analysis. We propose a dependability analysis process in order to analyze the behavior of AR-equipped socio-technical systems. Based on the modeling and analysis extensions, we propose a safety-centered framework for risk assessment of AR-equipped socio-technical systems and we apply it in two different domains. First, we conduct a case study in the automotive domain in cooperation with our industrial partner and we show how the required activities in the related safety standards are supported by different steps of our framework. Then, we use a digitalized socio-technical factory system in robotic domain containing both organizational and technological changes as a case in a new domain in order to evaluate applicability and effectiveness of our framework for capturing new kinds of accidents due to dependability threats caused by organizational changes and AR. Furthermore, we conduct a systematic literature review to position our contributions and to compare our work with other related works (we had preliminary literature reviews in the initial steps, nevertheless in this step we conduct a systematic literature review for positioning and comparing our work).

Populärvetenskaplig sammanfattning

Under de senaste två till tre decennierna har organisationer förändrats extremt mycket. Postnormal olycksteori handlar om hur nya organisatoriska förändringar, såsom digitalisering, kan leda till nya typer av olyckor som kallas postnormala olyckor. Det finns också tekniska förändringar, som exempelvis användandet av förstärkt verklighet (AR) som människa-maskin-gränssnitt inom olika typer av säkerhetskritiska system. De organisatoriska förändringarna, utöver tekniska förändringar och deras effekter på människor, kan introducera nya systemrisker och bör beaktas i riskbedömningsaktiviteterna i enlighet med relaterade säkerhetsstandarder. System som består av både tekniska enheter och sociala enheter (människor och organisationer) kallas sociotekniska system. Sociotekniska system som innehåller förstärkt verklighet kallar vi AR-utrustade sociotekniska system. För att på ett adekvat sätt kunna bedöma risker i sådana system är det viktigt att överväga nya pålitlighetshot orsakade av förstärkt verklighet och nya organisatoriska förändringar. I litteraturen finns olika modellerings- och analystekniker för riskbedömning. Vidare, i samband med säkerhetskritiska system, är det avgörande att ta hänsyn till säkerhetsstandarder och bedöma risken enligt dessa.

Denna avhandling syftar till att stärka riskbedömning i AR-utrustade sociotekniska system enligt säkerhetsstandarder som beaktar postnormala olyckor genom att tillhandahålla ett säkerhetscentrerat ramverk för riskbedömning (vi kallar det säkerhetscentrerat på grund av det stöd det ger för säkerhetsstandarder). Vårt arbete tillhandahåller modelleringsmöjligheter för att modellera pålitlighetshot orsakade av organisatoriska förändringar som leder till postnormala olyckor i AR-utrustade sociotekniska system. Dessa nya modelleringsmöjligheter tillhandahålls genom metamodelltillägg. Tilläggen används

för att utöka analystekniker för att möta kraven för AR-utrustade sociotekniska systemanalyser med hänsyn till säkerhetsstandarder och postnormala olyckor. För att uppnå detta identifierar vi pålitlighetshot som leder till postnormala olyckor genom nya modelleringselement, som utgör en utökning av SafeConcert, en konceptuell metamodell för modellering av sociotekniska system, och dess AR-relaterade tillägg. Den utökade metamodellen används sedan för att förstärka en riskanalysteknik som används för socioteknisk systemanalys. Vi föreslår en tillförlitlighetsanalysprocess för att analysera beteendet hos AR-utrustade sociotekniska system. Baserat på modellerings- och analysutökningarna föreslår vi ett säkerhetscentrerat ramverk för riskbedömning av AR-utrustade sociotekniska system och tillämpar det inom två olika domäner. Först genomför vi en fallstudie inom fordonsområdet i samarbete med vår industriella partner och visar hur de nödvändiga aktiviteterna i de relaterade säkerhetsstandarderna stöds av olika steg i vårt ramverk. Sedan använder vi ett digitaliserat sociotekniskt fabrikssystem i robotdomänen, som innehåller både organisatoriska och tekniska förändringar, för att utvärdera giltigheten och effektiviteten av vårt ramverk för att fånga nya typer av olyckor på grund av tillförlitlighetshot orsakade av organisatoriska förändringar och AR. Vidare genomför vi en systematisk litteraturstudie för att positionera våra bidrag och för att jämföra vårt arbete med andra relaterade verk (vi genomförde preliminära litteraturstudier i de inledande stegen, men i detta steg genomför vi en systematisk litteraturstudie för att positionera och jämföra vårt arbete).

To my family

Acknowledgments

First and foremost, I would like to express my immense gratitude to my main supervisor, Barbara Gallina. Thank you for your patience, guidance and support that have inspired me during my research. I also wish to express my gratitude to Karin Laumann and Martin Rasmussen Skogstad from NTNU University, Marko Vidović, Predrag Vidas, Davor Kovačec from Xylon Company, Atanas Gotchev, Robert Bregovic, Soili Pakarinen from Tampere University for the support and feedback during the ImmerSAFE project.

I also want to take the opportunity to be grateful with the head of our division, Radu Dobrin, as well as Federico Ciccozzi, Gunnar Widforss, Jan Carlson, Elisabeth Uhlemann and Muhammad Atif Javed for facilitating all the MDU routines. My gratitude is also for all the people, who are, or have been colleagues at MDU. In particular, I thank Julieth Castellanos, Irfan Sljivo and Zulqarnain Haider, for taking their time to answer my questions and sharing their knowledge. Special thanks to Cristina Seceleanu for reviewing my thesis and giving me valuable comments.

Above all, I thank God for helping me in my whole life, then I give special thanks to my parents for always believing in me, offering their most caring support and enthusiasm as well as my family and family-in-law for their inspiration and endless love during these years. Finally, and most important, I would like to express my gratitude and love to my husband Hamed and my son, Daniel. Words would not suffice to describe what you mean to me and the support and love you have given me. Your company, unconditional support and love have strengthened me through this challenging experience.

The work in this Ph.D. thesis has been supported by EU H2020 MSC-ITN grant agreement No 764951, via the project ImmerSAFE [1].

Soheila Sheikh Bahaei, January 2023, Västerås, Sweden

List of Publications

Papers Included in the Doctoral Thesis[1]

Paper A: *A Metamodel Extension to Capture Post Normal Accidents in AR-equipped Socio-technical Systems*, Soheila Sheikh Bahaei and Barbara Gallina. In Proceedings of the 31th European Safety and Reliability Conference (ESREL-2021), Research Publishing, Singapore, September 2021.

Paper B: *A Case Study for Risk Assessment in AR-equipped Socio-technical Systems*, Soheila Sheikh Bahaei, Barbara Gallina and Marko Vidović. Journal of Systems Architecture (JSA-2021), Elsevier, October 2021.

Paper C: *Towards Qualitative and Quantitative Dependability Analyses for AR-equipped Socio-technical Systems*, Soheila Sheikh Bahaei and Barbara Gallina. In Proceedings of the 5th International Conference on System Reliability and Safety (ICSRS-2021), IEEE, November 2021.

Paper D: *Technical Report on Risk Assessment of Safety-critical Socio-technical Systems: A Systematic Literature Review*, Soheila Sheikh Bahaei, Barbara Gallina. Technical Report, ISRN MDH-MRTC-345/2022-1-SE, Mälardalen Real-Time Research Center, Mälardalen University, December 2022.

Paper E: *Technical Report on Assessing Risk of AR and Organizational Changes Factors in Socio-technical Robotic Manufacturing*, Soheila Sheikh Bahaei, Barbara Gallina. Technical Report, ISRN MDH-MRTC-346/2022-1-SE, Mälardalen Real-Time Research Center, Mälardalen University, December 2022.

[1] The included papers have been reformatted to comply with the thesis layout

Additional Peer-reviewed Publications Related to the Thesis[2]

Paper 1: *Augmented Reality-extended Humans: Towards a Taxonomy of Failures - Focus on Visual Technologies*, Soheila Sheikh Bahaei and Barbara Gallina. In Proceedings of the 29th European Safety and Reliability Conference (ESREL-2019), Research Publishing, Singapore, September 2019.

Paper 2: *Towards Assessing Risk of Reality Augmented Safety-critical Socio-technical Systems* Soheila Sheikh Bahaei and Barbara Gallina. Published as Proceedings Annex in the 6th International Symposium on Model-Based Safety and Assessment (IMBSA-2019) website, Thessaloniki, Greece, October 2019.

Paper 3: *Effect of Augmented Reality on Faults Leading to Human Failures in Socio-technical Systems*, Soheila Sheikh Bahaei, Barbara Gallina, Karin Laumann and Martin Rasmussen Skogstad. In Proceedings of the 4th International Conference on System Reliability and Safety (ICSRS-2019), IEEE, November 2019.

Paper 4: *Extending SafeConcert for Modelling Augmented Reality-equipped Socio-technical Systems*, Soheila Sheikh Bahaei and Barbara Gallina. In Proceedings of the 4th International Conference on System Reliability and Safety (ICSRS-2019), IEEE, November 2019.

Paper 5: *A Case Study for Risk Assessment in AR-equipped Socio-technical Systems*, Soheila Sheikh Bahaei, Barbara Gallina and Marko Vidović. Technical Report, ISRN MDH-MRTC-332/2020-1-SE, Mälardalen Real-Time Research Center, Mälardalen University, May 2020.

Paper 6: *A Framework for Risk Assessment in Augmented Reality-equipped Socio-technical Systems* Soheila Sheikh Bahaei. Accepted at the Doctoral Forum hosted by the 50th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2020), IEEE, June 2020.

[2]These papers are not included in this thesis

Licentiate Thesis

A Framework for Risk Assessment in Augmented Reality-equipped Socio-technical Systems, Soheila Sheikh Bahaei. Mälardalen University Press, Licentiate Theses, ISSN 1651-9256; 293. May 2020.

Contents

I Thesis **1**

1 Introduction **3**
1.1 Thesis Outline 6

2 Background and Prior Work **11**
2.1 Fundamental Definitions for Risk Assessment and Dependability 11
2.2 Organizational Changes 16
2.2.1 Post Normal Accident Theory 16
2.2.2 Global Distance 17
2.3 Augmented Reality 18
2.4 Safety Standards 19
2.4.1 ISO 26262, SOTIF, SEooC and SAE 20
2.4.2 Robotic Safety Standards 23
2.5 Risk Assessment in AR-equipped Socio-technical Systems . . 27
2.5.1 modeling AR-equipped Socio-technical Systems . . . 27
2.5.2 Analyzing Socio-technical Systems 33
2.6 Goal Question Metric method 41

3 Research Summary **43**
3.1 Research Process 43
3.2 Problem Formulation 45
3.3 Research Goals 46

4 Thesis Contributions **49**
4.1 Metamodel Extension to Capture Post Normal Accidents . . . 49
4.1.1 Extracted Influencing Factors 50
4.1.2 Extended Modeling Elements 51

4.1.3 Potential Usage on an Example 52
4.2 Process for Dependability Analysis Based on Our Extensions . 55
4.2.1 Define Components and subcomponents 56
4.2.2 Define FPTC Rules for All Components 57
4.2.3 Create Error Models for the Component 57
4.2.4 Define Stochastic Behavior Parameters 57
4.2.5 Potential Usage on an Example 57
4.3 Risk Assessment Framework for AR-equipped Socio-technical Systems 64
4.4 Applying the Framework in Automotive Domain 66
4.4.1 Objectives of Case Study 66
4.4.2 Research Methodology of Case Study 66
4.4.3 Case Study Selection and Description 68
4.4.4 Case Study Execution: System Modeling 69
4.4.5 Case Study Execution: System Analysis 72
4.4.6 Compliance with ISO 26262 and SOTIF 84
4.5 Applying the Framework in Robotic Domain 86
4.5.1 Research Methodology 87
4.5.2 Planning the Study 88
4.5.3 Executing the Study 92
4.5.4 Discussion on the Results and Their Validity 103
4.6 Systematic Literature Review 111

5 Related Work **123**
5.1 Modeling Socio-technical Systems 123
5.2 Risk Analysis in Socio-technical Systems 125
5.3 Literature Reviews on Safety/Risk Analysis 127
5.4 Case Studies in Safety/Risk Analysis 129

6 Conclusion and Future Work **131**
6.1 Conclusions 131
6.2 Future Work 133

Bibliography **137**

II Included Papers **153**

7 Paper A:
A Metamodel Extension to Capture Post Normal Accidents in AR-

equipped Socio-technical Systems **155**
7.1 Introduction 157
7.2 Background 158
7.2.1 Modeling AR-equipped Socio-technical Systems . . . 158
7.2.2 Post Normal Accident Theory 159
7.2.3 Global Distance Metric 161
7.3 Proposed Extended Metamodel 162
7.3.1 Extracted Influencing Factors 162
7.3.2 Extended Modeling Elements 164
7.3.3 Potential Usage on an Example 166
7.4 Discussion 167
7.5 Conclusion and Future Work 168
Bibliography 169

8 Paper B:
A Case Study for Risk Assessment in AR-equipped Socio-technical Systems **173**
8.1 Introduction 175
8.2 Background 177
8.2.1 CHESS Framework 177
8.2.2 SafeConcert and its Extension of AR 177
8.2.3 The FPTC Syntax 181
8.2.4 Concerto-FLA Analysis Technique 182
8.2.5 ISO 26262, SOTIF, SEooC and SAE 182
8.3 An Integrated Framework for Assessing Risk of AR-equipped Socio-technical Systems 186
8.4 Case Study Design and Execution 189
8.4.1 Objectives 189
8.4.2 Research Methodology 189
8.4.3 Case Study Selection and Description 191
8.4.4 Case Study Execution: System Modeling 192
8.4.5 Case Study Execution: System Analysis 195
8.4.6 Compliance with ISO 26262 and SOTIF 207
8.4.7 Lessons Learnt 209
8.5 Threats to Validity 211
8.6 Discussion 212
8.7 Related Work 213
8.8 Conclusion and Future Work 215
Bibliography 217

9 Paper C:
Towards Qualitative and Quantitative Dependability Analysis for AR- equipped Socio-technical Systems 223
9.1 Introduction . . . 225
9.2 Background . . . 226
9.2.1 Metamodel extensions for AR-equipped Socio-technical Systems . . . 226
9.2.2 Toolchain for Automated Dependability Evaluation . . . 227
9.2.3 Synergy of Qualitative and Quantitative Dependability Analysis Techniques . . . 229
9.2.4 Analyzing Socio-technical systems . . . 231
9.3 Proposed Analysis Process . . . 232
9.3.1 Define Components and Sub-components . . . 234
9.3.2 Define FPTC rules for All Components . . . 234
9.3.3 Create Error Models for the Component . . . 235
9.3.4 Define Stochastic Behavior Parameters . . . 235
9.4 Case study . . . 235
9.4.1 Modeling the System . . . 236
9.4.2 Qualitative Analysis . . . 238
9.4.3 Quantitative Analysis . . . 240
9.5 Discussion . . . 242
9.6 Related Work . . . 243
9.7 Conclusion . . . 243
Bibliography . . . 245

10 Paper D:
Technical Report on Risk Assessment of Safety-critical Socio-technical Systems: A Systematic Literature Review 249
10.1 Introduction . . . 251
10.2 Background and Related Work . . . 253
10.2.1 Risk Assessment of Safety-critical Socio-technical Systems - Basic Concepts . . . 253
10.2.2 Related Work . . . 255
10.3 Research Method . . . 257
10.3.1 Planning the SLR . . . 258
10.3.2 Conducting the SLR . . . 265
10.4 Results and Analysis . . . 270
10.5 Discussion . . . 293
10.5.1 Discussion on the Results . . . 293

10.5.2 Threats to Validity 299
10.6 Conclusions and Future Work 300
Bibliography 303

11 Paper E:
Technical Report on Assessing Risk of AR and Organizational Changes Factors in Socio-technical Robotic Manufacturing 309
11.1 Introduction 311
11.2 Background and Related Work 313
11.2.1 Background 313
11.2.2 Related Work 320
11.3 Research Methodology 322
11.4 Planning the Study 323
11.4.1 Objectives 323
11.4.2 Selected Case 323
11.4.3 Study Protocol 326
11.5 Executing the Study 327
11.5.1 System Modeling 327
11.5.2 System Analysis 331
11.6 Discussion on the results and their validity 340
11.6.1 Discussion on the results 340
11.6.2 Discussion on the validity 344
11.7 Conclusion and Future Work 346
Bibliography 347

I

Thesis

Chapter 1

Introduction

Significant changes in organizations over the past twenty to thirty years besides increasing usage of new technologies such as AR are on one hand improving the system functioning, but on the other hand they may introduce new kinds of risk. The theory of post normal accident [2], which is an extension of normal accident theory [3], has highlighted the important changes in organizations over the last two to three decades. Based on this theory, technology and task are more digitalized and standardized in comparison to 1980s, which were automated. Organizational structures are more networked (externalized, horizontal) in comparison to 1980s, which were more integrated (internalized, vertical). In addition, organizational strategies are more financialized in comparison to 1980s, which were only industrial. Furthermore, environments are more globalized and self-regulated, while they were national and state regulated during 1980s. Effect of these changes on humans and organizations is not negligible and thus it is crucial to investigate it, since both human and organization take part as the socio entities of socio-technical systems. Furthermore, global distance metric [4] is a metric for capturing new influencing factors on human communication and collaboration. Global distance is defined as distances in geographical, temporal and cultural features of people working in an organization [5]. It is now well established that these new influencing factors affect on human and system performance [6]. There is a need to address these factors originated in recent changes by strengthening current risk assessment techniques along with considering related safety standards.

In order to perform safety-centered risk assessment in AR-equipped socio-technical systems, we review safety standards in addition to the risk assess-

ment techniques in the literature. Based on the standard ISO 31000:2018 [7], which is a general standard in risk management, *risk* is "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". According to ISO 26262 [8], the automotive standard for functional safety, risk assessment is a "method to identify and categorize hazardous events of items and to specify safety goals and ASILs (Automotive Safety Integrity Level) related to the prevention or mitigation of the associated hazards in order to avoid unreasonable risk". The focus in this standard is on risks emanated from malfunctions of electrical and/or electronic (E/E) system. In contrast, ISO 21448:2022 [9], defined as safety of the intended functionality (SOTIF), 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 may 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 [9].

Modeling the system entities and their behavior plays a vital role in risk assessment. UML (Unified Modeling Language)[1]-based metamodels are the most widely used groups of metamodels and have been extensively used for defining constructs required for modeling system entities and their important aspects. SafeConcert [10] is a conceptual metamodel proposed for modeling socio-technical systems. This metamodel is implemented as part of CHESS ML (CHESS Modeling Language) [11], which is a UML-based modeling language in CHESS framework [12]. Effects of Augmented Reality (AR) as a new technology on human and organizational factors and their modeling have been explored in our previous studies [13] [14].

So far, an integrated framework for risk assessment of AR-equipped socio-technical systems has not been proposed in compliance with safety standards considering post normal accidents. More specifically, there are no investigations into the effects of organizational changes leading to post normal accidents on modeling and analysis of system behavior in compliance with safety standards. Current frameworks do not contain modeling and analysis constructs for modeling and analyzing dependability threats leading to post normal accidents. In addition, there has been little investigation about modeling and analysis capabilities of current techniques for risk assessment of systems containing augmented reality and if the proposed assessment activities support safety stan-

[1] www.uml.org : accessed 2022-12-27

dards.

The objectives of this research are investigating the effects of the new organizational changes on modeling and updating available metamodels to enable capturing dependability threats leading to post normal accidents in AR-equipped socio-technical systems. In addition, we provide a process for dependability analysis to be used for risk assessment of AR-equipped socio-technical systems. Then, we propose a safety-centered risk assessment framework based on the modeling and analysis extensions and we validate it by conducting two case studies in two different domains. We use an industrial case study for verifying the modeling and analysis capabilities of the framework in capturing risks caused by AR-related dependability threats in automotive domain. Moreover, we apply our framework in socio-technical robotic manufacturing to demonstrate applicability and effectiveness of our contributions in a new domain with respect to considering effects of AR, organizational changes and provided support for safety standards. Finally, we explore literature and provide a positioning and comparison of our proposed framework with other related work through a systematic literature review.

More specifically, in the first step, we extract the new influencing factors on system behavior based on post normal accident theory and global distance metric, and we integrate these factors in the SafeConcert, which is a conceptual metamodel for modeling AR-equipped socio-technical systems.

In the second step, we propose an extension for a synergy of qualitative and quantitative dependability analysis technique. The extension is based on modeling extensions related to effects of AR and new organizational changes. It is also based on Concerto-FLA analysis technique [15], which is implemented as Eclipse plugin for dependability analysis and risk assessment in CHESS project [16]. We use SafeConcert metamodel and Concerto-FLA analysis technique, because they include constructs for integrating concepts of socio-technical systems. In order to include concepts related to effects of AR and organizational changes we propose modeling extensions to model dependability threats related to AR and organizational changes. Then, we provide a safety-centered risk assessment framework for AR-equipped socio-technical systems containing modeling and analysis phases in compliance with safety standards considering post normal accidents.

In the third step, we evaluate our contributions by applying our proposed framework in two different domains. We conduct a case study in automotive domain and we show how different steps of the framework can support required activities in ISO 26262 and SOTIF safety standards. In addition, we apply the framework in robotic domain to demonstrate the applicability and effectiveness

of our contributions in robotic domain with respect of effects of AR and organizational changes. We show how different steps of the framework can support required activities in robotic safety standards.

Finally, we conduct a systematic literature review and explore literature to position our contributions and to compare our contributions with other related studies.

We focus on AR-equipped socio-technical systems because of the context of the ImmerSAFE project [1] and also due to the increased AR applications. We use safety standards from automotive domain because most of our examples are from automotive domain. However, it is possible to use our contributions in other domains as we show it in the second application which is in robotic domain.

The resulting research efforts aim at planting the seeds for formal practices in the context of AR-equipped socio-technical risk assessment considering new post normal accidents which can be beneficial for development of tools that support safety analysts moving towards more comprehensive and finally automated risk assessments.

1.1 Thesis Outline

We organize this thesis in two parts. In the first part, we summarize the research as follows: In Chapter 2, we recall essential background and prior work. In Chapter 3, we present the research summary. In Chapter 4, we describe the specific research contributions of this thesis. In Chapter 5, we discuss related work. Finally, in Chapter 6, we present conclusions and future work. The second part is a collection of the papers included in this thesis. We present a brief overview of the included papers.

Paper A: ***A Metamodel Extension to Capture Post Normal Accidents in AR-equipped Socio-technical Systems***, Soheila Sheikh Bahaei, Barbara Gallina. In Proceedings of the 31th European Safety and Reliability Conference (ESREL), Research Publishing, Singapore, September 2021.

Abstract: In the past twenty to thirty years, organizations have extremely changed and these changes in addition to technological changes such as use of augmented reality (AR) introduce new system risks. Post normal accidents theory describes that organizations are more globalized and digitalized and are formed as networks of organizations, which would lead to post normal acci-

dents such as network failure accident. In addition, it states that strategies and organizational structures are more financialized and networked respectively and technology and task are more digitalized and standardized. These organizational factors affect also on human performance. Organization and human are considered as the socio parts of socio-technical systems. Metamodels should provide the modeling elements required for modeling human and organizational factors in new AR-equipped socio-technical systems. Current metamodels do not consider factors that would lead to post normal accidents. In this paper, we elaborate the theory of post normal accidents and we extract the influencing factors leading to post normal accidents. We also consider global distance including geographical, temporal and cultural distances, as an influencing factor on human performance. Then, we use the extracted influencing factors for extending modeling elements in our previously proposed conceptual metamodel for modeling AR-equipped socio-technical systems. Our proposed extended metalmodel can be used by analysis techniques in order to perform risk assessment for AR-equipped socio-technical systems.

My contribution: I was the main author of the paper under the supervision of the co-author. My specific contributions included extracting the influencing factors leading to post normal accidents and influencing factors based on global distance and using them for proposing new modeling elements. Both authors contributed equally in discussions and developing the paper contribution. The co-author contributed with reviews and comments for preparing the paper and suggestions for improvements.

Paper B: ***A Case Study for Risk Assessment in AR-equipped Socio-technical Systems***, Soheila Sheikh Bahaei, Barbara Gallina and Marko Vidović. Journal of Systems Architecture (JSA), Elsevier, October 2021.

Abstract: Augmented Reality (AR) technologies are used as human-machine interface within various types of safety-critical systems. Several studies have shown that AR improves human performance. However, the introduction of AR might introduce risks due to new types of dependability threats. In order to avoid unreasonable risk, it is required to detect new types of dependability threats (faults, errors, failures). In our previous work, we have designed extensions for the SafeConcert metamodel (a metamodel for modeling socio-technical systems) to capture AR-related dependability threats (focusing on faults and failures). Despite the availability of various modeling techniques, there has been no detailed investigation of providing an integrated framework

for risk assessment in AR-equipped socio-technical systems. Hence, in this paper, we provide an integrated framework based on our previously proposed extensions. In addition, in cooperation with our industrial partners, active in the automotive domain, we design and execute a case study. We aim at verifying the modeling and analysis capabilities of our framework and finding out if the proposed extensions are helpful in capturing system risks caused by new AR-related dependability threats. Our conducted qualitative analysis is based on the Concerto-FLA analysis technique, which is included in the CHESS toolset and targets socio-technical systems.

My contribution: I was the main author of the paper under the supervision of the second co-author. My specific contributions included preparing the framework, designing and executing the case study and writing the paper. The second co-author contributed to the design of the paper, provided reviews, comments for improving the paper and suggestions/ideas on how to accomplish the task. The third co-author contributed with reviews and comments for providing the model based on the system used in their company.

Paper C: ***Towards Qualitative and Quantitative Dependability analysis for AR-equipped Socio-technical Systems***, Soheila Sheikh Bahaei and Barbara Gallina. In Proceedings of the 5th International Conference on System Reliability and Safety (ICSRS), IEEE, November 2021.

Abstract: Augmented Reality technologies are becoming essential components in various socio-technical systems. New kinds of risks, however, may emerge if the concertation between AR, other technical components and socio-components is not properly designed. To do that, it is necessary to extend techniques for risk assessment to capture such new risks. This may require the extension of modeling languages and analysis techniques. In the literature, modeling languages have been already extended by including specific language constructs for socio aspects in relation to the AR-impact. No satisfying contribution is available regarding analysis techniques. Hence, to contribute to filling the gap, in this paper, we propose an extension of previously existing analysis techniques. Specifically, we build on top of the synergy of qualitative and quantitative dependability analysis techniques and we extend it with the capability of benefiting from AR-related modeled aspects. In addition, we apply our proposed extension to an illustrative example. Finally, we provide discussion and sketch future work.

My contribution: I was the main author of the paper under the supervision of the co-author. My specific contributions included providing the extension for analysis technique, applying it on an example and writing the paper. The co-author contributed by providing ideas on developing the paper contribution and by providing reviews and comments for improving the paper.

Paper D: ***Technical Report on Risk Assessment of Safety-critical Socio-technical Systems: A Systematic Literature Review***, Soheila Sheikh Bahaei and Barbara Gallina. Technical Report, ISRN MDH-MRTC-345/2022-1-SE, Mälardalen Real-Time Research Center, Mälardalen University, December 2022.

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/globalization, 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, AR effects, risk assessment process, their safety perspective, modeling formality, type of analysis, tool support, application domain and supported standards. Finally, we also provide our findings and possible future works based on the analysis of the primary studies, their potential applications and their challenges.

My contribution: I was the main author of the paper under the supervision of the second co-author. My specific contributions included conducting the systematic literature review and writing the paper. The second co-author contributed to the design of the paper, provided reviews, comments for improving the paper and suggestions/ideas on how to characterize the works and research questions.

Paper E: ***Technical Report on Assessing Risk of AR and Organizational Changes Factors in Socio-technical Robotic Manufacturing***, Soheila Sheikh Bahaei and Barbara Gallina. Technical Report, ISRN MDH-MRTC-346/2022-1-SE, Mälardalen Real-Time Research Center, Mälardalen University, December 2022.

Abstract: Technological changes such as the use of Augmented Reality (AR) along with the advent of new organizational changes such as digitalization are on the one hand positively changing the way of working but on the other hand they are introducing new risks, potentially leading to not only normal but also post-normal accidents. In our previous work, we have incrementally proposed a novel framework called FRAAR for risk assessment of AR-equipped socio-technical systems (i.e., systems integrating human, organisational and technical entities). We have also partly evaluated our framework via an industrial automotive case study and by providing comparison and positioning with respect to other related works in a systematic literature review. In this paper, we conduct a new study to evaluate the applicability and effectiveness of our framework in a different domain. To do that, we choose a digitalized socio-technical factory system, focusing on the human-robot collaboration for a realistic diesel engine assembly task using AR-based user interface. Finally, we discuss about validity of our work and we provide our findings and possible future works.

My contribution: I was the main author of the paper under the supervision of the second co-author. My specific contributions included conducting the study and writing the paper. The second co-author provided reviews, comments for improving the paper and suggestions/ideas on how to develop the paper contribution and how to evaluate the work.

Chapter 2

Background and Prior Work

This section introduces the background required by the current research, helping in the understanding of its content. Section 2.1 provides fundamental definitions for risk assessment and dependability. Section 2.2 presents organizational changes and Section 2.3 provides an overview of augmented reality. Then, Section 2.4 recalls safety standards. Section 2.5 provides an overview of risk assessment in AR-equipped socio-technical systems containing modeling AR-equipped socio-technical systems using AR-related extensions and analyzing system behavior. Finally, Section 2.6 presents Goal Question Metric approach (GQM), which is a method for measuring based on specific purpose.

2.1 Fundamental Definitions for Risk Assessment and Dependability

In this section, we provide some fundamental definitions for risk assessment and dependability that will be used during the research.

Based on the definition provided by Aven, risks are "consequences and uncertainties." [17] and risk analysis is a "tool for dealing with uncertainty". Lowrance defines risk as a measure of probability and severity of adverse effects [18]. Based on ICH (International Conference on Harmonization) guidelines [19], risk assessment consists of risk identification, risk analysis and risk evaluation. Risk identification deals with identifying the risks. Risk analysis deals with assigning likelihood and severity to identified risks and risk evaluation deals with comparing the identified and analyzed risks against risk criteria

to determine whether residual risk is tolerable. According to ISO 26262 [8] standard, risk is "combination of the probability of occurrence of harm and the severity of that harm". Risk assessment is a "method to identify and categorize hazardous events of items and to specify safety goals and ASILs (Automotive Safety Integrity Level) related to the prevention or mitigation of the associated hazards in order to avoid unreasonable risk". Based on society for risk analysis glossary [20], risk assessment is a "systematic process to comprehend the nature of risk, express and evaluate risk, with the available knowledge". Our study is based on ISO 31000: 2018 [7] standard, which is a generic approach and is not for a specific domain. Based on this standard, 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".

Risk assessment can be used for measuring dependability [21]. Based on dependability terminology provided by Avizienis et al. [21]:

- *System* is "an entity that interacts with other entities, i.e. other systems, including hardware, software, humans, and the physical world with its natural phenomena".
- *System function* is "what the system is intended to do".
- *Correct service* "is delivered when the service implements the system function".
- *Service failure* or failure is "an event representing a transition (a deviation) from correct service to incorrect service" (shown in Figure 2.1).
- *Human failure* is deviation from correct human function to incorrect human function.
- *Error* "is the part of the total state of the system that may lead to its subsequent service failure" (shown in Figure 2.1).
- *Fault* is "the adjudged or hypothesized cause of an error" (shown in Figure 2.1). It would be internal, if it is emanated from system itself or external, if it is emanated from other systems.
- *Failure mode* is a form in which a failure may manifest itself. In literature [22], service's failure modes have been categorized based on:
    1. *Provisioning*

- *Omission*: No output is provided.
- *Commission*: Output is provided when not expected.

2. *Timing*
   - *Early*: Output is provided too early.
   - *Late*: Output is provided too late.
3. *Value*
   - *Coarse*: The output is not within the expected range of values and user can detect this deviation.
   - *Subtle*: The output is not within the expected range of values and user cannot detect this deviation.

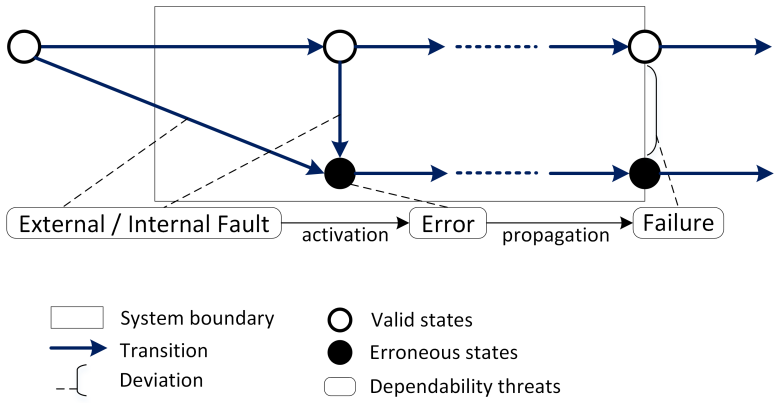


Figure 2.1: Causality chain among threats (adapted from [23])

Based on Avizienis et al. terminology [21], *dependability threats* are *faults, errors* and *failures* and as we explained, *fault* is cause of *error* and *error* is cause of *failure*. If there are service failures more frequently or more severely than acceptable, then this can be called failure in dependability. Based on definitions provided for risk and dependability threats, we can conclude that *dependability threats* are risk sources. *Hazard* is defined as "a circumstance which can lead to damage" [24] and standard ISO 12100:2010 [25], which is a standard for safety of machinery, defines hazard as "potential source of harm". For example, unexpected movement of a robot while collaborating with a human worker is a hazard. If the risk associated with the *hazard* is unacceptable then safety requirements should be defined. Thus, we can conclude

that *dependability threats* are hazard causes and hazard analysis is analysis of consequences of *dependability threats*, e.g. failures which are risk sources.

The causality chain of *dependability threats, hazard* and *harm* is shown in Figure 2.2. As it is shown in this figure, *dependability threats* can lead to *hazards*, which are associated with a specific *risk* and *hazard* can lead to *harm* (sometimes referred to as *accident*).

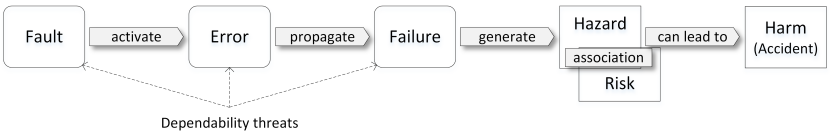


Figure 2.2: Relationships between dependability threats, hazard, risk and harm

Safety can be interpreted as "without unacceptable risk" and high level of safety means low level of risk [20]. 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 2.1. *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" [26]. 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" [26]. 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" [27]. In summary, based on [28], 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 [27], *Safety I* is not *reactive* as described in [26] 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 [27], *safety engineering today* is also introduced 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 is explained 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.

Table 2.1: Comparison between safety perspectives

Safety Perspective	**Definition**	**Defined by**	**Special focus on**	**Type of assumed causality**
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

Our perspective is more close to the *safety engineering today* perspective and the reason is that we consider linear chain of events and the root causes of an accident. However, we consider failures of technical components, human, organization and their interactions in the risk assessment process.

2.2 Organizational Changes

New organizational changes in the last two to three decades have been sources of new accidents which are called post normal accidents. In this subsection, we recall information about Post Normal Accident theory [2], which is an extension for Normal Accident (NA) theory [3]. In addition, we recall information about global distance metric and how it affects on system behavior.

2.2.1 Post Normal Accident Theory

Post normal accident theory [2], which is an extension for normal accident theory [3], is proposed by Jean-Christophe Le Coze. Perrow's normal accident theory argues that in tightly complex systems accidents are unavoidable or normal. Four analytical categories are also argued by Perrow to provide strong understanding of the situations which happen in organizations. These four categories are technology and task, structure, goal (later updated to strategy by Jean-Christophe Le Coze) and environment. Post normal accident theory argues that because of advent of new notions such as *globalization*, an update or adaptation for normal accident theory is required. In this theory, goal category is updated to strategy and features of the four categories (environment, strategy, structure, technology and task) are compared during 1980s and 2010s (Shown in Figure 2.3). Post normal accident theory, illustrates implications of trends such as *digitalization*, *standardization*, *financialization* and *self-regulation* on these four layers.

It discusses that environment was national and state regulated during the time normal accident theory was proposed (1980s). However, it is more *globalized* and *self-regulated* during 2010s. Based on definition provided in [29], *globalization* refers to "extended financial environment and greater exposure, worldwide competition, work and labour flexibility, incentives to breakdown vertical structures to gain flexibility through novel and expanding ICT networked infrastructure, normalized practices and dependence on a growing service activity (e.g. consulting)". *Self-regulation* refers to "industry regulating itself through the production of its own standards and internal control".

Strategy was more industrial during 1980s, while it is more *financialized* and industrial during 2010s. *Financialization* refers to "increasing the influence of financial actors (e.g. hedge funds) in companies' managerial decision-making processes".

Structure was more integrated during 1980s, while it is more networked during 2010s.

Analytical categories	Features based on NA (1980s)	Features based on Post NA (2010s)
Environment	*National & state regulated*	*Globalized & self-regulated*
Strategy	*Industrial*	*Financialized and Industrial*
Structure	*Integrated (internalized, vertical)*	*Networked (externalized, horizontal)*
Technology and task	*Automated*	*Digitalized & standardized*

Figure 2.3: Post normal accident theory [2]

Finally technology and task were more automated during 1980s, while they are more *digitalized* and *standardized* during 2010s. *Digitalization* refers to "the progressive replacement or extension of human activities by a combination of ICT systems and machines (or robots) which can perform an increasingly wide range of manual and cognitive tasks more and more independently". *Standardization* refers to "widespread management principles promoted by outsourcing and self-regulation, consulting firms and certification schemes for global markets".

As it is discussed in [29], recent changes introduce new safety challenges and besides their provided progress, they may be source of harm. It is also stated that looking into new categories of system risks is required as a complementary perspective for the study.

2.2.2 Global Distance

Global distance metric [4] has been suggested by Noll and Beecham, for global distance measurement between distributed sites on Global Software Development (GSD) [30]. Geographic, temporal and cultural distances are considered and quantified in this metric [5]. For example, for organization buildings in different countries a higher impact value is considered in comparison to buildings

in the same region or in the same campus. Similarly, for temporal and cultural distance different impact values are considered. It is also discussed that global distance would obstruct the communication among people in distributed teams.

In [31], an evaluation is designed to test cultural difference in understanding graphical symbols such as icons used in technological devices. US and Swedish subjects are evaluated and the results show that culture influences on their certainties for graphical symbol understanding. In [32], empirical evidence is provided showing that geographical proximity influences on social interactions and these effects even have increased by IT revolution. In [33], it is discussed that temporal distance influences on information diffusion processes in social and technological networks.

Based on these studies, global distance can be considered as an influencing factor on human performance. For example, a safety manager would live in a country with a culture that human safety is not so critical, while for another safety manager, it is highly critical based on the culture of the country he is living in. Thus, there would be some misunderstanding in discussions between these two people, if they work in two different buildings of a same organization located in different countries.

2.3 Augmented Reality

Augmented reality is any kind of extra information superimposed to reality and provided to user [34]. It would be visual, haptic, auditory, etc. For example, visual augmented reality refers to using graphics and digital content to juxtapose with what an individual is seeing in real-time [35]. However our research is not limited to visual augmented reality, we use visual augmented reality as an example throughout the research, because it is more apprehensible. AR displays can be categorized to three types including head-worn, hand-held and spatial. Head-worn displays are attached to the head, hand-held displays are displays that can be used by hand like mobile phones and spatial displays are placed in the environment like head-up displays (HUDs) [36]. HUD is “any transparent display that presents data without requiring users to look away from their usual viewpoint” [37]. For example, Figure 2.4 shows an example of using augmented reality information illustrating navigation information with the aim of increasing driving efficiency and driver reaction time.

Using augmented reality can improve user awareness and reaction time efficiency, meanwhile it can increase cognitive-processing or distract the driver [39], for example, if it covers important parts of the real world view of the

Figure 2.4: Using AR on head up display to show navigation information [38]

driver.

In [40], augmented reality is used in a driver simulator study with 88 participants and results show that visual warnings increase driver performance. Augmented reality can contribute to treatment of several mental and physical disorders [41] and for jobs with demanding situations and repetitive tasks, which threaten mental and physical health, AR can be used to upkeep mental and physical healthy state [42]. Neurological effects of AR, earned by brain-imaging technology show that brain cognitive activity increases and memory encoding is 70% higher while using AR [43]. AR integrates elements from virtual reality with elements from real world [44] leading to improvement in training by providing interactive ways for engaging learners and motivating them to have a better experience through the augmented environment [45].

Augmented reality may introduce new types of dependability threats. For example, if the expected improvement is not gained through AR because of distracting the user. AR affects on interpersonal communications and decreases social presence [46], which would lead to risk.

2.4 Safety Standards

IEC 61508 [47] is the primary functional safety standard for electrical, electronic and programmable electronic (E/E/PE) safety-related systems. Domain-specific standards are proposed for different domains based on IEC 61508 standard. For example, ISO 26262 [8] is proposed as an adaption of IEC 61508 for

automotive electric/electronic systems. ISO 10218 [48] is proposed as an adaption of IEC 61508 for Robots. Subsection 2.4.1 provides detail information about ISO 26262, SOTIF, SEooC and SAE, which are standards and taxonomy in automotive domain. Then, Subsection 2.4.2 provides detail information about safety standards and technical specification in robotic domain used in this thesis.

2.4.1 ISO 26262, SOTIF, SEooC and SAE

ISO 26262 [8] is a functional safety standard that provides the set of activities that should be performed during the safety lifecycle of safety-related systems. This standard specifies risk due to malfunctioning behavior of items. On the other hand, ISO/PAS 21448-SOTIF [9] is the standard for specifying risks due to other types of hazardous behavior related to functional insufficiencies of the intended functionality or its implementation. ISO 26262 and SOTIF addresses complementary aspects of safety. For example, the E/E random hardware faults are addressed in ISO 26262 and lack of driver attention while driving a car is addressed in SOTIF. In ISO 26262, ASIL (Automotive Safety Integrity Level) is determined and used for applying the requirements to avoid unreasonable residual risk. ASIL specifies item's necessary safety requirements to achieve an acceptable residual risk. Residual risks are remaining risks after using safety measures. An ASIL value is one of four levels (A-D) and it is determined based on three factors: severity, exposure and controllability. The severity factor indicates class of severity in case of hazard occurrence and it is classified from 0 to 3 (shown by S0-S3). S3 shows the category with the highest severity and it is related to situations with life threatening injuries. The exposure factor indicates class of probability of exposure with respect to operational situations and it is classified from 0 to 4 (shown by E0-E4). E4 shows the category with the highest probability of exposure (exposure duration more than 10% of average operating time). The controllability factor indicates the class of driver controllability and it is classified from 0 to 3 (shown by C0-C3). C3 shows the category with the highest controllability (more than 99% of drivers can control). ASIL classification based on these three factors is shown in Figure 2.5. QM (quality management) shows that no safety requirement is necessary. ASIL value A shows the lowest safety requirements and ASIL value D shows the highest safety requirements.

Safety element out of context (SEooC), introduced by ISO 26262, refers to an element that is not defined in the context of a special vehicle, but it can be used to make an item, which implements functions at vehicle level. For ex-

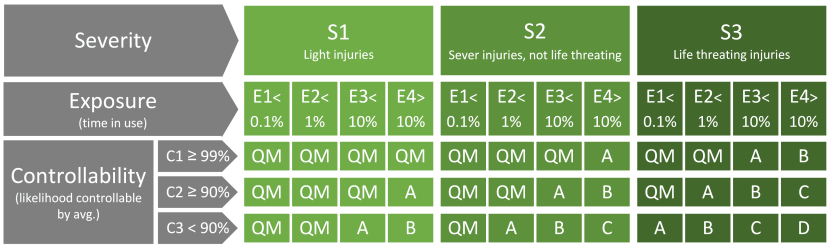


Figure 2.5: ASIL classification [49]

ample, system controller can be an SEooC. It can be a system, a combination of systems, a subsystem, a software/hardware component or a part. SEooC is based on ISO 26262 safety process and information regarding system context such as interactions and dependencies on the elements in the environment should be assumed [50].

The SEooC development contains 4 main steps:

1. (a) Definition of the SEooC scope: assumptions related to the scope, functionalities and external interfaces of the SEooC should be defined.

   (b) Definition of the assumptions on safety requirements for the SEooC: assumptions related to item definition, safety goals of the item and functional safety requirements related to SEooC functionality, which are required for defining technical safety requirements of the SEooC should be defined.

2. Development of SEooC: based on the assumed functional safety requirements, technical safety requirements are derived and then SEooC is developed based on ISO 26262 standard.

3. Providing work products: work products are documents that show the fulfilled functional safety requirements and assumptions on the context of SEooC.

4. Integration of the SEooC into the item: safety goals and functional safety requirements defined in item development should match with assumed functional safety requirements for the SEooC. In case of a SEooC assumption mismatch, change management activity based on ISO 26262 standard should be conducted.

The process required for improving the intended functionality to ensure safety includes eight activities. Possible interactions between these activities and ISO 26262 activities and SEooC are shown in Figure 2.6.

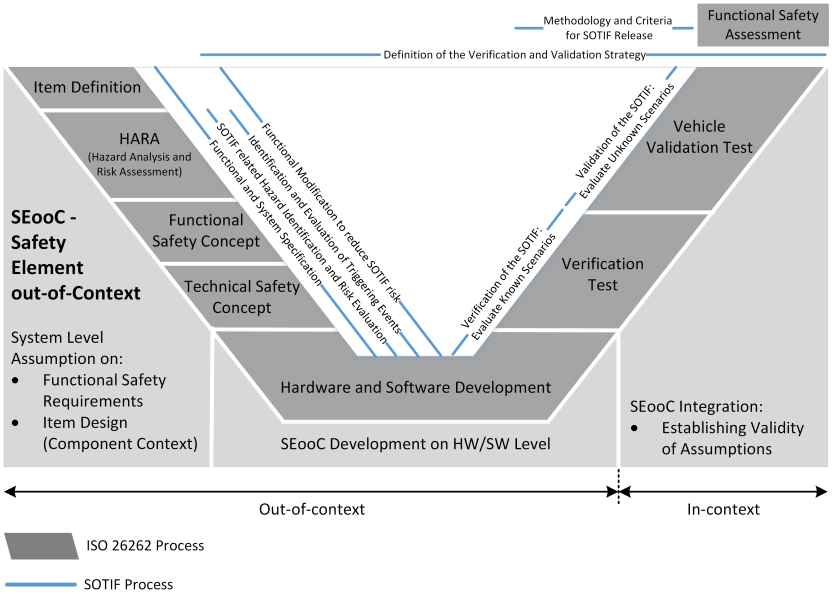


Figure 2.6: Alignment of SOTIF activities to ISO 26262 activities and SEooC (adapted from [50] and [51])

Safety process of the ISO 26262 standard starts with *concept phase* containing *item definition*, *hazard analysis and risk assessment* and *functional safety concept* [50]. An *item* implements a vehicle level function. In *item definition* the main objective is defining items. Defining items requires defining the dependencies and interactions with environment. Then, related hazards should be identified and functional safety requirements should be obtained. In SEooC, assumptions related to the system context are the main output of the *concept phase*. *Functional safety concept* includes providing functional safety requirements. Output provided by *Functional safety concept* is used by *technical safety concept*. *Technical safety concept* includes defining technical safety requirements and system design. Then, *hardware and software development* should be done based on *technical safety concept*. *HW/SW development* can

be done based on assumptions provided in concept phase. Next steps in the process are *verification test*, *validation test* and *functional safety assessment*. In SEooC, these steps require establishing validity of assumptions.

SOTIF process starts with *functional and system specification*. In this step the main objective is providing functional description and considerations on system design and architecture. Then, potential hazardous events should be identified in *SOTIF related hazard identification and risk evaluation*. If the harm is possible for the identified potentially hazardous events, then analysis of their triggering events should be conducted (*identification and evaluation of triggering events*). Functional modification is the next activity for avoiding the hazards or for reducing the resulting risk (*functional modification to reduce SOTIF risk*). Next activities are verification and validation strategy specification (*definition of the verification and validation strategy*) and then in verification and validation activities arguments are provided to illustrate that the residual risk is below acceptable level by testing on various known and unknown scenarios (*verification of SOTIF*,*validation of SOTIF*). Finally, SOTIF activities should be reviewed and evaluation on residual risk should be performed based on the verification and validation results and specified criteria (*Methodology and criteria for SOTIF release*).

Based on the taxonomy and definitions related to driving automation systems for on-road motor vehicles performing part or the entire dynamic driving task (DDT) on a sustained basis, there are six levels of driving automation. SAE level 0 refers to no driving automation and SAE level 5 refers to full driving automation [52]. These levels with description and example are shown in Figure 2.7. Assessing human factor in driver-vehicle interface is not only important on lower SAE levels, but also on higher levels because of the importance of safe transition between automated and non-automated vehicle operation [53]. In order to improve safety, various scenarios of driver/vehicle interaction should be considered.

2.4.2 Robotic Safety Standards

There are five main relevant standards and technical specification for risk assessment in human robot collaboration domain:

- ISO 12100:2010, Safety of machinery - General principles for design - Risk assessment and risk reduction [25]
- ISO 10218-1:2011, Robots and robotic devices – Safety requirements for industrial robots – Part 1: Robots [48]

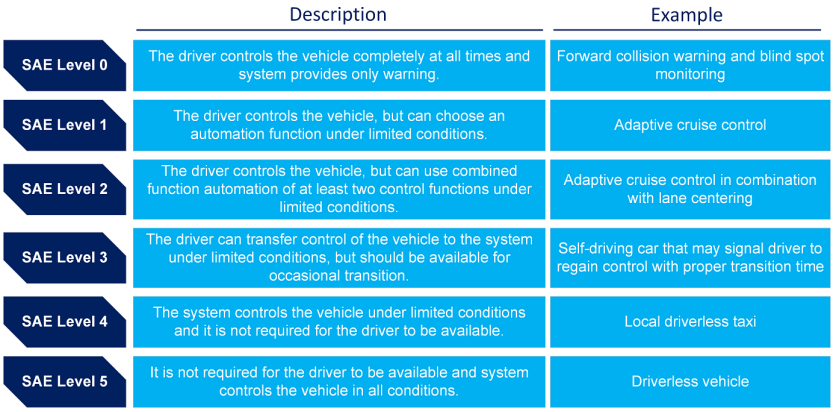


Figure 2.7: SAE levels of driving automation [54]

- ISO 10218-2:2011, Robots and robotic devices – Safety requirements for industrial robots – Part 2: Robot systems and integration [55]
- ISO/TS 15066:2016, Robots and robotic devices - Collaborative robots [56]
- ISO 13849-1:2015, Safety of machinery – Safety-related parts of control systems - Part 1: General principles for design [57]

Based on standard ISO 12100:2010 [25] risk is “combination of the probability of occurrence of harm and the severity of that harm”. Severity of the harm (S) is classified as S1 (for occasions with slight injuries which are reversible) and S2 (for occasions with serious injuries or death which are irreversible). Probability of occurrence of harm (P) is classified as P1 for occasions where there is chance of avoidance or significant decrement in effects, otherwise it is classified as P2. Based on standard ISO 13849-1:2015 [57], safety-related PLr (required performance level) is determined based on severity of injury (S), possibility of avoiding or limiting harm and probability of occurrence (P) and frequency and/or exposure to hazard (F). Frequency and/or exposure to hazard is classified as F1 for occasions with exposure time less than or equal to 1/20 of overall operating time or frequency of less than or equal to once per 15 min, otherwise it is classified as F2. Determining the required performance level is shown in Figure 2.8.

Standard ISO 10218-1:2011 [48], provides guidelines and requirements for design, measures and use of industrial robots. Basic hazards are recognized for

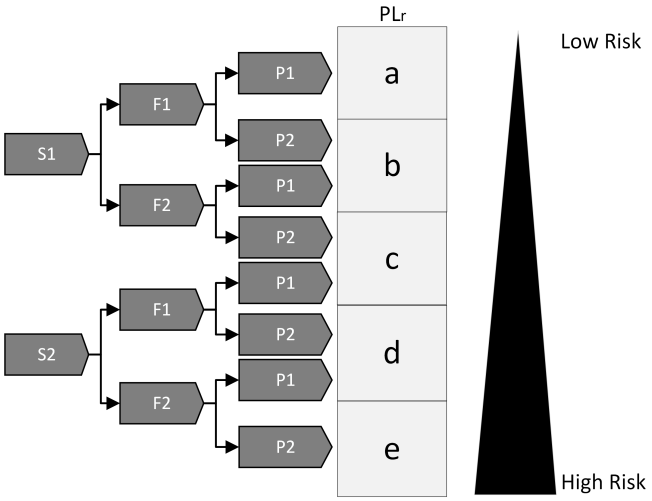


Figure 2.8: Determining required performance level based on [57]

industrial robots and industrial robot systems. However, it is discussed that the numbers and types of hazards are different for various kinds of robots with different automation process and installation complexity. In addition, the sources of the hazards are specific for each particular robot. Standard ISO 10218-2:2011 [55], which is complementary part of ISO 10218-1:2011 specifies the requirements for robot systems, integration and their installation. It also contains significant hazards for robot and robot systems. However, other hazards for specific applications must be addressed based on individual basis.

Based on technical specification ISO/TS 15066:2016 [56] collaborative operation means "state in which a purposely designed robot system and an operator work within a collaborative workspace". The aim of using collaborative robots is to integrate the competencies of robots such as repetitive performance, precision, power and endurance with the skills and abilities of human. Traditional applications prevented human intervention during the robot activity and it caused lower speed and not being able to automate some operations. In order to have collaboration between human and robot operations, it is essential to consider safety related issues and assess the risk during the collaboration.

Based on standard ISO 12100:2010 [25], risk assessment is the process containing risk analysis and risk evaluation. Risk analysis is the process con-

taining defining the limits of the machine, identifying hazards and estimating the risk. Risk evaluation is "judgment, on the basis of risk analysis, of whether the risk reduction objectives have been achieved". This process is more extended in ISO 10218-2: 2011 by considering robot system which contains industrial robot, end-effector(s) and any supporting machinery, equipment or sensors. In addition, task identification is considered during the risk assessment process to determine the potential occurrence of hazardous situations. Finally, in ISO/TS 15066:2016 the risk assessment is defined containing the following actions:

- Risk analysis
  - Determining the limits of the robot system (intended use and foreseeable misuse)
  - Identifying the hazards and associated hazardous situations
    - considering robot related hazards
    - considering hazards related to the robot system
    - considering application related hazards
    - identifying tasks
  - Estimating the risk of each hazard and hazardous situation
- Risk evaluation
  - Evaluating the risk and taking decision about necessity of reducing the risk based on risk analysis results

In traditional robot system installations, it was not possible for human to work in close proximity to robots unless the power of the robot was disconnected. Since in human robot collaboration they can operate in the same workspace while the power of the robot is connected, it is of high importance to take into account potential hazards and their related risk. Technical measures for risk reduction are based on main principles defined in ISO/TS 15066:2016: 1) hazard elimination by design or hazard reduction by substitution. 2) preventing the human to face the hazards or providing a safe state before human come to the hazardous situation, 3) risk reduction during the interventions.

2.5 Risk Assessment in AR-equipped Socio-technical Systems

In order to assess risk in AR-equipped socio-technical systems, it is crucial to model the system and analyze its behavior based on the provided model. Subsection 2.5.1 contains essential background related to modeling socio-technical systems and AR-related extensions. Then, Subsection 2.5.2 recalls essential background related to analyzing socio-technical systems.

2.5.1 modeling AR-equipped Socio-technical Systems

There are different modeling languages in the literature for system modeling to be used for risk assessment by proposing UML extensions. EAST-ADL2 [58] extends UML and SysML (System Modeling Language) [59] and provides modeling language for automotive domain. DAM (Dependability Analysis Modeling) [60] also provides dependability modeling on UML profile, which is coupled with MARTE (Modeling and Analysis of Real Time and Embedded systems) [61]. We base our work on SafeConcert metamodel [10] because of the support this metamodel provides for modeling socio-technical systems and also because it is integrated within the AMASS platform [62], the first open-source platform for supporting engineering and certification of safety-critical systems [63].

SafeConcert [10] is a conceptual metamodel for modeling socio and technical entities in socio-technical systems. This metamodel is implemented as part of CHESS ML (CHESS Modeling Language) [11], which is a UML-based modeling language used in CHESS framework [64]. In SafeConcert metamodel, software, hardware and socio entities can be modeled as components in component-based systems representing socio-technical systems. SERA taxonomy [65] is used for modeling human and organization, which are the socio entities of the system.

In [13], extensions are proposed for this metamodel in order to incorporate AR related factors. As it is shown in Figure 2.9, AR-equipped socio-technical system is a system which has augmented reality technology in addition to usual socio and technical entities. This technology affects on human and organization. Human using augmented reality would have extended capabilities, which are required to be modeled in order to consider their failure behavior while doing risk assessment. For example, with the use of augmented reality a person can sense surrounding environment, thus surround sensing is an AR-extended characteristic for human.

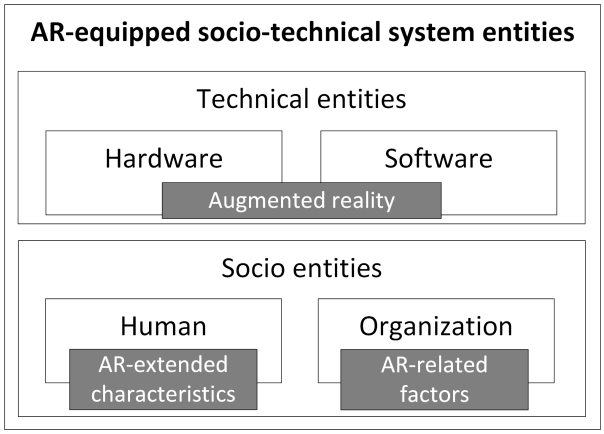


Figure 2.9: AR-equipped socio-technical system entities [13]

As it is shown in Figure 2.10, entities, their characteristics and their relations are modeled using components, subcomponents and connectors. subcomponents of human and organization are selected based on SafeConcert human and organization modeling elements and AR-related modeling extensions. The factors with gray color are the conceptual extensions. Organizational factors are based on several state-of-the-art taxonomies such as Rasmussen [66], HFACS [67], SERA [65] and SPAR-H [68] and AR-related factors are added based on studies and experiments on AR such as [44] and [45].

In [70], the human modeling elements are extended based on AREXTax, which is an AR-extended human function taxonomy [71]. This taxonomy is obtained by harmonizing about six state-of-the-art human failure taxonomies (Norman [72], Reason [73], Rasmussen [66], HFACS (Human Factor Analysis and Classification System) [67], SERA (Systematic Error and Risk Analysis) [65], Driving [74]) and then extending the taxonomy based on various studies and experiments on augmented reality. These extended modeling elements are divided to four categories, shown in Figure 2.11. Three of these categories are human functions including *human process unit, human SA (situational awareness) unit,* and *human actuator unit*. The one other category is *human fault unit*, which is related to human internal influencing factors affecting on human functions. We explain these modeling elements in the next two paragraphs. In the first paragraph we explain modeling elements related to human functions and in the second paragraph we explain modeling elements related to *human*

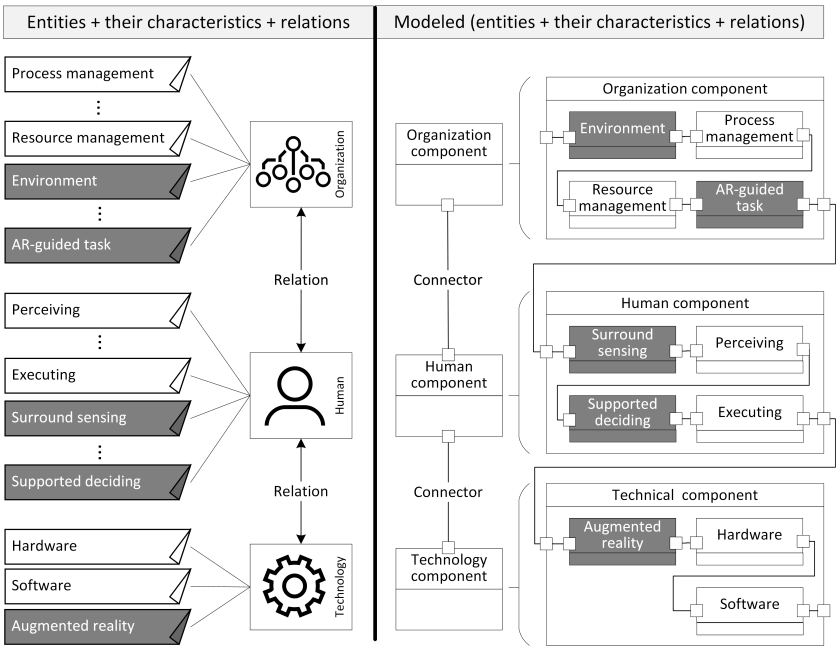


Figure 2.10: AR-equipped socio-technical system modeling [69]

fault unit and also other fault categories. Extended modeling elements are shown with white color and AR-stemmed modeling elements are shown with dotted line border.

The extended modeling elements in *human process unit, human SA unit*, and *human actuator unit* enable modeling of AR-extended human functions. For example, detection failure, which represents a failure in *detecting* human function, is a human failure introduced by several human failure taxonomies such as Reason [73] and Rasmussen[66] taxonomies. Based on experiments and studies on augmented reality including [75] and [76], *detecting* function can be extended to *surround detecting* while using AR (surrounding information would be augmented on real world view of the user by AR). Thus, *surround detecting* can be considered as an extended subcomponent of human component; in other words *surround detecting* is an extended modeling element proposed to be used for modeling and analyzing AR-equipped socio-technical systems.

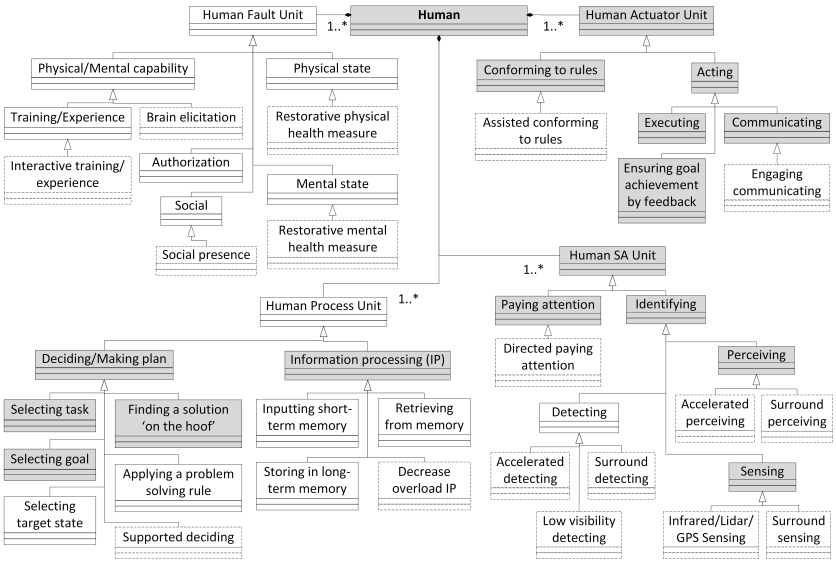


Figure 2.11: Extended human modeling elements [13]

In [13], organization and human modeling elements are extended based on AREFTax, which is a fault taxonomy including AR-caused faults [38]. This taxonomy is obtained by harmonizing about five state-of-the-art fault taxonomies (Rasmussen [66], HFACS [67], SERA [65], Driving [74] and SPAR-H (Standardized Plant Analysis Risk Human Reliability Analysis)[68]) and then extending the taxonomy based on various studies and experiments on augmented reality. These extended modeling elements are shown in Figure 2.12 and human fault unit of Figure 2.11. Extended modeling elements are shown with white color and AR-stemmed modeling elements are shown with dotted line border. These extended modeling elements enable modeling of various faults leading to human failures including AR-caused faults. Faults would be caused by human, environment, organization, etc. Human related faults are categorized as human fault unit of Figure 2.11 and non-human faults are categorized as three categories of organizational factors including organization and regulation unit, environment unit and task unit. For example, failure in *physical state* of a human is a human internal fault leading to human failure. This is shown as an extended human modeling element in human fault unit category shown in Figure 2.11. Another example is *condition*, which is a non-human

factor and it is categorized as extended modeling elements for organization components shown in Figure 2.12. One example of the AR-extended modeling elements is *social presence* shown in Figure 2.11. Based on studies on augmented reality [46], using AR would decrease social presence and failure in *social presence* can be considered as fault leading to human failure.

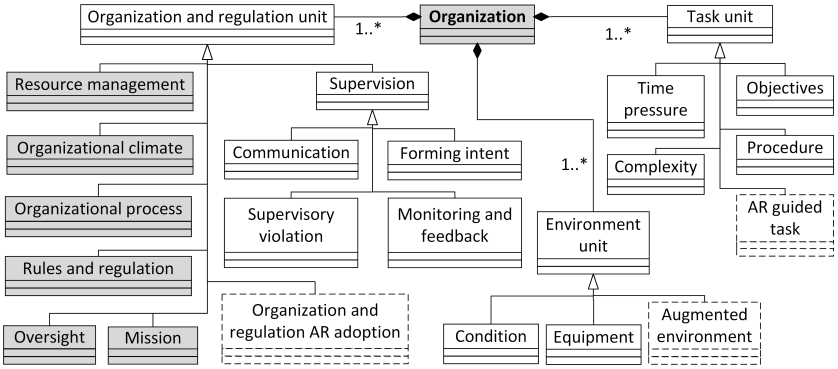


Figure 2.12: Extended organizational modeling elements [13]

We use the extended SafeConcert for modeling the example shown in Figure 2.4. We can consider three composite components including human component, organization component (road transport organization) and AR-HUD component. We consider organization component also to take into account effect of organizational factors (such as environmental factors). Organizational factors influencing human functioning are selected from extended SafeConcert organization modeling elements shown in Figure 2.12. Our selected elements are:

- Organization and regulation AR adoption: it refers to upgrading rules and regulations of road transport organization based on AR technology [77].
- Condition: it refers to road condition.
- AR guided task: it refers to the task, which AR is used for guiding driver to do that [78]. For example, if AR is used to guide driver to park the car more safely, parking safely is the AR-guided task.

Organization component is affected by influences from regulation authorities. In order to model these influences, we consider an input for the organization component connected to system input (shown by REG in Figure 2.13).

We consider four subcomponents for human composite component selected from extended SafeConcert human modeling elements shown in Figure 2.11. These four subcomponents are:

- Surround detecting: it refers to an AR-extended function, because driver can detect surround environment through AR technology.
- Deciding: it refers to human decision making function.
- Executing: it refers to human executing function.
- Social presence: it refers to an AR-caused factor, because AR may decrease social presence and lead to human failure.

Surround detecting affects on deciding and deciding affects on executing. Social presence input is connected to system input with the name human communication input (HCI) and affects on human executing. Human output, which is output of the system is human function shown by HF.

An AR-HUD component contains three primary subcomponents [37]:

- Projector unit: it refers to the subcomponent that produces an image on a combiner.
- Combiner: it refers to the subcomponent that is a flat piece of glass and can be the windshield of the car.
- Computer: it refers to the subcomponent that generates the information that should be displayed by projector unit.

Another system input, which is input of the computer subcomponent is raw data (RD) provided by sensors.

We explain about three scenarios depicted in Figure 2.13. AR-extended function and AR-caused factors are shown by gray color, to show the effect of AR and the contribution of the proposed modeling elements.

In the first scenario (S1), content provided by AR-HUD is wrong and it leads to the driver's failure. For example, there is failure in combiner of AR-HUD, which is a technical component. This failure is an external fault for human component and causes system failure.

In the second scenario (S2), content provided by AR-HUD is correct, but there is failure in organization and regulation AR adoption, which is an external fault for human component. For example, when the organization updates rules and regulation based on AR requirements. This leads to failure in organization

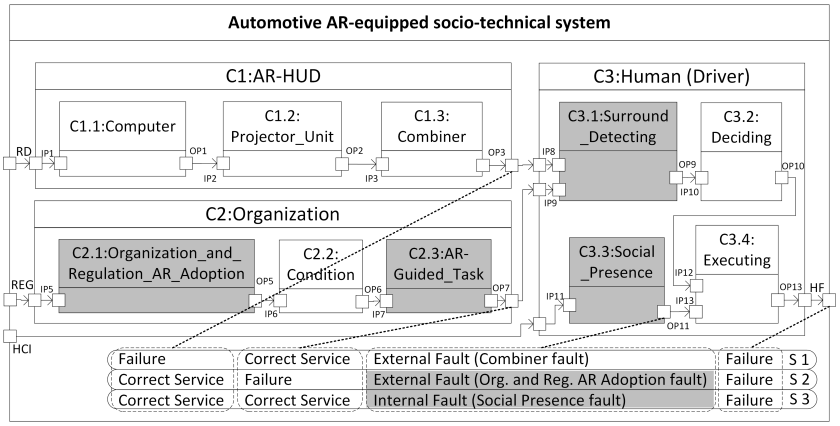


Figure 2.13: Using extended SafeConcert for modeling an AR-equipped socio-technical system

output. This failure is also an external fault for human component and causes system failure.

In the third scenario (S3), there is failure in social presence subcomponent of the driver component, which is an internal fault leading to failure in executing subcomponent and leading to system failure. For example, driver would miss the common ground with other people, this failure would lead to wrong action. Failure in social presence is an internal fault for human component and would lead to system failure.

As it is shown in this example, the proposed extended modeling elements can be used for enhancing modeling of internal and external faults leading to human failures, used in risk analysis tools.

2.5.2 Analyzing Socio-technical Systems

In socio-technical systems the output is the result of human and technology interaction embedded within social structures such as organizational goals and environmental aspects. Standard techniques in risk analysis such as Fault Tree Analysis (FTA) [79], Failure Modes and Effects Analysis (FMEA) [80], formal methods and probabilistic safety analysis are not sufficient [81]. The problem with traditional methods such as FTA and FMEA is that they should be done manually, which requires a huge amount of time and work for recent com-

plicated systems. Model-driven risk analysis techniques such as Fault Propagation and Transformation Calculus (FPTC) [82], Failure Propagation and Transformation Analysis (FPTA) [83], Hierarchically Performed Hazard and Operability Studies (HiP-HOPS) [84], CHESS-FLA [85] (Failure Logic Analysis within the CHESS project [16]) and Concerto-FLA [15] (Failure Logic Analysis within the Concerto project [86]) are developed based on traditional methods to automatically provide FTA and FMEA results based on system architecture and modeling of components failure behavior.

In our research, we use Concerto-FLA analysis technique as a qualitative technique containing socio aspects and we also use a synergy of qualitative and quantitative dependability analysis techniques in order to incorporate quantitative analysis. We explain these two approaches in the following subsections.

Concerto-FLA analysis technique

Concerto-FLA [15] is a model-based analysis technique that provides the possibility for analyzing failure behavior of humans and organizations in addition to technical entities by using SERA [65] classification of socio-failures. This approach is provided as a plugin within the CHESS toolset and allows users to define component-based architectural models composed of hardware, software, human and organization. For each component, FPTC (Failure Propagation Transformation Calculus) [82] rules (logical expressions that relate output failures to input failures) are used to model a component's failure behavior.

FPTC syntax for modeling failure behavior at component and connector level is as follows:

behavior = expression+
expression = LHS '->' RHS
LHS = portname'.' bL | portname '.' bL (',' portname '.' bL) +
RHS = portname'.' bR | portname '.' bR (',' portname '.' bR) +
failure = 'early' | 'late' | 'commission' | 'omission' | 'valueSubtle' | 'valueCoarse'
bL = 'wildcard' | bR
bR = 'noFailure' | failure

Failure used in this syntax is the form in which a failure may manifest itself, which is called failure mode based on dependability terminology provided by Avizienis et al. [21] (explained in Subsection 2.1).

Wildcard in an input port shows that the output behavior is the same re-

gardless of the failure mode on this input port. noFailure in an input port shows normal behavior.

Components' behavior can be classified as source (if component generates a failure), sink (if component is able to detect and correct input failure), propagational (if component propagates failures received in its input to its output) and transformational (if component transforms the type of failure received in its input to another type in its output) [87].

Based on this syntax, "IP1.noFailure $\rightarrow$ OP1.omission" shows a source behavior and should be read as follows: if the component receives noFailure (normal behavior) on its input port IP1, it generates omission on its output port OP1.

Concerto-FLA analysis technique, which uses FPTC syntax includes five main steps.

1. Modeling architectural elements including software, hardware, human, organization, connectors, interfaces and etc.

2. Using FPTC syntactical rules to model failure behavior at component and connector level. Concerto-FLA has adopted FPTC syntax for modeling failure behavior at component and connector level.

3. Modeling failure modes at system level by injection of inputs.

4. Performing qualitative analysis through automatic calculation of the failure propagations. This step is similar to FPTC technique that system architecture is considered as a token-passing network and set of possible failures that would be propagated along a connection is called tokenset (default value for each tokenset is noFailure, which means normal behavior). In order to obtain system behavior, maximal tokenset is calculated for each connection through a fixed-point calculation.

5. Interpreting the results at system level. Based on the interpretation, decision for changing system design would be taken.

We show these steps by providing an example to clarify how the technique works. HUD system explained in Subsection 2.3 is used as an example.

1. In the first step of Concerto-FLA technique, model of architectural elements should be provided. Based on system description, HUD and human are considered as two composite components of the system. Architectural model of the system using SafeConcert modeling elements

is shown in Figure 2.14. HUD is composed of three main elements: combiner, projector unit and computer. Combiner is any transparent display for illustrating AR information. AR information is projected on the combiner by projector unit and is produced by computer [37]. Computer receives raw data from sensors. Because of the presence of AR technology, we call the composite component AR-HUD. To model human composite component, three human modeling elements are selected, including HSPerception, HAKnowledgeDecision and HAResponse, which are modeled as three subcomponents of human composite component. Output of the HAResponse component is output of the system, which is shown by human function.

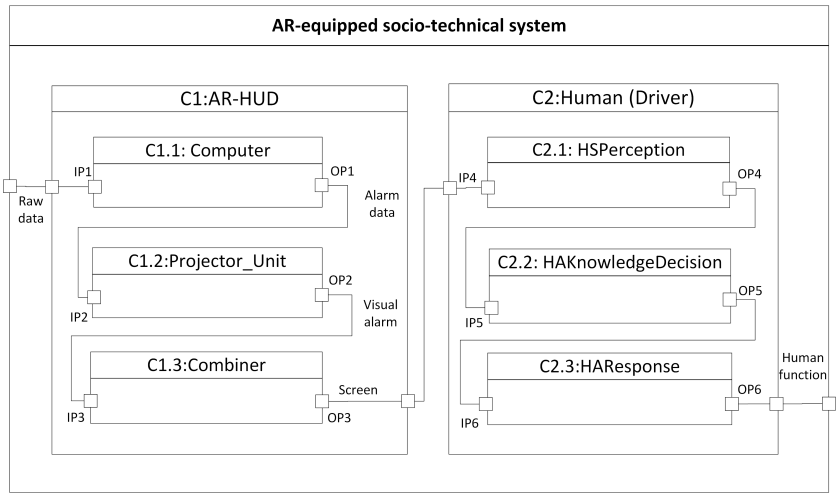


Figure 2.14: Concerto-FLA modeling for AR-HUD example

2. In the second step, failure behaviors of each component should be provided using FPTC rules, which are based on studying each component in isolation. Incoming and outgoing failures can be classified by related domain failure categorization. For example, timing, value, commission and omission failures are considered in this approach. "IP1.noFailure → OP1.noFailure" behavior shows that if there is normal behavior in input of computer subcomponent, then there is normal behavior in its output. Some sample rules for subcomponents are shown in Figure 2.15.

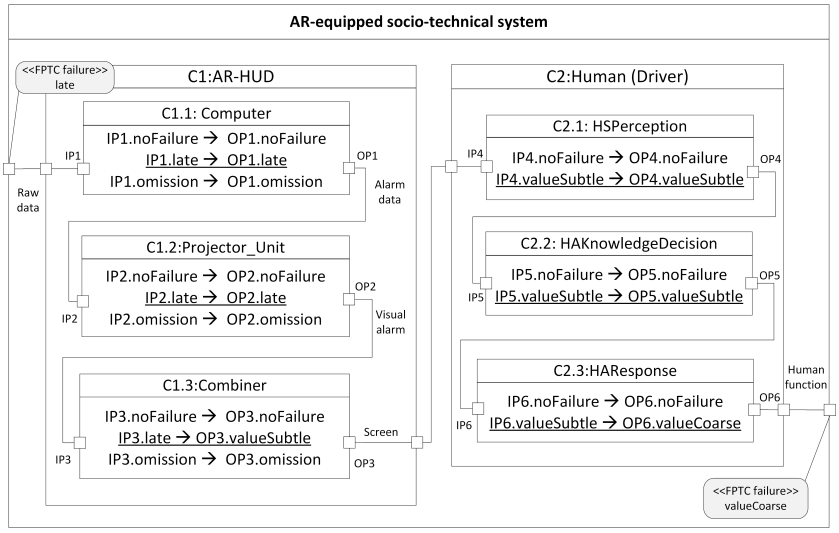


Figure 2.15: Concerto-FLA modeling and analysis results on AR-HUD example

3. In the third step, we assume that the raw data is provided late by sensor and late will be considered as the input failure for computer subcomponent (IP1 in Figure 2.15).

4. In the fourth step, we calculate the failure propagations, which is shown in Figure 2.16. Based on the analysis algorithm provided in Concerto-FLA technique, each subcomponent is considered as a point and default tokenset is assigned to all connections between subcomponents. Tokenset for each connection is defined with a noFailure token. Then, maximal tokenset is provided based on FPTC expressions and by comparing input failure mode with left hand side of the FPTC expressions. Right hand side of the matched expressions will be added to tokenset of the outgoing connection. For example, possible failure modes for IP1 are noFailure and late (noFailure is the default failure mode for all connections and late is shown in the picture as the possible input failure mode). Based on the FPTC expressions in computer component, noFailure and late match with left hand side of the first two expressions, thus their right hand side will be added to IP2 tokenset. The failure propagation is cal-

culated for all connections and maximal tokenset is calculated (shown in Figure 2.16). The failure propagation leads to valueCoarse failure in system output. This step is done automatically in CHESS toolset.

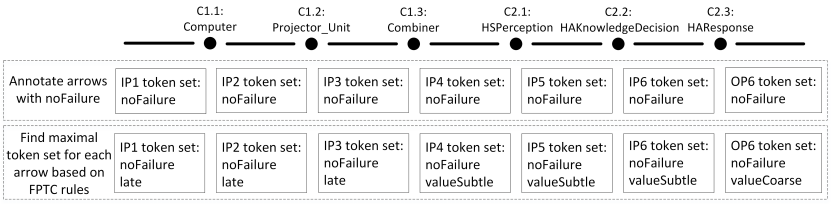


Figure 2.16: Concerto-FLA analysis on AR-HUD example

5. Finally, in the last step results can be interpreted. ValueCoarse on OP6 is because of valueSubtle on IP6 that is because of valueSubtle on IP5 and we continue this back propagation to find the origin of the failure that is late on IP1 in this case (shown in Figure 2.17). By using this method, it is possible to find the effect of components' failure behavior on critical systems' failure behavior considering the origin of the failure. Then, mitigation methods can be used to mitigate the failures and the analysis can be used iteratively to reach the required level of safety.

valueCoarse on OP6 -> valueSubtle on IP6 -> valueSubtle on IP5 -> valueSubtle on IP4 -> late on IP3 -> late on IP2 -> late on IP1

Figure 2.17: Back propagation of the results on AR-HUD example

Synergy of qualitative and quantitative dependability analysis techniques

A synergy of qualitative and quantitative dependability analysis techniques is proposed in [88]. It contains State-based analysis and Failure Logic Analysis (FLA). State-based analysis technique [89] is a quantitative technique and FLA is a qualitative analysis based on qualitative behavior of components and their causes.

It is required to have information or assumptions about the system architecture to be used for modeling system architecture. Formalism used in state-based analysis is Stochastic Petri Nets (SPNs) [90] with general probability

distributions. There are three types of behavior modeling used in these two analysis techniques, which are simple stochastic behavior, error model and Failure Logic Analysis (FLA) [88]. These three types of behavior modeling are described in the following paragraphs.

Simple stochastic behavior uses probability distribution for specifying the time to the occurrence of a failure and the time required to fix the component after failure occurrence, if available. Possible failure modes and their probabilities also can be provided. As it is shown in Figure 2.18, exponential distribution with rate of 1.0e-6 per hour of operation is used for illustrating time to failure of this hardware component. Possible failure modes in case of failure in the output and their probabilities are shown in this example, which are omission (means output is not provided when expected) with probability of 80% and valueSubtle (means output is not in the expected range and it is not detected by user) with probability of 20%.

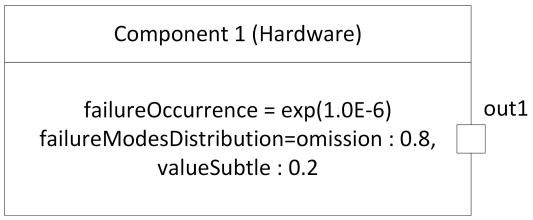


Figure 2.18: Modeling a hardware component with stochastic behavior [88]

Error model is defined by using a set of finite state machines modeling internal faults, external faults and their probabilities. It also models transitions between states. Error models are used when there are detail information about the component's failure behavior [88]. For example in Figure 2.19, a software is modeled by two error models modeling internal fault occurrence and effect of external faults. In the top part of the picture, probability of occurrence of internal fault is defined as exp(6.0E-6) and it would propagate to an undetected error state leading to output failure mode omission with weight 0.8 or it would propagate to an error state incorrect value with weight 0.2. In the bottom part of the picture, omission external fault is considered propagated to undetected error state leading to omission failure mode in the output.

FLA behavior is defined by assigning possible failure modes in the input to possible failure modes in the output (the same as FPTC rules). In this type of behavior modeling probabilities are not considered. For example, in Figure 2.20, a software is modeled by defining FLA behavior. In this example,

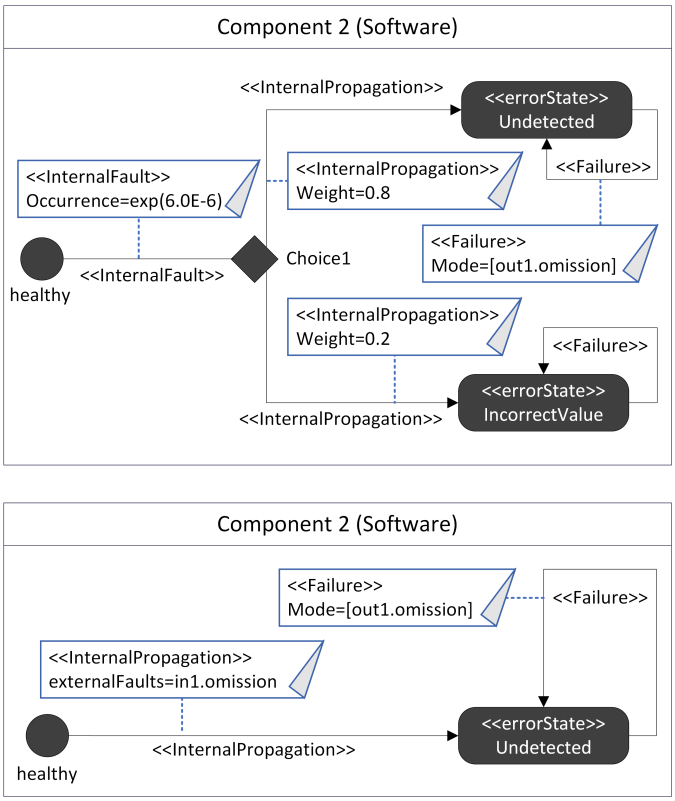


Figure 2.19: Modeling a software component with error models [88]

there are two inputs (In1, In2) and two outputs (out1, out2) for the software component. NoFailure (normal behavior) on input In1 and valueSubtle on input In2 will lead to valueSubtle on out1 and noFailure on out2. Relationship of other possible failure modes on inputs and outputs are defined similarly.

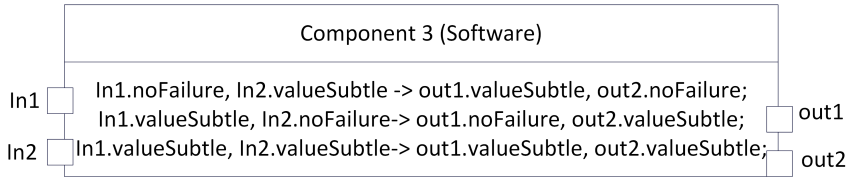


Figure 2.20: Modeling a software component with FLABehavior [88]

The following metrics can be measured by the quantitative analysis:

- Reliability: the probability that system continuously remains in proper state from the time 0 up to time t.
- Availability:
  - Immediate: the probability that system is in proper state at time t.
  - In a time interval: the fraction of time that system is in proper state in a given time interval.
- Probability of Failure on Demand (PFD): the probability that the system fails to provide a requested service. It can be obtained by calculating 1 minus immediate reliability.

We use these techniques as the bases of our research and provide the required contributions in order to assess risk of AR-equipped socio-technical systems in compliance with safety standards.

2.6 Goal Question Metric method

The Goal Question Metric approach (GQM) [91] is a method for measuring based on specific purpose. Based on this method, goals should be defined at the first step. Then, research questions should be defined based on the goals. Finally, metrics should be defined based on the research questions and in a way to reach the defined goals. In this way the metrics provide the possibility to analyze goal achievement. It has been used in several projects such as NASA Goddard Space Flight Centre environment [92].

Chapter 3

Research Summary

In this chapter, we present a summary of the research addressed in this thesis. In particular, we present the research process used (See Section 3.1), the problem formulation (see Section 3.2) and the research goals (see Section 3.3).

3.1 Research Process

Conducting research in a particular area requires comprehending related research methods and being able to apply them. A framework for research methods within computing area is shown in Figure 3.1 [93]. There are four main steps including *problem identification, data collection, data processing* and *evaluating the result*. The research starts with *identifying the problem* and defining what we want to achieve and what is happening. This step can be conducted through study of state-of-the-art. The next step is *data collection*, where it is required to define how and where to collect data. This step can be conducted through literature review. Once the data is collected, it should be processed through the step *processing data*. Processing data can be conducted through methodologies such as classifying data and creating taxonomy. Finally, the last step is *evaluating the result*, where goal achievement can be analyzed and limitations can be identified. Evaluating the results can be conducted through conducting a case study.

An overview of the adapted research method used in this thesis is shown in Figure 3.2. First, we identify the research problem and define the main goal. Then, we divide the research problem to sub-problems and identify the sub-

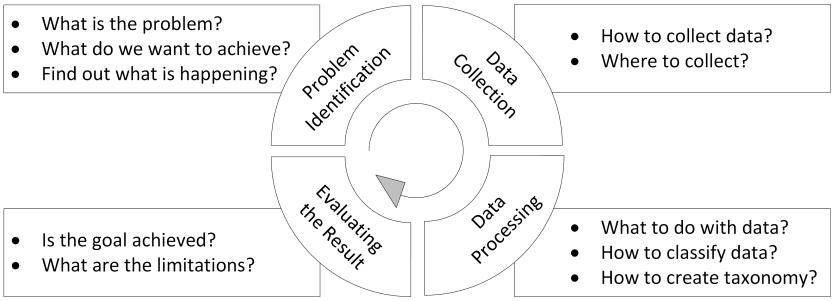


Figure 3.1: A framework for research methods within computing area

goals and study the state-of-the-art. After that, we propose a solution for the gap identified on the study. Next, we implement the solution and evaluate on an academic example. After this step, the results can be published as a paper. Integration and communication with industry can also be considered as a step after academic evaluation of the proposed solution, to enable evaluating the solution on real world problem. Finally, if the result is accepted for the real world problem, it will provide the possibility of publishing a paper, otherwise problem should be identified to repeat the iterative task for the new research problem.

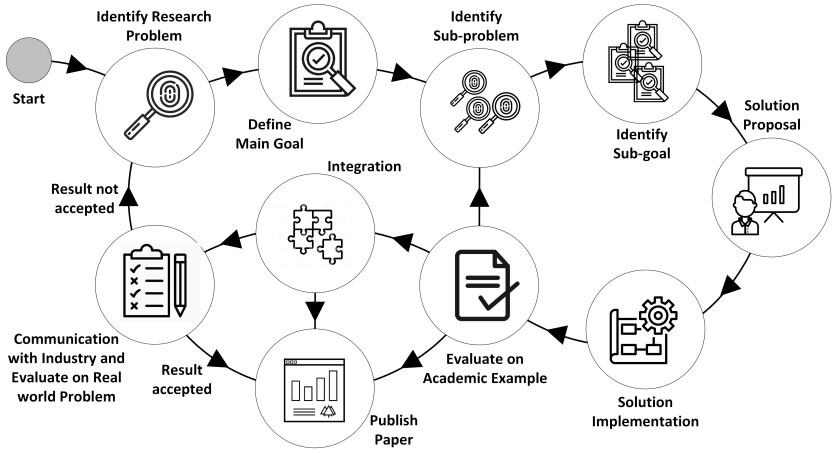


Figure 3.2: Overview of our research methodology

3.2 Problem Formulation

New organizational changes in the past twenty to thirty years may introduce new kinds of risks. In addition, new technologies, such as augmented reality are used with the aim of increasing human performance and extending human capabilities, meanwhile failing of these extended capabilities introduces new types of failures. Thus, these new organizational changes in addition to technological changes may cause new types of human failures and new types of faults leading to human failures. A socio-technical system contains socio entities including human and organization and technical entities including software and hardware. In order to do the risk assessment, it is crucial to identify possible dependability threats and to provide modeling and analysis constructs for modeling and analysis of dependability information. There are various safety standards, which we require to consider while we use the modeling and analysis techniques to perform the risk assessment. However, there is no data on effects of new organizational changes on modeling and analysis techniques. In addition, we lack a dependability analysis approach to be used for risk assessment of AR-equipped socio-technical systems considering new organizational changes. There is also no investigation on the capabilities of the current techniques for risk assessment of AR-equipped socio-technical systems in compliance with safety standards.

This thesis aims at strengthening risk assessment in augmented reality-equipped socio-technical systems in compliance with safety standards considering post normal accidents by providing a safety-centered risk assessment framework. More specifically, the thesis identifies the dependability threats related to new organizational changes leading to post normal accidents to get involve modeling elements and analysis techniques based on these threats and by providing extensions. Based on these extensions a risk assessment framework is proposed in compliance with safety standards. To reach these goals, in the first step we reviewed post normal accident theory to extract influencing factors which would act as dependability threats leading to post normal accidents. We proposed new modeling elements based on the extracted factors and we extended an existing metamodel to enrich it with more expressive power to be used for modeling AR-equipped socio-technical systems considering post normal accidents. In the second step, we proposed an analysis process for qualitative and quantitative analysis of system failure behavior to assess the risk of AR-equipped socio-technical systems considering proposed extensions. Then, we proposed a risk assessment framework containing modeling and analysis phases to be used for assessing risk of AR-equipped socio-technical systems

considering organizational changes and safety standards. In the third step, we applied the proposed framework in two different domains. We applied the proposed framework in automotive domain to validate modeling and analysis capabilities of the framework and to find out if the proposed steps can support activities of related safety standards. Then, we applied the proposed framework in robotic domain to demonstrate applicability and effectiveness of the proposed framework in a new domain. In the fourth step, we conducted a systematic literature review to position our work and to compare it with other related works.

3.3 Research Goals

As presented in Section 3.2, this thesis aims at strengthening risk assessment of safety-critical AR-equipped socio-technical systems in compliance with safety standards considering post normal accidents by providing a risk assessment framework. To reach this goal, we define the main research goal as follows:

Overall Research Goal:
Strengthening risk assessment of safety-critical AR-equipped socio-technical systems in compliance with safety standards considering post normal accidents.

In order to address the overall research goal, we define concrete subgoals that address the specific challenges. The subgoals are described as follows:

Subgoal 1: ***Capturing dependability threats leading to post normal accidents in AR-equipped socio-technical systems.*** Influencing factors on system behavior, which would act as dependability threats leading to post normal accidents are extracted based on post normal accident theory. Metamodel extensions are provided based on these extracted influencing factors by proposing new modeling elements, which are helpful for capturing the new dependability threats leading to post normal accidents.

Subgoal 2: ***Integrating captured dependability threats in risk assessment of AR-equipped socio-technical systems.*** Once the required constructs for capturing dependability threats leading to post normal accidents are available, we can conduct dependability analysis. For doing that, we

propose a process for qualitative and quantitative dependability analysis that can be used to assess risk of AR-equipped socio-technical systems. Then, based on the modeling and analysis extensions, we propose a safety-centered risk assessment framework.

Subgoal 3: ***Validating modeling and analysis capabilities, applicability and effectiveness of contributions for risk assessment in AR-equipped socio-technical systems.*** Once we have the required extensions and they are integrated in a risk assessment framework, it is required to validate modeling and analysis capabilities in risk assessment of AR-equipped socio-technical systems with respect to safety standards. We applied the framework on two AR-equipped socio-technical systems from two different domains and we showed how different steps of the framework can support required activities in safety standards in order to demonstrate applicability and effectiveness of the framework.

Subgoal 4: ***Positioning and comparing the contributions of our work.*** Finally, we position our work and compare it with other related studies by conducting a systematic literature review. We have done preliminary literature review in different stages of our research, but at this stage of the research we conducted a systematic literature review on the topic and we provided a comprehensive review and assessment on the literature based on our defined research questions.

Chapter 4

Thesis Contributions

In this chapter, we present a brief description of the technical contributions provided by this thesis. In particular, in Section 4.1, we describe the first contribution, which is a metamodel extension to capture dependability threats leading to post normal accidents. In Section 4.2, we describe the second contribution, which is our proposed process for dependability analysis by extending a synergy of qualitative and quantitative analysis. In Section 4.3, we propose the third contribution, which is a risk assessment framework for AR-equipped socio-technical systems in compliance with safety standards considering post normal accidents. Then, we describe our fourth and fifth contributions in Section 4.4 and Section 4.5, in which we apply our proposed framework in two different domains to demonstrate the applicability and effectiveness of the framework. We show how different steps of the framework would support required activities in the related safety standards. Finally, in Section 4.6, we describe the final contribution, which is a systematic literature review for risk assessment of safety-critical socio-technical systems considering development of conceptualization of socio-technical systems.

4.1 Metamodel Extension to Capture Post Normal Accidents

In this section, first, based on the post normal accident theory and global distance metric, discussed in Subsections 2.2.1 and 2.2.2, we extract the factors influencing on human performance leading to accident. Then, we extend the

organization and human modeling elements of the extended conceptual meta-model, which is explained in Subsection 2.5.1. Finally, we provide a potential usage on an example.

4.1.1 Extracted Influencing Factors

Influencing factors are selected, if they have the potential to influence on human performance leading to accidents. Definitions and safety effects discussed in [29] and [5] are used for identifying these factors.

Extracted influencing factors on human performance are divided into two groups. The first group is organizational factors and the second group is human factors. These two groups are as follows:

1. Group 1 (organizational factors)
   - **Globalized environment**: It may cause complex interactions between different entities. These implications may affect on human performance and would lead to an accident.
   - **Self-regulated environment**: It may cause missing of independent oversights by states that may affect on human performance and would lead to an accident.
   - **Organizational strategy**:
     - **Financialized strategy**: It may cause pressure for returning the investment and shifting power to financial actors. These implications may affect on human performance and would lead to an accident.
     - **Industrial strategy**: It may cause changes in industrial relations that may affect on human performance and would lead to an accident.
   - **Organizational structure**:
     - **Networked structure**: It may cause increase in complexity of interactions across organizations and other entities of the system that may affect on human performance and would lead to an accident
   - **Digitalized task**: It may cause complexity in human and machine interactions and development of new information structures. These changes may affect on human performance and would lead to an accident.
   - **Standardized task**: It may cause change in practices that may affect on human performance and would lead to an accident.
2. Group 2 (human factors)

- **Global distance**:
  - **Geographic distance**: It may cause difficulties in managing physical places that may affect on human performance and would lead to an accident.
  - **Temporal distance**: It may cause difficulties in time management that may affect on human performance and would lead to an accident.
  - **Cultural distance**: It may cause difficulties in communications that may affect on human performance and would lead to an accident.

4.1.2 Extended Modeling Elements

The first group of factors can be used for extending organization modeling elements and the second group can be used for extending human modeling elements.

Based on the provided definitions for each of the extracted influencing factors and based on the three categories of organization modeling elements proposed in [38], we add new modeling elements to the categories, shown in Figure 4.1. The modeling elements with dotted line border are elements for modeling AR-extended organizational aspects. Extended modeling elements in this section are shown with gray color and the previous categorization of meta classes are shown with white color.

For example, time pressure is an organizational modeling element using to model scenarios that time pressure may influence on human performance and may lead to system failure or an accident. AR guided task refers to a task that AR is used for guiding the operator for doing the task. If this task is not defined correctly, it may influence on human performance leading to system failure. Standardized task is an extended modeling element proposed in this section based on post normal accident theory. Standardization may influence on human performance and may lead to an accident.

We also use global distance metric for extending human modeling elements, shown in Figure 4.2. The modeling elements with dotted line border are elements for modeling AR-extended organizational aspects. Extended modeling elements in this section are shown with gray color. For example, social modeling element is a human modeling element. This modeling element can be used for modeling scenarios that problem in communication between people would lead to misunderstanding and failure in human performance. Thus, it would lead to an accident. Social presence modeling element can be used for modeling scenarios that using AR would decrease social presence, mean-

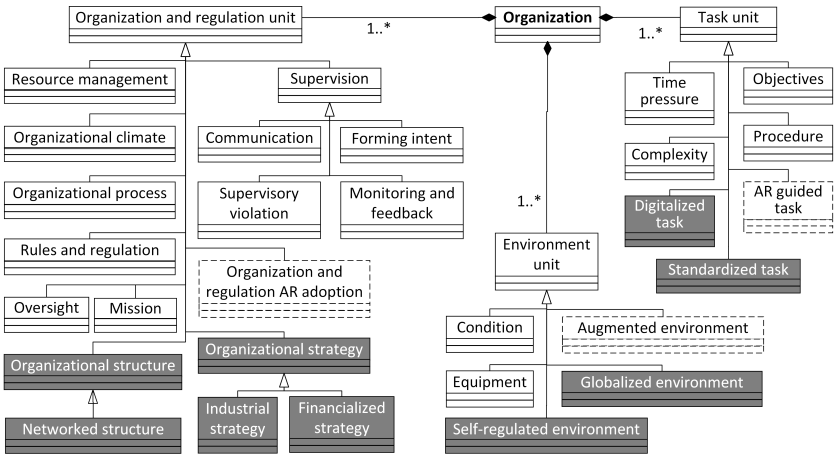


Figure 4.1: Extended organization modeling elements [69]

ing that people miss their communication because of AR. Thus, it may influence on human performance and it may lead to an accident. Global distance is the extended modeling element proposed in this section. This modeling element can be used for modeling scenarios that for example cultural distance between people causes misunderstanding. Thus, it may influence on human performance and it may lead to an accident.

4.1.3 Potential Usage on an Example

British Petroleum (BP) is one of the biggest multinational companies in the world. A series of accidents between 2005 and 2010 in multinational BP in different branches occurred. We use this example to show our extension contribution in modeling conditions leading to these accidents.

Based on the analysis of these accidents using commission reports and social concepts for interpretation [94], identified potentials for these accidents are as follows:

- Networked structure of BP
- Lack of appropriate learning from experience
- Fault in control authority

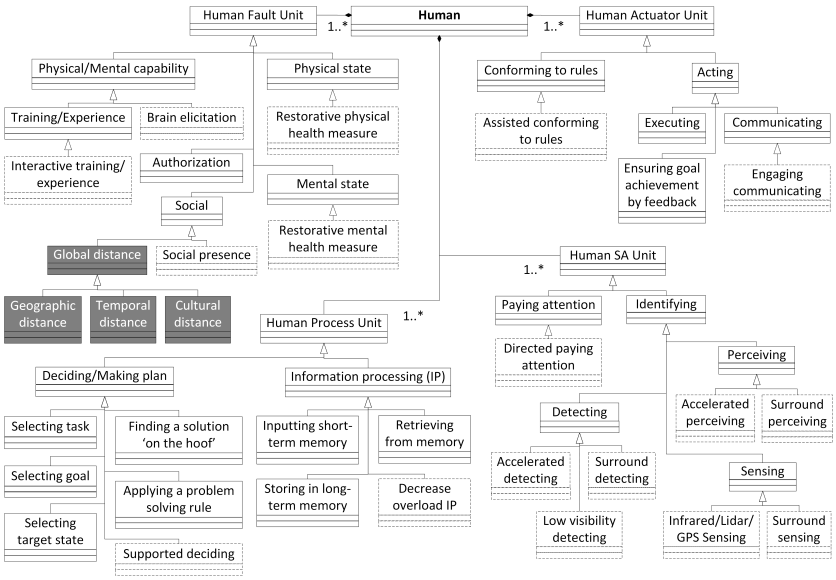


Figure 4.2: Extended human modeling elements [69]

• Strategies of CEO of the company

In Figure 4.3, we show a modeling example using our extended modeling elements. The modeling elements representing three factors leading to accident in BP, as examples, are shown using networked structure, experience and organizational strategy components. Two of these three used modeling elements are the modeling elements extended in this section. These modeling elements, which are based on the factors explained in the post normal accident theory as factors leading to post normal accidents are shown in gray. We show three scenarios and in each of them failure in one of the three components has contributed to accident. For example, in the first scenario (S1), output of networked structure produces a failure and the other three provide correct service, which means no failure in their outputs. Final output of the system, which is shown by OP14 produces failure because of the failure in networked structure component. In the second scenario (S2), the reason for failure in the output of the system is failure in OP4, which is output of organizational strategy component. In the third scenario (S3), the reason for failure in the output of the system is failure in OP11, which is output of the experience component. Similarly, different scenarios can be modeled and discussed using different representation constructs proposed in our conceptual metamodel.

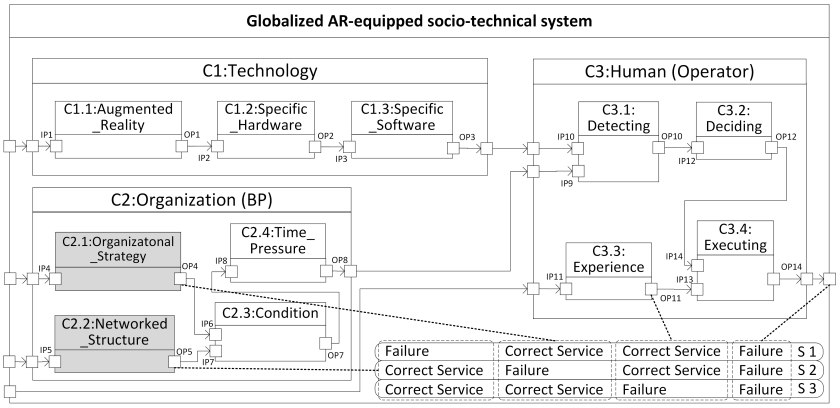


Figure 4.3: Globalized AR-equipped socio-technical system modeling [69]

Another interpretation is proposed by Jean-Christophe Le Coze in [2], in the context of globalization. In this interpretation, the author explains how deregulation, externalization, standardization, digitalization and financializa-

tion have contributed to the accidents in BP. Since our extension contains the representation constructs required for modeling these concepts, different scenarios using these modeling elements can be considered and discussed during modeling and risk assessment to improve system behavior.

Managing multinationals is a big challenge for companies like BP. Considering technological factors and organizational factors in SafeConcert metamodel were not enough for describing such events. We show in this example that the new proposed modeling elements can be helpful for modeling recent factors such as networked structure of an organization in order to incorporate their effect while performing risk assessment.

This contribution is presented in Paper A (see Chapter 7).

4.2 Process for Dependability Analysis Based on Our Extensions

In this section, we propose an extension based on extended modeling elements and Concerto-FLA analysis technique [15], explained in Subsection 2.5.2. We build on top of the synergy of qualitative and quantitative analysis in [88] explained in Subsection 2.5.2. We aim at extending this synergy by incorporating socio-related and AR-related aspects explained in Subsection 2.5.1 and 4.1. Our proposed analysis process is illustrated in Figure 4.4.

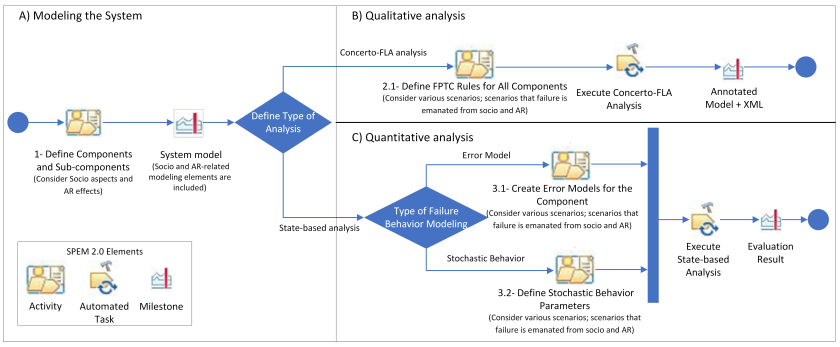


Figure 4.4: The proposed extended analysis process [95]

The added value with respect to the synergy of quantitative and qualitative analysis is the possibility of analyzing various socio, AR-related aspects and

aspects related to organizational changes and their effects on system behavior. Metamodel extensions related to AR and organizational changes are used in the system modeling by including modeling elements related to AR and organizational changes in the system model. In case of using qualitative analysis, Concerto-FLA analysis can be used for defining FPTC rules for AR-related components and automated analysis is used for obtaining the annotated model by analysis results. In case of quantitative analysis, error model or stochastic behavior are used for analyzing system behavior including AR-related and organizational changes effects.

Part A of Figure 4.4 contains the activity that should be done for preparing the system model. This activity is defining components and subcomponents. Then, we need to decide about analysis type. If we need to perform qualitative analysis, we perform the next activities based on Concerto-FLA analysis technique (Part B), otherwise we perform based on State-based analysis technique (Part C).

Based on Part B of the figure, FPTC rules should be defined for all components. Then, Concerto-FLA analysis will be executed and model annotated by analysis results will be provided.

Based on Part C of the figure, failure behavior modeling type should be defined. If we choose to use error model, then we need to create the error model of the desired component. If we choose to use stochastic behavior, then we should define the related parameters. Next step is to execute the state-based analysis and to measure the evaluation result.

Result of the analysis can be used for hazard identification, hazard analysis, defining safety goals and safety requirements.

We explain the activities of all the steps in the following subsections and in Table 4.1, we compare these steps of our proposed extended process with the previous process in [88].

4.2.1 Define Components and subcomponents

Main entities incorporating in a system are considered as the main components. It is important to consider socio entities, which are human and organization. Defining subcomponents are based on important aspects of each entity. In technical components, important aspects are defined based on technical description of the system. Human important aspects are defined based on human functions and human internal states. Organization important aspects are defined based on organizational important aspects. Human and organization modeling elements introduced in the extended metamodel explained in Sub-

section 4.1 are the modeling constructs that can be used for defining human and organization subcomponents. For example, condition, environment and any other influencing factor on human performance can be considered as organizational important aspects. The extensions include organizational changes aspects, which should be considered in defining subcomponents.

4.2.2 Define FPTC Rules for All Components

This activity should be done based on the syntax explained in Subsection 2.5.2. In order to define FPTC rules, each component/subcomponent should be analyzed individually. We should define the possible failure modes at each of their inputs and outputs for various scenarios. Then, FPTC rules can be used for relating the failure modes at inputs to the failure modes at outputs. For example, a camera would not receive the input (raw image) because of the obstacle in front of it. Input failure mode in this example is omission as explained in Subsection 2.5.2. Based on technical analysis of the camera, we would model it as propagational (explained in Subsection 2.5.2). It means that the failure mode in input propagates to the output port and it does not provide the output.

4.2.3 Create Error Models for the Component

This activity should be done based on the syntax explained in Subsection 2.5.2. In order to define error models, the intended component/subcomponent should be analyzed individually. State machine for each component including internal and external faults and their probabilities should be defined for various scenarios.

4.2.4 Define Stochastic Behavior Parameters

This activity should be done based on the syntax explained in Subsection 2.5.2. In order to define stochastic behavior parameters, the intended component/subcomponent should be analyzed individually. Possible failure modes and their probabilities should be defined for various scenarios.

4.2.5 Potential Usage on an Example

In this subsection, we provide an example with the objective of presenting the analysis capabilities provided by the proposed process. First step is to model the system, as shown in part A of the process. Then, Concerto-FLA analysis

Table 4.1: Comparison of our Proposed Extended Process with the Previous Process in [88]

Steps	**In the previous process in** [88]	**In our proposed extended process**
Define components and subcomponents	Technical components/subcomponents are defined.	Technical + socio + AR-related + components/subcomponents related to organizational changes are defined.
Define FPTC rules for all components	Scenarios including failures emanated from technical components/ subcomponents are considered.	Scenarios including failures emanated from technical + socio + AR-related + components/subcomponents related to effects of organizational changes are considered.
Create error models for the component		
Define stochastic behavior parameters		

can be used for qualitative analysis (Part B) and state-based analysis can be used for quantitative analysis (Part C). We consider an industrial monitoring system introduced in [96]. We use this system as an example for analyzing AR-equipped socio-technical system.

The industrial monitoring system uses a sensor for receiving raw data. Raw data is processed in server and it is organized to be represented to the user for making decisions. AR can be used for providing graphical or textual instructions for solving a problem, configuring an equipment or maintenance activities. In this example, we consider using AR for providing visual alarm in case of problem in a special equipment under control.

Modeling the System: This system includes technical and socio entities. Technical entity is the monitoring system and socio entities are the user and organization. We model each of these entities based on their description and based on their important aspects.

The technical components of this system are defined based on description of monitoring system as follows:

- Sensor: it is a hardware component. It can be various sensors, for example a camera receiving raw data of a specific equipment, which is considered for monitoring.

- Server: it is a hardware component. It is a computer that contains processing unit for processing the data.

- Processing unit: it is a software component. It processes the received data from sensor and organizes it in a format to be used by the user.
- AR application interface: it is a hardware component. It is the interface between the user and the server. It is a screen containing AR technology notations.

The user can be characterized based on its important aspects, which are human functions and human internal states. We use four following modeling elements of the extended metamodel explained in Subsection 4.1.

- Directed paying attention: it refers to an AR-extended human function. It models the function paying attention when it is directed to a specific position by using AR technology. For example, in this case, if there is something strange related to the equipment which is under monitoring, then AR technology can be used for displaying a red circle around the strange area. Thus, the user attention will be directed to the position to make a decision to prevent any probable risk.
- Training: it refers to training received by the human.
- Deciding: it refers to human deciding function.
- Executing: it refers to human executing function.

The organization can be modeled based on important organizational aspects. We use the following modeling elements of the extended metamodel explained in Subsection 4.1.

- Condition: it refers to the condition of the organization where the monitoring task is performed.
- Organization and regulation AR adoption: it refers to an AR-extended aspect. It models the adoption process needed in the organization to be able to use AR.
- AR guided task: it refers to the task that AR is used for guiding the human to do that. For example, a task should be defined in an organization that in case of special AR alarm the user should react.

Based on the described entities and their important aspects, we provided the model shown in Figure 4.5. Sensor receives raw data (shown by RD in Figure 4.5) and provides the output for processing unit. Data is processed in

processing unit and its output is shown in AR application interface to the user. Organization and regulation AR adoption is influenced by regulation authorities (REG) and it affects on AR guided task defined by organization. AR guided task is also influenced by condition of the organization, which is influenced by condition out of the organization (shown by CON). Output of monitoring system which is a visual description on a screen influences on human directed paying attention and output of the organization influences on training. Finally, human deciding function is influenced by directed paying attention and training. Human executing function is influenced by human deciding function. Output of the system, which is output of the human component is human function (HF).

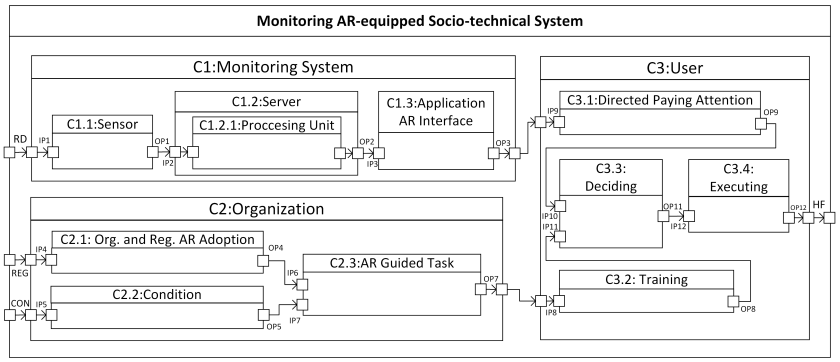


Figure 4.5: Modeling the system [95]

Qualitative Analysis: As it is shown on part B of the Figure 4.4, in order to provide the qualitative analysis, we need to define FPTC rules for all components. These rules should be defined based on individual analysis of components and based on the assumptions of various scenarios. For example, we provide the FPTC rules for a specific scenario and we provide the system behavior based on failure propagation.

- **Definition of scenario**: We assume that the equipment under monitoring is in a situation that it can harm a person. The information is received by the sensor and it is processed by the processing unit and a visual alarm is displayed on the AR display. However, we assume that there is a failure in organization and regulation AR adoption. For example, organization should

update regulations in order to include AR related considerations and trainings. Since there is no rule defined in the organization, the required training is not provided for the user. The user's attention is directed to the alarm, but the user does not take the correct decision and does not provide the required execution function to prevent the harm.

- **Modeling of the failure behavior**: In this scenario the organization and regulation AR adoption is behaving as a source (source behavior is explained in Subsection 2.5.2). The input of this component receives noFailure, but in the output it provides valueSubtle. The reason is that organization has not updated rules and regulations to adopt AR (valueSubtle) and the user does not receive the required AR-related training (omission). Since the user does not receive the required AR-related training, the deciding component provides valueSubtle failure mode in its output. Thus, the user does not provide the required execution (omission). Monitoring system components are behaving as propagational and propagate noFailure from input to output.
- **Analyzing the system behavior**: Analysis annotations are shown in Figure 4.6. ValueSubtle in OP4 means that the AR adoption in organization and regulation is not performed correctly. ValueSubtle failure mode transforms to omission in AR guided task and it propagates in training. Then, in deciding it transforms to valueSubtle and in executing transforms to omission. The failure propagation is shown by blue color.

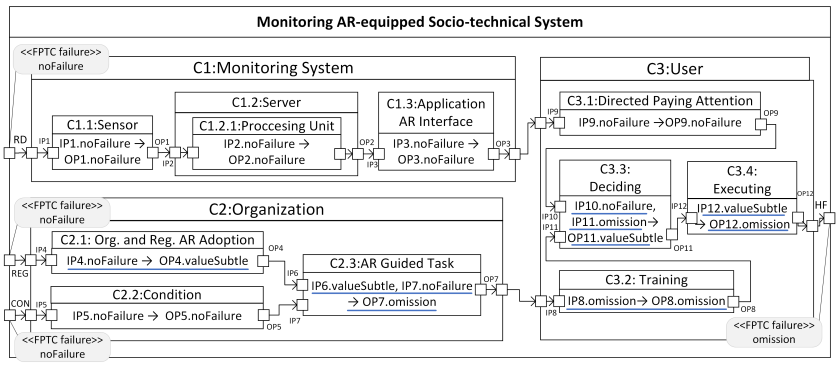


Figure 4.6: Qualitative analysis of the system [95]

- **Interpreting the results**: Based on the back propagation of the results, we can explain how the rules have been triggered. Omission in HF is because of

valueSubtle in OP11. ValueSubtle in OP11 is because of omission in IP11 and we continue until IP4, which is input port of organization and regulation AR adoption. Thus, this component caused the failure in the system output. The identified hazard is not providing correct deciding and the reason for this hazard is failure in organization and regulation AR adoption. System failure in this scenario may lead to fatal injuries for people around the intended equipment. Thus, safety goal should be defined to overcome this risk. For example, for this scenario, safety goal can be defined as follows:

– **Safety goal**: The organization should update rules and regulations based on AR and should provide the required AR training.

By using the qualitative analysis and by considering various possible scenarios, various safety goals can be defined. Based on safety goals, system design can be updated and analysis of system behavior can be performed for more iterations to reach the accepted level of safety.

Quantitative Analysis: Based on part C of the Figure 4.4, in order to provide the quantitative analysis, we should model the failure behavior using error models or stochastic parameters. Similar to qualitative analysis these models should be defined based on individual analysis of components and based on the assumptions of various scenarios. For example, we provide stochastic behavior modeling for a specific scenario and we provide the analysis result.

- **Definition of scenario**: Similarly, we assume that the equipment under monitoring is in a situation that it can harm a person. The information should be received by the sensor and it should be processed by the processing unit. Then, a visual alarm should be illustrated through AR display and the user should decide based on illustrated alarm and based on received training from organization to execute a needed task preventing the risk.
- **Modeling of the failure behavior**: In this scenario, for each component we consider possible failure modes and their probabilities as it is shown in Figure 4.7. Probabilities can be defined based on previous accident reports or based on expert opinion. For example, in this scenario, organization has not updated rules and regulations based on AR technology. Thus, failure probability in the Org. and Reg. AR adoption component is high (0.9).
- **Analyzing the system behavior**: In order to perform the analysis, we can consider the hazard related to this scenario and calculate the intended measure or failure mode probability in system output. We consider the same

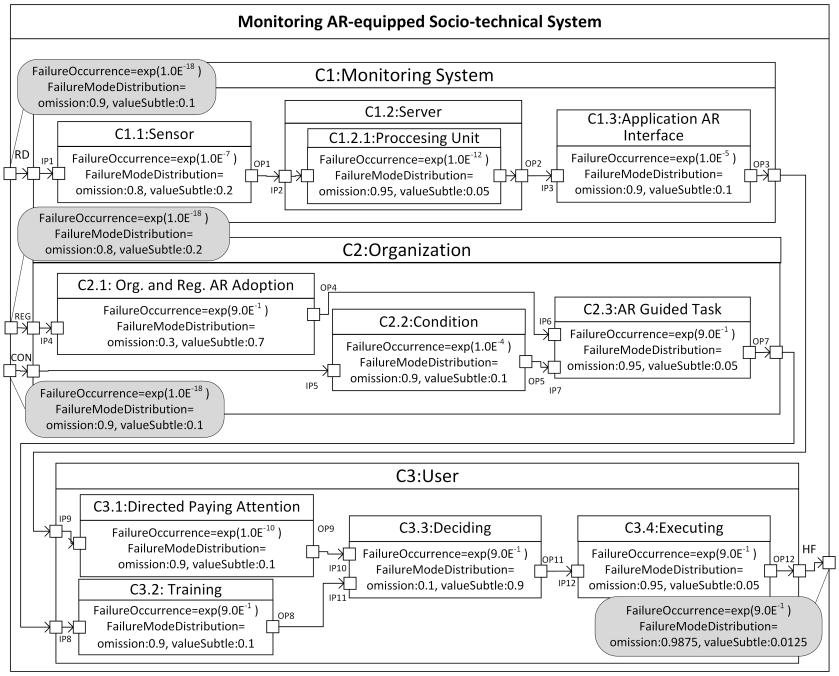


Figure 4.7: Quantitative analysis of the system [95]

hazard as the one we considered in qualitative analysis, which is not providing correct deciding. In this case, we want to calculate the probability of omission failure mode in system output. The result for this assumed scenario is shown in Figure 4.7. Calculation is an automatic task, which can be performed by running the analysis in the toolset. For example, failure in output of executing function would be of type omission or valueSubtle. The probability of omission failure mode is calculated based on the probability of executing function providing an omission failure mode while its input can be different failure modes with different probabilities and all the possible conditions should be considered in the calculation. In this example, probability of failure occurrence in system output (human function) is 0.9, which shows that the reliability of the system from time 0 up to time 1000 hours is around $1 - 0.900 = 0.100$. The probability for omission failure mode will be $0.9 * 0.9875 = 0.88875$.

- **Interpreting the results**: Based on the back propagation of the results, we can explain how the hazard would happen and how much is the probability. For example, in this scenario the probability of omission failure mode in output is 0.88875 and the reason is high probability of failure in organization and regulation AR adoption.

  Similar to the previous scenario, safety goal can be defined in order to decrease the probability and prevent the risk. The probability can be helpful to decide if a special failure mode in the system output should be overcome or it can be ignored due to low probability of its occurrence.

  By using the quantitative analysis and by considering various possible scenarios, various safety goals can be defined. Based on safety goals, safety requirements can be defined and system design can be updated. Then, analysis of system behavior can be performed for more iterations to reach the accepted level of safety.

This contribution is presented in Paper C (see Chapter 9).

4.3 Risk Assessment Framework for AR-equipped Socio-technical Systems

Our proposed framework for assessing risk of AR-equipped socio-technical systems is based on the proposed modeling extensions and the extended anal-

ysis process. We name this framework FRAAR (Framework for Risk Assessment in AR-equipped socio-technical systems).

The proposed modeling extensions on SafeConcert is explained in Subsection 4.1. The proposed extended analysis process is explained in Section 4.2. Essentially, the added value with respect to SafeConcert and synergy of quantitative and qualitative analysis is the availability of modeling and analysis capabilities for modeling and analyzing various socio aspects, AR-extended human functions, AR-related influencing factors on human functions and factors in relation to post normal accidents.

We use V-model structure to illustrate methodology of the provided framework. Different steps of the methodology are shown in Figure 4.8.

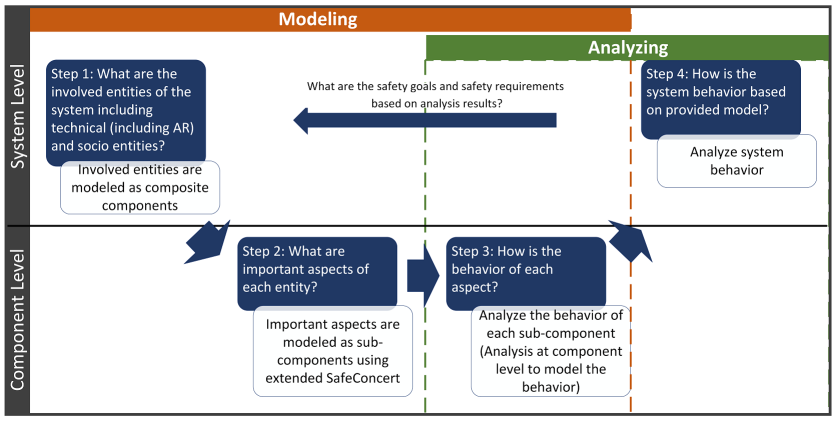


Figure 4.8: Methodology of the provided framework for assessing risk of AR-equipped socio-technical systems [54]

As it is shown in Figure 4.8, there are four main steps. In the first step, we need to answer to the question of what are the involved entities in the system. Since we model the system as a component-based system, defining involved entities determines the composite components. In an AR-equipped socio-technical system, involved entities include technical (including AR) and socio entities.

In the second step, we need to identify important aspects of each entity. These important aspects are used to determine subcomponents of each composite component. In this step, our proposed extended modeling elements explained in Subsection 4.1 can be helpful to have a list of important aspects.

Based on the scenario and the selected case study, required subcomponents can be selected. For example, paying attention can be considered as an important aspect of a human driving a car. Not paying attention would lead to failure in deciding, which is a hazardous behavior that would lead to system risk.

Third step is to model the behavior of each subcomponent, which should be done based on analysis of each subcomponent individually. FPTC syntax explained in Subsection 2.5.2, can be used for modeling the behavior of each subcomponent. Error model and stochastic behavior can also be used in case of availability of quantitative data to be used for quantitative analysis.

Finally, last step is analyzing system behavior, which provides system behavior based on the provided model. We can do this step using Concerto-FLA analysis technique explained in Subsection 2.5.2 based on part B of Figure 4.4, and we can do this step based on part C of Figure 4.4 as explained in Section 4.2.

Based on the analysis results there would be feedback for changing the system behavior in order to decrease risk. This feedback can be suggestions for safety requirements or functional modifications.

This contribution is presented in Paper B (see Chapter 8).

4.4 Applying the Framework in Automotive Domain

4.4.1 Objectives of Case Study

Our objectives include presenting the modeling and analysis capabilities of our framework in compliance with related safety standards explained in Subsection 2.4.1. In other words, we aim at estimating the effectiveness of modeling and analysis capabilities in predicting risk caused by new dependability threats and we also aim at presenting how the framework can be supportive for safety standards. In order to do that, we use an industrial case study from automotive domain. Analysis results can be used for defining related safety requirements.

4.4.2 Research Methodology of Case Study

We use case study research methodology based on [97]. The steps carried out for the presented research are presented in Figure 4.9. In the first step, objectives and the structure of the research are discussed.

In the second step, we asked Xylon Company for a case study in the context of augmented reality socio-technical systems. Surround view system as a case study was suggested by this company and a meeting was organized to decide about the collaboration. We also discussed about system description.

In the third step, system architecture was provided based on information provided by the company and it was reviewed in several iterations for improvement.

In the fourth step, analysis of the case study was provided based on Concerto-FLA analysis technique and it was reviewed in iterations for improvement.

In the fifth step, a discussion about results and lessons learnt was provided. Then, the results were reviewed and a discussion about validity of the work was provided.

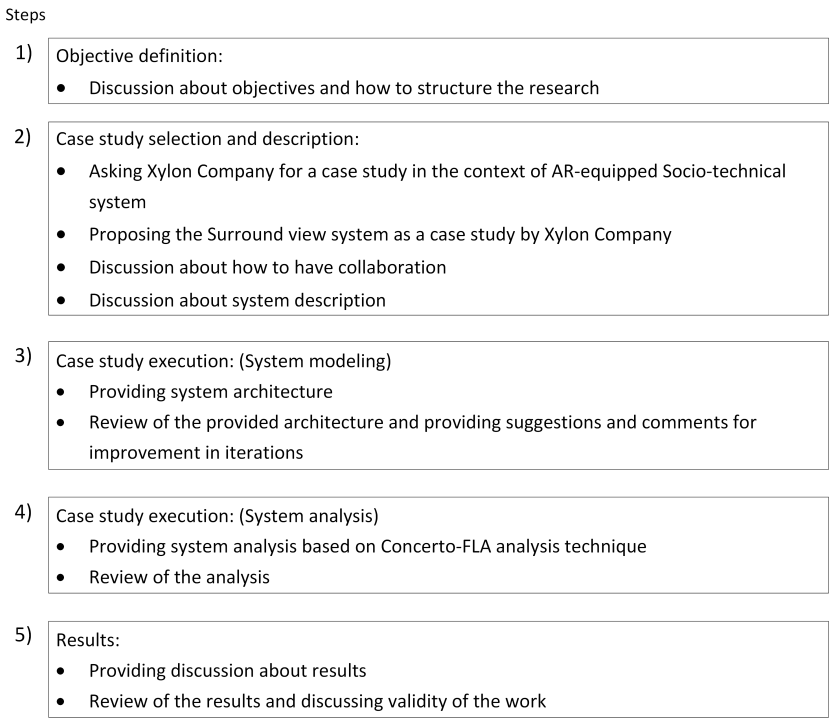


Figure 4.9: Steps taken for the carried out research [54]

4.4.3 Case Study Selection and Description

The case study is conducted in collaboration with Xylon, an electronic company providing intellectual property in the fields of embedded graphics, video, image processing and networking.

In this study, we select as case study subject a socio-technical system containing the following entities:

- Road transport organization (socio entity): representing the organization responsible for providing transport rules and regulations, proper road conditions and etc.
- Driver (socio entity): representing a human who is expected to drive a vehicle and park it safely by utilizing augmented reality technology used in the surround view system of the vehicle.
- Vehicle (technical entity): representing vehicle containing surround view system (a SEooC with the potential for using in vehicles with high levels of driving automation. However, currently it is used at driving automation level 0 (SAE level 0). It includes augmented reality technology to empower drivers).

Surround view systems are used to assist drivers to park more safely by providing a 3D video from the surrounding environment of the car. In Figure 4.10, it is illustrated how the 3D video is shown to the driver. As it is shown in Figure 4.10, driver can have a top view of the car while driving. This top view is obtained by compounding 4 views captured by 4 cameras mounted around the car and by changing point of view. It is like there is a flying camera visualizing vehicle's surrounding, which is called virtual flying camera feature. A picture of a virtual car is also augmented to the video to show the position of the car. Navigation information and parking lines also can be annotated to the video by visual AR technology. The current surround view system is a SEooC of driving automation level 0. However, Xylon plan to develop automated driving system features in higher levels for the future versions of the system.

Assumptions on the scope of the SEooC are:

- The system can be connected to the rest of the vehicle in order to obtain speed information. In case of drawing parking path lines, steering wheel angle and information from gearbox would also be obtained to determine reverse driving.

Assumptions on functional requirements of the SEooC are:

Figure 4.10: Sample images from 3D videos provided in surround view system [54]

- The system is enabled either at low speed or it can be activated manually by the driver.
- The system is disabled either when moving above some speed threshold or it can be deactivated by driver.

Assumptions on the functional safety requirements allocated to the SEooC are:

- The system does not activate the function at high vehicle speed automatically.
- The system does not deactivate the functionality at low speed automatically.

4.4.4 Case Study Execution: System Modeling

This subsection reports on how we model the described system in Subsection 4.4.3 using extended SafeConcert.

Subsection 4.4.3 provides the required information for the first step of the risk assessment process, which is identifying the entities for defining composite components. Based on the selected case study explained in Subsection 4.4.3, organization, driver and vehicle containing an automotive surround view system are three composite components of this system. In this subsection, we provide information for the second and third steps of risk assessment process.

Important aspects of each entity are modeled as subcomponents of each composite component. For socio entities, the important aspects are selected from extended modeling elements explained in Subsection 4.1 and for vehicle,

which is a technical entity the important aspects are based on system description.

- Important aspects of road transport organization (selected from Figure 4.1):
  - *Organization and regulation AR adoption*: it refers to upgrading rules and regulations of road transport organization based on AR technology.
  - *Condition*: it refers to road condition.
  - *Monitoring and feedback*: it refers to the monitoring task and feedback provided by organization.
- Important aspects of driver (selected from Figure 4.2):
  - *Surround detecting*: it is an AR-extended function, because driver can detect surround environment through AR technology.
  - *Supported deciding*: it is an AR-extended function, because driver can decide with the support of AR technology.
  - *Executing*: it is human executing function.
  - *Interactive experience*: it is an AR-caused factor, because AR provides interactive ways for enhancing user experience.
  - *Social presence*: it is an AR-caused factor, because AR may decrease social presence and lead to human failure.
- Important aspects of vehicle containing surround view system (selected based on system description received from Xylon Company):
  - A set of speed sensors: each sensor is a hardware for providing speed of the vehicle based on its movement.
  - A set of cameras: each camera is a hardware for providing raw data for a video receiver. Usually there are four cameras that can be attached to four sides of the car.
  - Switch: switch is a hardware for receiving on/off command from driver. It is also possible to send on/off command automatically based on driving requirement.
  - Peripheral controller: peripheral controller includes hardware and driver for receiving user inputs such as speed and on/off command and for sending them to user application implementation.

- A set of video receivers: each video receiver includes a hardware and a driver. Its hardware is used for transforming raw data to AXI-stream based on the command from its driver implementation.
- Video storing unit: video storing unit includes a hardware and a driver. Its hardware is used for receiving AXI-stream and storing it to the memory by means of DDR memory controller based on the command received form its driver.
- DDR controller: DDR controller is a hardware for accessing DDR memory, which stores video in DDR memory and provides general memory access to all system IPs.
- Video processing IP: Video processing IP includes hardware and driver for reading prepared data structures and video from memory, for processing video accordingly and finally for storing the processed video to memory through DDR controller. The prepared data is stored to memory by video processing IP driver based on the data structures received from memory.
- Display controller: Display controller includes hardware and driver for reading memory where processed video is stored and for converting it to the format appropriate for driving displays.
- Processing unit: processing unit includes hardware and software, which its software contains all the software and drivers of all other IPs. The software also contains user application implementation and video processing engine implementation. User application implementation receives inputs from peripheral unit and controls operation of all IPs by means of their software drivers. Video processing engine implementation prepares data structures to be stored in DDR memory through DDR controller.

Figure 4.11 provides an overview of integration of some of important aspects of the human, organization and vehicle.

In Figure 4.12, we show how this AR-equipped socio-technical system is modeled using extended SafeConcert. Driver is composed of five subcomponents. Driver has four inputs and two of its inputs are from system inputs with the names human detection input (HDI) and human communication input (HCI). Two other inputs are from organization and surround view system. We consider *interactive experience* and *social presence* as two subcomponents of human component, which are influencing factors on human functions. *Interactive experience* affects on *supported deciding* and is affected by *surround detecting*. *Social presence* affects on human *executing*. Driver output, which is output of the system is human action shown by HA.

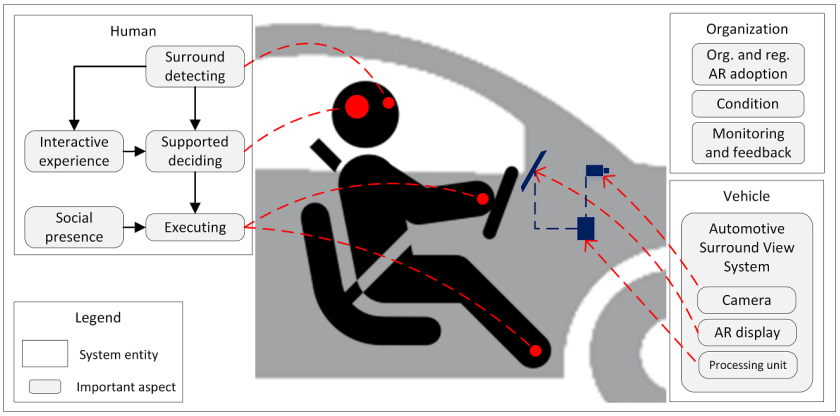


Figure 4.11: Integration of the human, organization and vehicle important aspects [54]

Organization and regulation AR adoption, *condition* and *monitoring and feedback* are three subcomponents of organization composite component. Organization component receives input from system, which represents influences from regulation authorities on the organization (REG).

Vehicle is also modeled with three inputs including user command shown by CMD, vehicle movement shown by VMV and camera input shown by CAM. Green color is used to show the AR-extended modeling elements used in this system.

4.4.5 Case Study Execution: System Analysis

This subsection reports on the analysis of the system using AR-related extensions. We follow the five steps of Concerto-FLA analysis technique explained in subsection 2.5.2 for system analysis.

1. First step is provided in Figure 4.12. We explained how the system is modeled in Subsection 4.4.4.

2. Second step is shown by providing FPTC rules, which are used for linking possible failure modes on the input of each component to the possible failure modes on the output. ”IP.variable → OP.variable” shows propagational behavior of the component, which means that any failure mode in its input is

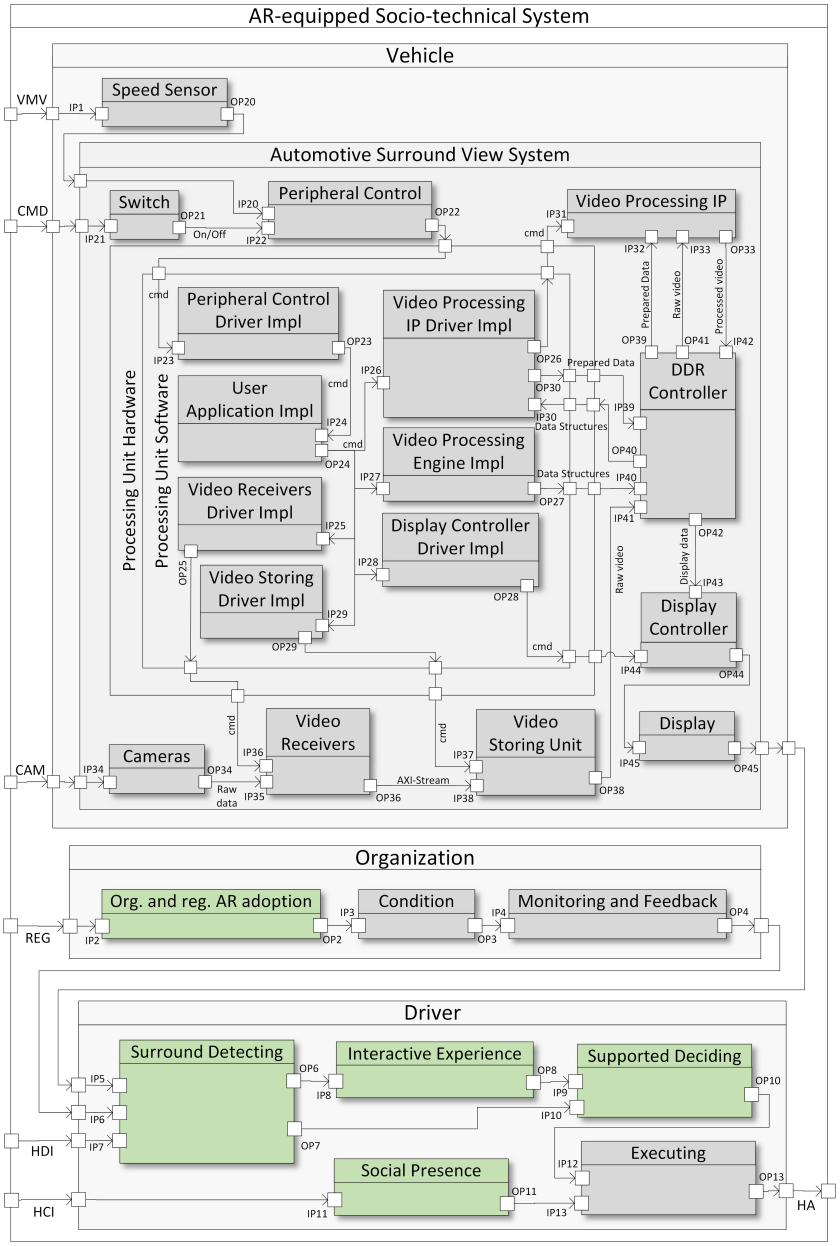


Figure 4.12: AR-equipped socio-technical system modeling [54]

propagated to its output. FPTC rules of modeled subcomponents are shown in Figures 4.13-4.16. There is one box for each component. The left part of the box shows the name of the component. The right part of the box shows possible failure modes in the input (up left), possible failure modes in the output (up right) and FPTC rules (bottom). Based on dependability-related terminology in literature such as [21], [98] and [22], we consider omission, commission, etc. as failure modes. However, these are named failures in FPTC terminology.

In this paragraph, we explain how the possible failure modes at input and output are identified/defined in Figures 4.13-4.16. For example, the camera takes in input a raw image. Based on the definitions of failure modes in Subsection 2.1, omission and valueSubtle are the possible failure modes for the case of camera. The reason for having omission as a possible failure mode at input is the possibility of an occlusion in front of the camera, which prevents receiving raw image as input. The reason for having valueSubtle as possible failure mode at input is the possibility of intervene, which leads to receiving input not in the expected range. For example, when image is blurred because of foggy weather. Possible failure modes at output can be obtained by considering the possible input failure modes in the FPTC rules. Defining the FPTC rules are explained in the next paragraph.

In this paragraph, we explain how the FPTC rules are defined in Figures 4.13-4.16. FPTC rules show how the component behaves. For example, the camera would not produce any failure, but if the input image is not in the expected range, then the output would not be in the expected range either. Moreover, if the input is not provided when expected, then the output would not be provided when expected. Thus, the camera propagates possible input failure modes to the output and it does not behave as source, sink and transformational (explained in Subsection 2.5.2).

In scenarios, we may change some components' failure behavior to source based on assumptions related to that scenario. For example, if we assume that an AR-related component is producing failure, then we need to change its failure behavior to source and update its FPTC rules.

3. Third step is to consider failure modes in inputs of the system to calculate failure propagation. In this case study, we inject noFailure to four inputs of the system, because we aim at analyzing system for scenarios that failure is originating from our modeled system and more specifically from our AR-related part of the system.

4. Fourth step is calculating the failure propagations. We consider three scenarios and show the analysis results in Figures 4.17 - 4.19.

5. Last step is back propagation of results. Interpretation of the back-propagated results can be used to make decision about design change or defining safety barrier, if it is required.

Scenario 1:

- **Description of the scenario:** In this scenario, we assume that failure in the system is emanated from the technical part of the system. We assume video processing IP produces processed video incorrectly. For example, we assume that the expected visual mark for parking lot striping is assigned on an incorrect position (value failure mode). As a consequence, the driver cannot detect the surround environment correctly and decides and executes incorrectly (value failure mode).

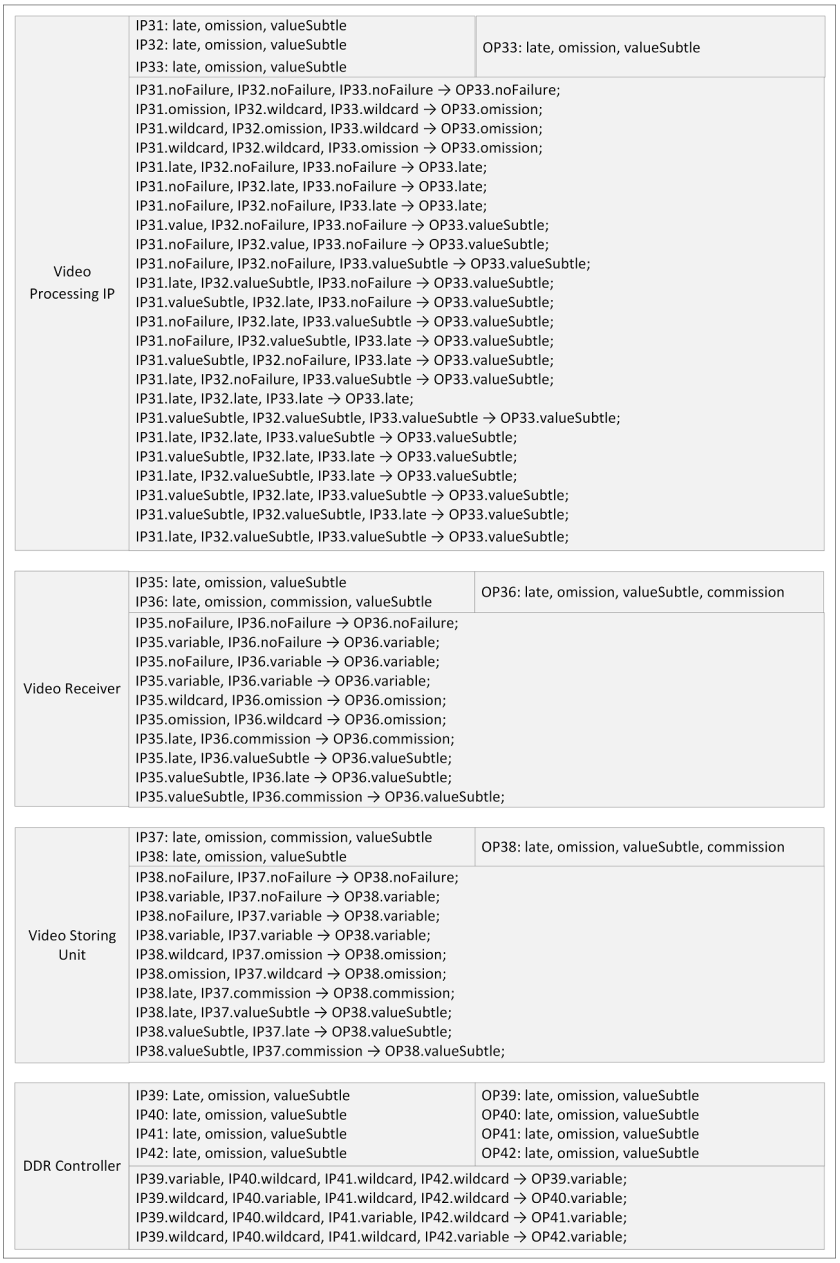
- **Modeling failure behavior:** We show the failure propagation with underlined FPTC rules, which are the rules that are activated, shown in Figure 4.17. In this scenario, video processing IP behaves as source and while its inputs are noFailure, it produces valueSubtle failure mode in its output. This activated rule is shown on its subcomponent. DDR controller, display controller and display subcomponents behave as propagational and propagate valueSubtle from inputs to outputs.

- **Analysis of system behavior:** ValueSubtle failure mode in IP5 means that displayed information on the display is not correct. ValueSubtle propagates to *surround detecting*, *interactive experience* and *supported deciding* and it transforms to valueCoarse in *executing*. The reason for this transformation is that if there is value failure mode in *executing* function, it can be detected by user, which means valueSubtle transforms to valueCoarse. We show the failure propagation by blue color of the underlined FPTC rules.

- **Interpreting the results:** Based on back propagation of the results, we can explain how the rules have been triggered. ValueCoarse on OP13 is because of valueSubtle on IP12. ValueSubtle on IP12 is because of valueSubtle on OP10 and we continue this back propagation to reach a component originating the failure, which is component with inputs IP31, IP32 and IP33. This component is video processing IP.

Name of the component	Possible failure modes at input	Possible failure modes at output
	FPTC rules	
Camera	IP34: omission, valueSubtle	OP34: omission, valueSubtle
	IP34.variable → OP34.variable;	
Speed Sensor	IP1: omission, valueSubtle	OP20: omission, valueSubtle
	IP1.variable → OP20.variable;	
Switch	IP21: late, omission, commission	OP21: late, commission, omission
	IP21.variable → OP21.variable;	
Peripheral Control	IP20: late, omission, valueSubtle IP22: late, omission, commission, valueSubtle	OP22: late, omission, commission, valueSubtle
	IP20.noFailure, IP22.noFailure → OP22.noFailure; IP20.variable, IP22.noFailure → OP22.variable; IP20.noFailure, IP22.variable → OP22.variable; IP20.variable, IP22.variable → OP22.variable; IP20.wildcard, IP22.omission → OP22.omission; IP20.omission, IP22.wildcard → OP22.omission; IP20.late, IP22.commission → OP22.commission; IP20.late, IP22.valueSubtle → OP22.valueSubtle; IP20.valueSubtle, IP22.late → OP22.valueSubtle; IP20.valueSubtle, IP22.commission → OP22.valueSubtle;	
Peripheral Control Driver Imp	IP23: late, omission, commission, valueSubtle	OP23: late, omission, commission, valueSubtle
	IP23.variable → OP23.variable;	
User Application Imp	IP24: late, omission, commission, valueSubtle	OP24: late, omission, commission, valueSubtle
	IP24.variable → OP24.variable;	
Video Receiver Driver Imp	IP25: late, omission, valueSubtle, commission	OP25: late, omission, commission, valueSubtle
	IP25.variable → OP25.variable;	
Video Processing Engine Imp	IP27: late, omission, valueSubtle	OP27: late, omission, valueSubtle
	IP27.variable → OP27.variable;	
Display Controller Driver Imp	IP28: late, omission, valueSubtle, commission	OP28: late, omission, commission, valueSubtle
	IP28.variable → OP28.variable;	
Video Storing Driver Imp	IP29: late, omission, valueSubtle, commission	OP29: late, omission, commission, valueSubtle
	IP29.variable → OP29.variable;	
Video Processing IP Driver Imp	IP26: late, omission, commission IP30: late, omission, valueSubtle	OP26: late, omission, commission, valueSubtle OP30: late, omission, valueSubtle
	IP26.noFailure, IP30.noFailure → OP26.noFailure, OP30.noFailure; IP26.variable, IP30.variable → OP26.variable, OP30.variable; IP30.valueSubtle, IP26.late → OP30.valueSubtle, OP26.late; IP30.wildcard, IP26.omission → OP26.omission, OP30.omission; IP30.omission, IP26.wildcard → OP30.valueSubtle, OP26.valueSubtle; IP30.late, IP26.commission → OP30.commission, OP26.valueSubtle; IP30.valueSubtle, IP26.commission → OP30.commission, OP26.valueSubtle;	

Figure 4.13: Modeling failure behavior of components [54]



Video Processing IP	IP31: late, omission, valueSubtle IP32: late, omission, valueSubtle IP33: late, omission, valueSubtle	OP33: late, omission, valueSubtle
	IP31.noFailure, IP32.noFailure, IP33.noFailure → OP33.noFailure; IP31.omission, IP32.wildcard, IP33.wildcard → OP33.omission; IP31.wildcard, IP32.omission, IP33.wildcard → OP33.omission; IP31.wildcard, IP32.wildcard, IP33.omission → OP33.omission; IP31.late, IP32.noFailure, IP33.noFailure → OP33.late; IP31.noFailure, IP32.late, IP33.noFailure → OP33.late; IP31.noFailure, IP32.noFailure, IP33.late → OP33.late; IP31.value, IP32.noFailure, IP33.noFailure → OP33.valueSubtle; IP31.noFailure, IP32.value, IP33.noFailure → OP33.valueSubtle; IP31.noFailure, IP32.noFailure, IP33.valueSubtle → OP33.valueSubtle; IP31.late, IP32.valueSubtle, IP33.noFailure → OP33.valueSubtle; IP31.valueSubtle, IP32.late, IP33.noFailure → OP33.valueSubtle; IP31.noFailure, IP32.late, IP33.valueSubtle → OP33.valueSubtle; IP31.noFailure, IP32.valueSubtle, IP33.late → OP33.valueSubtle; IP31.valueSubtle, IP32.noFailure, IP33.late → OP33.valueSubtle; IP31.late, IP32.noFailure, IP33.valueSubtle → OP33.valueSubtle; IP31.late, IP32.late, IP33.late → OP33.late; IP31.valueSubtle, IP32.valueSubtle, IP33.valueSubtle → OP33.valueSubtle; IP31.late, IP32.late, IP33.valueSubtle → OP33.valueSubtle; IP31.valueSubtle, IP32.late, IP33.late → OP33.valueSubtle; IP31.late, IP32.valueSubtle, IP33.late → OP33.valueSubtle; IP31.valueSubtle, IP32.late, IP33.valueSubtle → OP33.valueSubtle; IP31.valueSubtle, IP32.valueSubtle, IP33.late → OP33.valueSubtle; IP31.late, IP32.valueSubtle, IP33.valueSubtle → OP33.valueSubtle;	
Video Receiver	IP35: late, omission, valueSubtle IP36: late, omission, commission, valueSubtle	OP36: late, omission, valueSubtle, commission
	IP35.noFailure, IP36.noFailure → OP36.noFailure; IP35.variable, IP36.noFailure → OP36.variable; IP35.noFailure, IP36.variable → OP36.variable; IP35.variable, IP36.variable → OP36.variable; IP35.wildcard, IP36.omission → OP36.omission; IP35.omission, IP36.wildcard → OP36.omission; IP35.late, IP36.commission → OP36.commission; IP35.late, IP36.valueSubtle → OP36.valueSubtle; IP35.valueSubtle, IP36.late → OP36.valueSubtle; IP35.valueSubtle, IP36.commission → OP36.valueSubtle;	
Video Storing Unit	IP37: late, omission, commission, valueSubtle IP38: late, omission, valueSubtle	OP38: late, omission, valueSubtle, commission
	IP38.noFailure, IP37.noFailure → OP38.noFailure; IP38.variable, IP37.noFailure → OP38.variable; IP38.noFailure, IP37.variable → OP38.variable; IP38.variable, IP37.variable → OP38.variable; IP38.wildcard, IP37.omission → OP38.omission; IP38.omission, IP37.wildcard → OP38.omission; IP38.late, IP37.commission → OP38.commission; IP38.late, IP37.valueSubtle → OP38.valueSubtle; IP38.valueSubtle, IP37.late → OP38.valueSubtle; IP38.valueSubtle, IP37.commission → OP38.valueSubtle;	
DDR Controller	IP39: Late, omission, valueSubtle IP40: late, omission, valueSubtle IP41: late, omission, valueSubtle IP42: late, omission, valueSubtle	OP39: late, omission, valueSubtle OP40: late, omission, valueSubtle OP41: late, omission, valueSubtle OP42: late, omission, valueSubtle
	IP39.variable, IP40.wildcard, IP41.wildcard, IP42.wildcard → OP39.variable; IP39.wildcard, IP40.variable, IP41.wildcard, IP42.wildcard → OP40.variable; IP39.wildcard, IP40.wildcard, IP41.variable, IP42.wildcard → OP41.variable; IP39.wildcard, IP40.wildcard, IP41.wildcard, IP42.variable → OP42.variable;	

Figure 4.14: Modeling failure behavior of components (Cont.) [54]

Display Controller	IP43: late, omission, valueSubtle IP44: late, omission, commission, valueSubtle	OP44: late, omission, valueSubtle
	IP43.noFailure, IP44.noFailure → OP44.noFailure; IP43.variable, IP44.noFailure → OP44.variable; IP43.noFailure, IP44.variable → OP44.variable; IP43.variable, IP44.variable → OP44.variable; IP43.wildcard, IP44.omission → OP44.omission; IP43.omission, IP44.wildcard → OP44.omission; IP43.late, IP44.commission → OP44.commission; IP43.late, IP44.valueSubtle → OP44.valueSubtle; IP43.valueSubtle, IP44.late → OP44.valueSubtle; IP43.valueSubtle, IP44.commission → OP44.valueSubtle;	
Display	IP45: late, omission, commission, valueSubtle	OP45: late, omission, commission, valueSubtle
	IP45.variable → OP45.variable;	
Org. and Reg. AR Adoption	IP2: late, omission, valueSubtle, valueCoarse	OP2: late, omission, valueSubtle, valueCoarse
	IP2.variable → OP2.variable;	
Condition	IP3: late, omission, valueSubtle, valueCoarse	OP3: late, omission, valueSubtle, valueCoarse
	IP3.variable → OP3.variable;	
Monitoring and Feedback	IP4: late, omission, valueSubtle, valueCoarse	OP4: late, omission, valueSubtle, valueCoarse
	IP4.variable → OP4.variable;	
Surround Detecting	IP5: late, omission, valueSubtle IP6: late, omission, valueSubtle IP7: omission, valueSubtle, late	OP6: late, omission, valueSubtle OP7: late, omission, valueSubtle
	IP5.noFailure, IP6.noFailure, IP7.noFailure → OP6.noFailure, OP7.noFailure; IP5.omission, IP6.wildcard, IP7.wildcard → OP6.omission, OP7.omission; IP5.wildcard, IP6.omission, IP7.wildcard → OP6.omission, OP7.omission; IP5.wildcard, IP6.wildcard, IP7.omission → OP6.omission, OP7.omission; IP5.late, IP6.noFailure, IP7.noFailure → OP6.late, OP7.late; IP5.noFailure, IP6.late, IP7.noFailure → OP6.late, OP7.late; IP5.noFailure, IP6.noFailure, IP7.late → OP6.late, OP7.late; IP5.valueSubtle, IP6.noFailure, IP7.noFailure → OP6.valueSubtle, OP7.valueSubtle; IP5.noFailure, IP6.valueSubtle, IP7.noFailure → OP6.valueSubtle, OP7.valueSubtle; IP5.noFailure, IP6.noFailure, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.valueSubtle, IP7.noFailure → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.late, IP7.noFailure → OP6.valueSubtle, OP7.valueSubtle; IP5.noFailure, IP6.late, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.noFailure, IP6.valueSubtle, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.noFailure, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.noFailure, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.late, IP7.late → OP6.late, OP7.late; IP5.valueSubtle, IP6.valueSubtle, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.late, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.late, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.valueSubtle, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.late, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.valueSubtle, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.valueSubtle, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle;	
Interactive Experience	IP8: late, omission, valueSubtle	OP8: late, omission, valueSubtle
	IP8.variable → OP8.variable;	

Figure 4.15: Modeling failure behavior of components (Cont.) [54]

Supported Deciding	IP9: late, omission, valueSubtle IP10: late, omission, valueSubtle	OP10: late, omission, valueSubtle
	IP9.noFailure, IP10.noFailure → OP10.noFailure; IP9.variable, IP10.noFailure → OP10.variable; IP9.noFailure, IP10.variable → OP10.variable; IP9.variable, IP10.variable → OP10.variable; IP9.wildcard, IP10.omission → OP10.omission; IP9.omission, IP10.wildcard → OP10.omission; IP9.late, IP10.valueSubtle → OP10.valueSubtle; IP9.valueSubtle, IP10.late → OP10.valueSubtle;	
Social Presence	IP11: late, omission, valueSubtle	OP11: late, omission, valueSubtle
	IP11.variable → OP11.variable;	
Executing	IP12: late, omission, valueSubtle IP13: late, omission, valueSubtle	OP13: late, omission, valueCoarse
	IP12.noFailure, IP13.noFailure → OP13.noFailure; IP12.late, IP13.noFailure → OP13.late; IP12.noFailure, IP13.late → OP13.late; IP12.late, IP13.late → OP13.late; IP12.valueSubtle, IP13.noFailure → OP13.valueCoarse; IP12.noFailure, IP13.valueSubtle → OP13.valueCoarse; IP12.valueSubtle, IP13.valueSubtlev → OP13.valueCoarse; IP12.wildcard, IP13.omission → OP13.omission; IP12.omission, IP13.wildcard → OP13.omission; IP12.late, IP13.valueSubtle → OP13.valueCoarse; IP12.valueSubtle, IP13.late → OP13.valueCoarse;	

Figure 4.16: Modeling failure behavior of components (Cont.) [54]

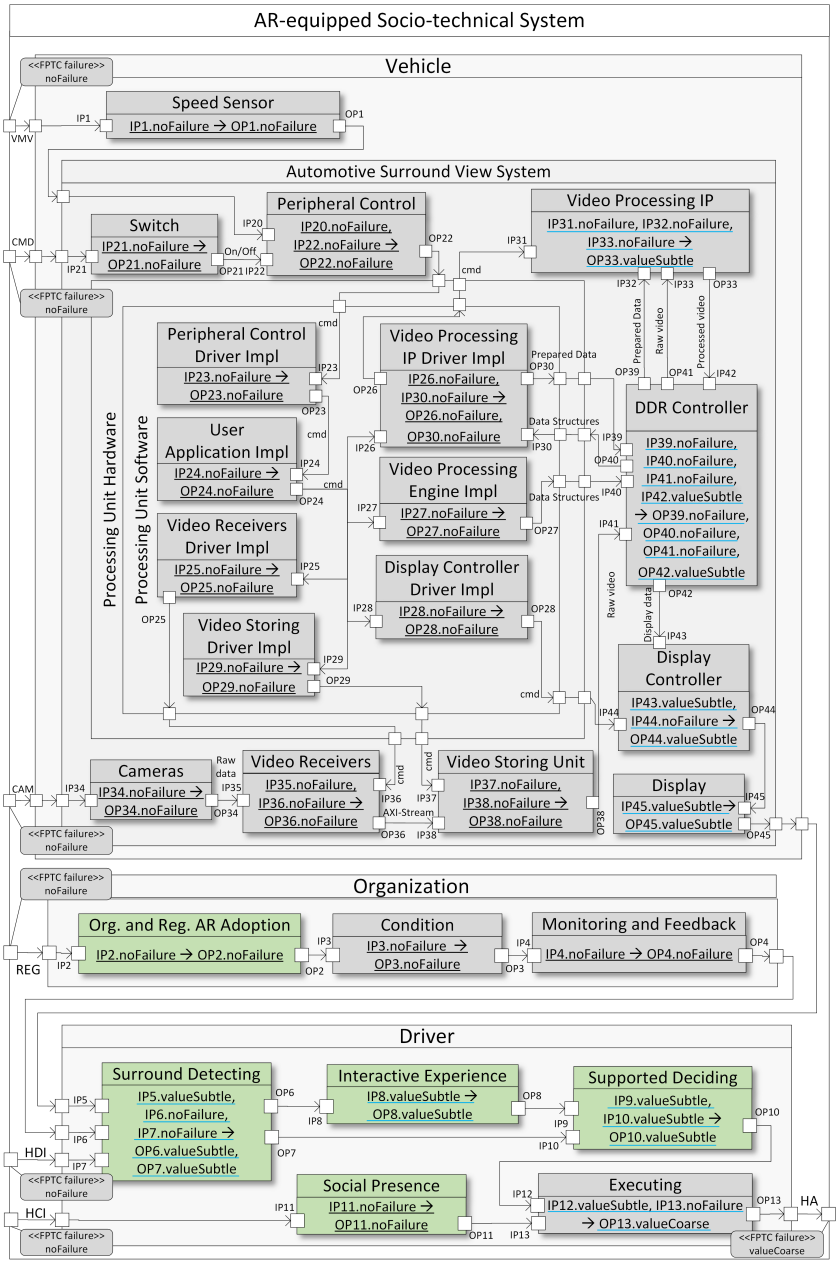


Figure 4.17: Analyzing AR-equipped socio-technical system (Scenario1) [54]

The analysis results can be helpful in hazard identification and categorization. Since the reason for system failure is a technical component, functional safety is addressed by ISO 26262.

In this case, unintended displayed information is the identified hazard and the reason is failure in video processing IP. System failure in this scenario would lead to light accident and light injuries. The reason is that the speed is not usually high while parking the car. Based on the explanation in Section 2.4 and Figure 2.5, severity in this case is S1. Class of exposure is E4, because probability of exposure is more than 10%. It means that it is more than 10 percent probable that a driver be exposed to parking situation while driving a car. Finally, class of controllability is C2 or normally controllable. It means that more than 90% of drivers can control this situation. Therefore, ASIL level for this case is A, based on Figure 2.5. Since ASIL level is A, then we should define safety goal, functional and technical safety requirements in order to overcome this risk. For example, for this scenario safety goal, functional safety requirement and technical safety requirement can be defined as follows to prevent failure in processing unit IP:

- **Safety goal**: The driver shall be notified, if there is failure in processing.
- **Safety requirement**:
  - **Functional safety requirement**: A monitoring component should be used to check the processing actively.
  - **Technical safety requirement**: Monitoring function should check the processing output every 10ms.

After interpreting the results and providing safety requirements, system design would be updated. Then, failure behavior can also be updated and failure propagation analysis can be repeated for another iteration.

Scenario 2:

- **Description of the scenario:** In this scenario, we assume that the technical part of the system works without failure, but driver doesn't have interactive experience. For example, it is the first time driver is working with systems containing AR and he/she can not understand the meaning of AR notations. Therefore, driver would decide and execute incorrectly.
- **Modeling failure behavior:** We show the failure propagation with underlined FPTC rules, which are the rules that are activated, shown in Figure 4.18.

Surround view sub-components behave as propagational and propagate no-Failure from inputs to outputs. *Interactive experience* behaves as source and while its input is noFailure, it has omission failure mode on its output. This activated rule is shown on this component.

- **Analysis of system behavior:** Omission failure mode in *interactive experience* transforms to valueSubtle in *supported deciding*, because lack of interactive experience causes wrong decision and in *executing*, it transforms to valueCoarse. Similar to the first scenario, the reason for this transformation is that if there is value failure mode in *executing* function, it can be detected by user, which means valueSubtle transforms to valueCoarse.
- **Interpreting the results:** Based on back propagation of the results, we can explain how the rules have been triggered. ValueCoarse on OP13 is because of valueSubtle on IP12. ValueSubtle on IP12 is because of valueSubtle on OP10 and we continue to IP8, which is related to *interactive experience* component.

  In this scenario, we considered failure in AR-related part of the system and since it refers to limitation in intended functionality (SOTIF related hazards), we do not determine ASIL level. If the expected severity and controllability of the scenario is higher than S0 and C0 respectively, we need to consider SOTIF safety process [99]. As we explained in the previous scenario, severity and controllability are higher than S0 and C0. Lack of interactive experience leads to system failure and incorrect deciding is the identified hazard. Safety goal and safety requirement can be defined as follows. Since the failure is not emanated from technical part of the system, we do not need to specify technical safety requirement:

  – **Safety goal**: Interactive experience shall be provided for the driver.
  – **Safety requirement**: The Company should provide a training video for all drivers at the first time of using the system.

  After applying the requirements the behavior of this component would change from source to other types and analysis can be repeated.

  It is not possible to detect risk originated from failure in interactive experience, without using the proposed representation constructs, because using these representation constructs or modeling elements provides the possibility to analyze their failure propagation and provides the possibility to analyze effect of these failures on system behavior. Then based on analysis results decision about design change or fault mitigation mechanisms would be taken.

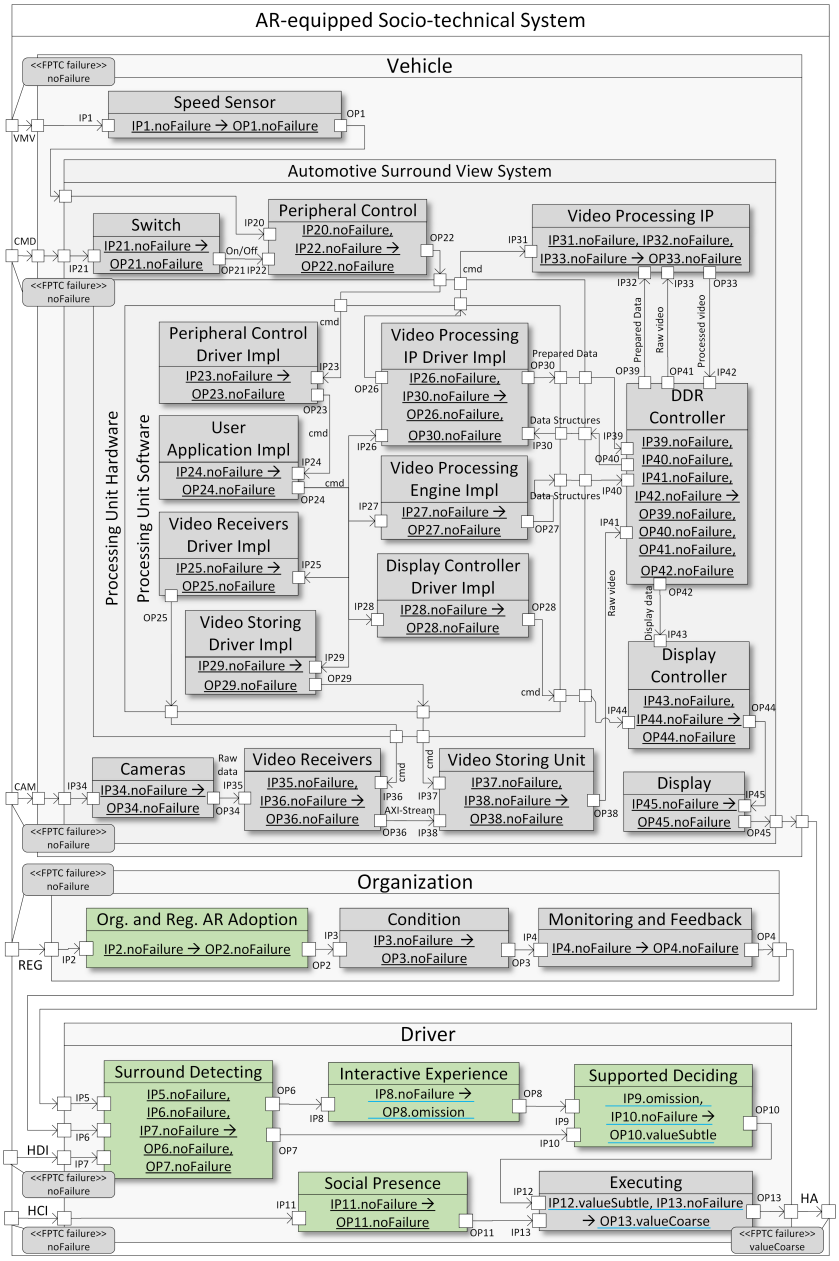


Figure 4.18: Analyzing AR-equipped socio-technical system (Scenario2) [54]

Scenario 3:

- **Description of the scenario:** In this scenario, we assume that road transport organization has not updated rules and regulations based on AR technology, which is a limitation in intended functionality. For example, parking lot striping is not updated to be used by AR applications and it affects on road condition, but *monitoring and feedback* component detect this problem and provide a feedback to driver. This feedback would be a visual text alarm showing that there is a problem in AR information. Therefore, driver will not depend on shown result and try to decide and execute correctly.
- **Modeling failure behavior:** We show the failure propagation with underlined FPTC rules, which are the rules that are activated, shown in Figure 4.19. Similar to the previous scenario, surround view subcomponents behave as propagational and propagate noFailure from inputs to outputs. *Organization and regulation AR adoption* behaves as source and while its input is noFailure, it has omission failure mode on its output. This activated rule is shown on this component. *Monitoring and feedback* component behaves as sink and while its input is omission, it has noFailure on its output.
- **Analysis of system behavior:** Omission failure mode propagates from *organization and regulation AR adoption* to *condition* and *monitoring and feedback*. In *monitoring and feedback* it will transform to noFailure. Then, noFailure is propagated from *surround detecting* to *interactive experience*, *supported deciding* and *executing*.
- **Interpreting the results:** In this scenario, system output is provided without failure. Thus, there is no hazard and no safety requirement is required.

4.4.6 Compliance with ISO 26262 and SOTIF

Proposed risk assessment activities support several ISO 26262 and SOTIF development process activities, shown in Table 4.2. Defining involved entities in step 1 and important aspects of each entity in step 2 supports *Item definition* activity of ISO 26262 standard and *functional and system specification* of SOTIF standard. In step 1 and 2 of our proposed activities, components and subcomponents are defined, which can support provision of items and functional specification. System model including all components and subcomponents support provision of system specification. Provided component-based model in step 1 and 2 of our proposed framework can be used as work products expected by the standards.

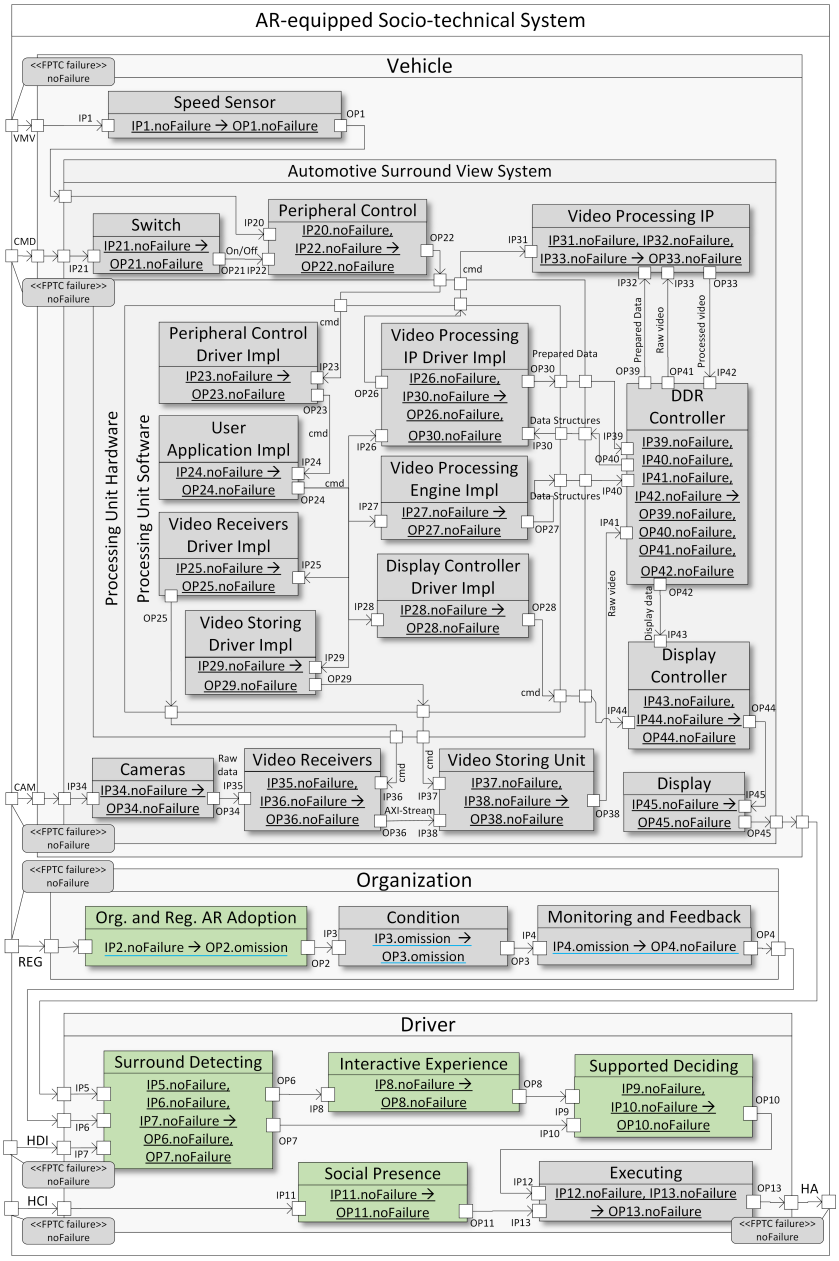


Figure 4.19: Analyzing AR-equipped socio-technical system (Scenario3) [54]

Modeling important aspects of each entity, analyzing their behavior and analyzing system behavior supports *hazard analysis and risk assessment (HARA)* of ISO 26262 standard and *SOTIF related hazard identification and risk evaluation* and also *identification and evaluation of triggering events* of SOTIF standard. In *hazard analysis and risk assessment* of ISO 26262, the aims are to identify the hazards and formulating safety goals. Step 2, 3 and 4 of our proposed activities support hazard identification by modeling failure propagation and by providing analysis results of different scenarios. These results support formulating safety goals to avoid unaccepted risks. In *SOTIF related hazard identification and risk evaluation*, the aims are identifying and evaluating SOTIF related hazards and their consequences. Modeling and analyzing activities in step 2, 3 and 4 provide the support for identification and evaluation of SOTIF related hazards and their consequences. For example, failing to pay attention leads to deciding incorrectly, which is a SOTIF related hazard and it leads to executing incorrectly. The modeling elements, used in step 2 and 3, provide the possibility to model and analyze paying attention, deciding and executing functions. Analysis in step 4 also provides the consequences at system level. Provided model in step two, three and analysis results in step four can be used as work products expected by the standards.

Analyzing system behavior in step 4 also supports defining functional and technical safety requirements, which are used in *functional and technical safety concept* of ISO 26262 standard and it also supports *functional modification to reduce SOTIF risk* of SOTIF standard. In addition, analysis results are based on considering various scenarios, which support *verification test* in ISO 26262 and *verification of the SOTIF*. Required work products for verification test in ISO 26262 and SOTIF standards can be prepared based on analysis results in step four of our proposed framework.

This contribution is presented in Paper B (see Chapter 8).

4.5 Applying the Framework in Robotic Domain

In this section, we apply our framework in a different domain to evaluate the applicability and effectiveness of our framework in a new domain. To do that, we choose a digitalized socio-technical factory system, focusing on the human-robot collaboration for a realistic diesel engine assembly task using AR-based user interface in an organization affected by organizational changes. Then, we discuss about the extent the robotic safety standards are supported (to demonstrate the applicability of the framework in the robotic domain), the extent

Table 4.2: Risk assessment activities of our provided framework and supported ISO 26262 and SOTIF development process activities [54]

The proposed activity	**ISO 26262 activity**	**SOTIF activity**
Defining involved entities and important aspects of each entity (Step1 and 2)	Item definition	Functional and system specification
Defining important aspects of each entity, analyzing its behavior and system behavior (Step2, 3 and 4)	HARA	SOTIF related hazard identification and risk evaluation and Identification and evaluation of triggering events
Analyzing system behavior (Step 4)	Functional safety concept	Functional modification to reduce SOTIF risk
Analyzing system behavior (Step 4)	Technical safety concept	Functional modification to reduce SOTIF risk
Analyzing system behavior (Step 4)	Verification test	Verification of the SOTIF

the conceptualizations provided by the framework are effective to capture the essential information for risk assessment in socio-technical robotic manufacturing and the extent the risk assessment is effective with respect to AR and organizational changes.

4.5.1 Research Methodology

This subsection describes the research method that we used for conducting and reporting our study. The research method is based on the guidelines for conducting and reporting case studies by Runeson and Höst [97]. There are five main steps for conducting and reporting a case study:

1. **Case study design:** In this step, objectives should be defined and the case study should be planned. In order to define objectives, a set of research questions can be defined. In order to plan the case study, the case (object of study) and case study protocol should be defined.

2. **Preparation for data collection:** In this step, procedures and protocols for data collection should be defined. The principal decisions on methods for collecting data are taken in the design step (defining the case study protocol) and the details of procedures are defined in this step.

3. **Collecting evidence:** In this step, the case study should be executed and data should be collected according to case study protocol. It is important to have several data sources to limit the effects of one data source interpretation. The collected data should provide the ability to address research questions.

4. **Analysis of collected data:** In this step, the collected data should be analyzed by defining an analysis methodology. There would be conclusions from the analysis such as recommendations for future studies.

5. **Reporting the results:** In this step, the results should be reported. The results include answers to the research questions, conclusions, suggestions for future research direction. Threats to validity can be analyzed with proposing countermeasures to reduce them.

we regroup these steps into 3 main activities as follows. Activity one, called planning the study, includes: step 1 and step 2; activity two, called executing the study, includes: step 3, step 4 and activity three, called discussion on the results and their validity which refers to step 5. We explain execution of these activities in the following sections.

4.5.2 Planning the Study

Objectives

We aim at evaluating the applicability and effectiveness of the FRAAR framework for the purpose of assessing risk of an AR-equipped socio-technical system in human robot collaboration domain with respect to considering effects of AR and organizational changes and support for standards. Based on this objective, we define the following research questions (Qs):

1. Q1: To what extent are the related safety standards in the robotic domain supported (which demonstrates the applicability of the framework in robotic domain)?

2. Q2: To what extent are the conceptualizations provided by the framework effective to capture the essential information for assessing risk in the socio-technical robotic factory?

3. Q3: To what extent is the risk assessment effective with respect to capturing factors related to effects of AR and organizational changes?

Based on these research questions, we define metrics for characterizing and answering the research questions.

Metrics based on Qs:

1. M1: Percentage of supported risk assessment steps provided by standards.
2. M2: Percentage of covered typical human robot interaction failures.
3. M3: Percentage of extensions on identified risk sources with respect to effects of AR and organizational changes.

We show the defined goal, questions and metrics based on GQM model in Figure 4.20.

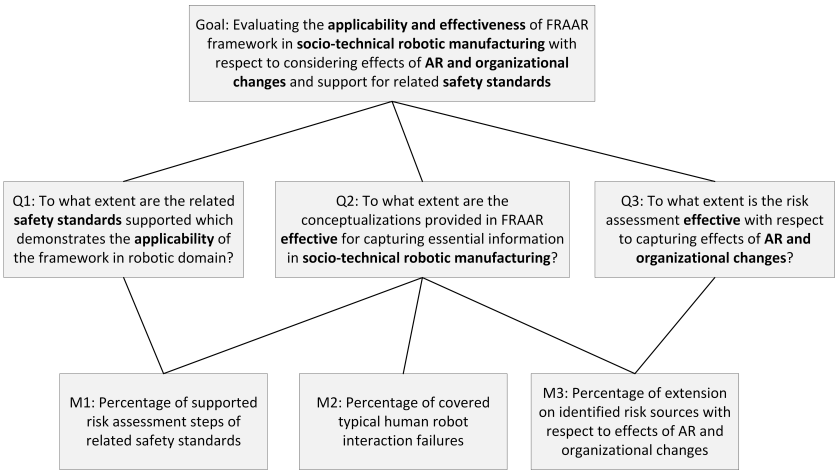


Figure 4.20: Defined goal, questions and metrics using GQM method

Selected Case

In this subsection, we describe an AR-equipped socio-technical system which we selected based on [100] and a taxonomy of typical failures in human robot collaboration proposed in [101].

The system contains the following entities:

- **Technical entities:**

- A robot collaborating with the human worker for the engine assembly task.
- An AR user interface for illustrating information such as instructions and robot status to the human worker.

- **Socio entities:**
  - A human worker who is working in local diesel engine manufacturing company.
  - Diesel manufacturing organization which is responsible for providing rules and regulations, proper work conditions and etc.

Interactive AR-based user interface (UI) proposed in [100] provides capabilities to improve safety of collaboration between human and robot in diesel manufacturing. There are two types of implementations for the AR-based UI: using projector-mirror setup (Figure 4.21) or wearable AR gear (HoloLens) (Figure 4.22). In projector-mirror setup the AR indications are shown on the table around the robot, while in wearable AR HoloLens the indications are shown on the display of the headset used by the human worker. We focus on projector-mirror setup.

Figure 4.21: Robot and AR-based UI using projector-mirror [100]

The AR-based UI provides six main indications: 1) danger zone which is the region the worker should avoid, 2) changes of human zone, 3) GO and STOP button for starting and stopping the robot, 4) CONFIRM button for verifying and changing of regions, 5) ENABLE button for enabling GO and CONFIRM buttons and 6) a graphical display box containing the instructions and status of the robot.

The considered task is based on [100] which is a part of a real engine assembly task taken from a local company. It contains five sub-tasks which one

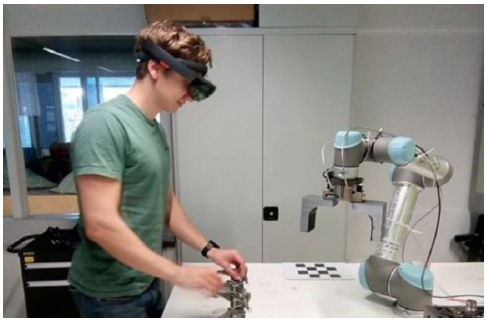


Figure 4.22: Robot and Human using AR-based wearable HoloLens UI [100]

of them (sub-task 4) is collaborative and we have the focus on that. These sub-tasks are: 1) installing 8 rocker arms (by human), 2) installing the engine frame (by robot), 3) Inserting 4 frame screws (by robot), 4) installing the rocker shaft (bringing and providing required force by robot and accurate positioning by human), 5) inserting the nuts on the shaft (by robot). The rocker shaft weights 4.3 kg and it is helpful to use a robot for bringing it. However, it is also crucial to consider safety issues while the human is in close distance and dropping the shaft on human worker's hands would lead to serious injuries.

In [101] a taxonomy of typical failures in human-robot collaboration is provided based on a literature review conducted in the paper. Based on this taxonomy there are two main types of failures in human robot collaboration: *technical failures* and *interaction failures*. *Technical failures* are categorized to *hardware* and *software failures*. *Interaction failures* are categorized to *human errors, environment and other agents*, and *social norm violations*. *Software failures* are categorized to *design failures, communication failures* (categorized to *incorrect data, bad timing, extra data* and *missing data*), and *processing failures* (categorized to *missing events, timing and ordering, abnormal terminations* and *incorrect logic*). *Hardware failures* are categorized to *effectors, power, control* and *sensors failures*. *Human errors* are categorized to *mistakes, slips, lapses* and *deliberate violations*. *Environment and other agents failures* are categorized to *group-level judgment, working environment* and *organizational flaws*.

Study Protocol

Based on [97], there are three types of data collection techniques: first degree (researcher in direct contact with the subjects collecting data in real time such as interview), second degree (researcher collects data without interacting with the subjects such as observation) and third degree (analysis of work artifacts such as using archival data). In this study, we use the third degree data collection technique. However, we use multiple sources of evidence in order to increase trustworthiness of the work. For selecting the case containing augmented reality in a real context, we use [100] which describes an AR-equipped socio-technical system with its real-life context. In order to model technical entities, we use technical details described in the related product websites. In addition, we collect data based on Goal Question Metric method (GQM) [91] which is a goal-oriented measurement technique as we explained in Section 2.6. Based on this technique, the goal of the study is defined and then research questions are defined based on the goal to trace goal to data intended to define the goal operationally. Finally, metrics are defined based on the research questions for characterizing and answering them to achieve the goal.

4.5.3 Executing the Study

System Modeling

Based on the first step of the FRAAR framework explained in Subsection 4.3, in order to model the system, we need to identify the system entities (as we identified in Subsection 4.5.2). Then, based on the second step, we need to identify the important aspects of each entity. Important aspects are required for modeling subcomponents of each composite component representing the related entity. We identify important aspects of the robot collaborating with human using the description provided in [100] and product technical specifications in [102] and [103]. For identifying human and organization important aspects, we use the extended modeling elements of FRAAR framework shown in Figure 4.2 and Figure 4.1.

- Important aspects of robot:
  - Control box hardware: it is a hardware for receiving command from computing system and providing control commands for controlling the arm and gripper using its related software.
  - Control box software: it is a software in relation to control box hardware for providing the commands.

  - Arm: it is a hardware for receiving command from control box and providing the required movement.
  - Gripper: it is a hardware for receiving command from control box and providing the required movement.

- Important aspects of projector-mirror UI:
  - RGB-D sensor: it is a hardware for capturing color image (RGB) and depth information from the scene and providing the required information to be sent to the computing system.
  - Computing system hardware: it is a hardware in relation to the computing system software for conducting the computations.
  - Computing system software: it is a software for providing command for robot and for providing the required input for 3LCD projector using the received information from RGB-D sensor.
  - A 3LCD video projector: it is a hardware for receiving information from computing system and providing a 1920*1080 color image with 50 Hz frame rate.
  - Mirror: it is a hardware for increasing the projection area.

- Important aspects of human worker:
  - Mental state: it refers to mental state of human that may influence on human behavior. For example, there may be problem in mental state because of time pressure and it may influence on worker behavior and it may lead to wrong decision and execution.
  - Detecting: it refers to human detecting function.
  - Deciding: it refers to human deciding function.
  - Executing: it refers to human executing function.
  - Information processing: it refers to human information processing function.
  - Communicating: it refers to human communicating function (for example with other people).
  - Cultural distance: it refers to a factor related to organizational changes. For example, if there is any misunderstanding between the worker and the manager due to distance between their cultures.

  - Interactive training/experience: it refers to a factor related to AR. When AR is used in the system, it is required for the worker to have training/experience to be able to work with AR interface.
  - Conforming to rules: it refers to a human function for conforming to rules.

- Important aspects of diesel manufacturing organization:
  - Financialized strategy: it refers to a factor related to the effects of new organizational changes that causes increasing power of financial actors leading to new strategies.
  - Time pressure: it refers to a factor that may influence on human behavior, because time pressure may cause wrong decision and execution by human.
  - Condition: it refers to the condition provided by the organization.
  - Augmented environment: it refers to the environment provided by using augmented reality. For example, when a projector is used for illustrating AR information, the augmented environment is the virtual displayed information along with the physical environment of the user.
  - Resource management: it refers to managing the resource in the organization.
  - Organization and regulation AR adoption: it refers to updating rules and regulations based on changes due to AR.
  - Equipment: it refers to equipment used for performing the task.
  - Organizational process: it refers to daily corporate decisions.
  - Oversight: it refers to providing feedback for managers.
  - Digitalized task: it refers to a factor integrating effects of organizational changes. It refers to task definition provided by organization while the task is digitalized as an organizational change.

An overview of the integration of human worker, AR-based projector-mirror UI, robot and organizational factors is provided in Figure 4.23.

In Figure 4.24, we show how the considered AR-equipped socio-technical system is modeled using the extended modeling language of FRAAR framework. Human worker contains nine subcomponents with four inputs. Three of human inputs are from organization and one is from system input as communicating input. Interactions between different subcomponents are shown in the figure. The output of human worker is Human Action shown by HA.

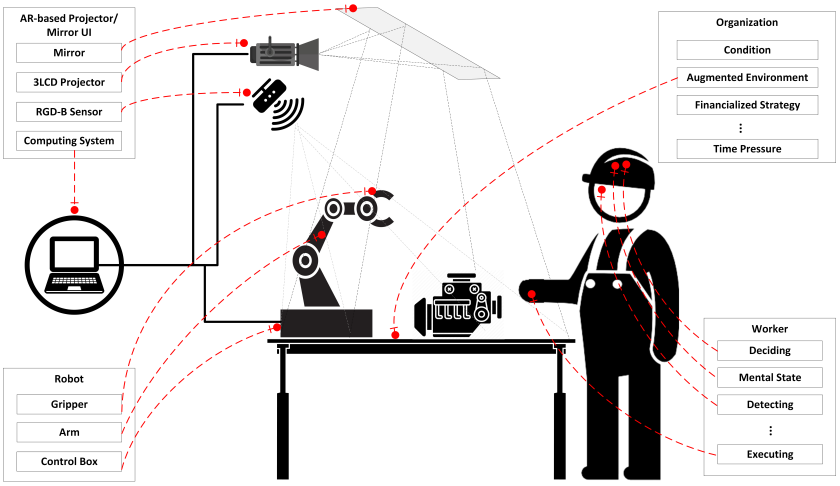


Figure 4.23: Integration of human worker, AR-based projector-mirror UI, robot and organizational factors (adapted from [104] and [105])

Robot has five subcomponents and one input coming from a computing system which contains the commands which should be executed by the robot. Output of the robot is robot action which is shown by RA. AR-based Projector-mirror UI has six subcomponents and one input which is input of the system containing the RGB-D data sensed by sensor, shown by RGB-D.

Organization has ten subcomponents and two inputs, one coming from mirror and the other input is connected to the input of the system. The input coming from the system input is influences from regulation authorities shown by REG. The organization has four outputs. One of them is connected to system output shown by OS, which is output of oversight subcomponent and provides the feedback for managers about the organization. The other three outputs are from augmented environment, time pressure and organization and regulation AR adoption, which are connected to worker inputs.

System Analysis

This subsection reports on the analysis of the system based on step 3 and step 4 of the FRAAR framework explained in Subsection 4.3. We assume that human worker and robot are collaborating to perform sub-task 4 explained in

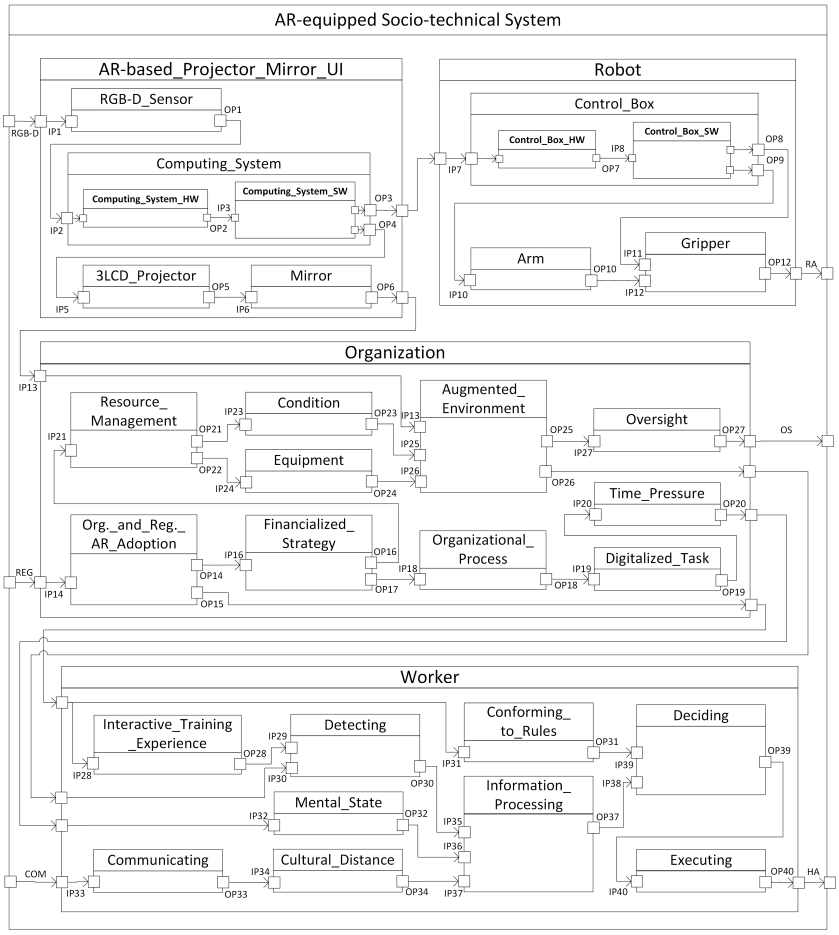


Figure 4.24: Modeling of the AR-equipped socio-technical system

Subsection 4.5.2 and we consider three scenarios as examples and we show the analysis results.

Scenario 1:

Description of the scenario: In this scenario, we assume that failure in the system is emanated from the financialized strategy. For example, because of increasing power of financial actors, new strategies are assigned to increase production. This can lead to changes on definitions of organization process and it causes changes in definition of the digitalized task (for example the collaboration between human and robot should be performed with higher speed). It can cause time pressure for worker. Time pressure can cause improper mental state, incorrect information processing, incorrect deciding and incorrect executing by the human worker and the human worker may move his/her hands under the rocker shaft when the robot is bringing it to install it (value failure mode). The result is a post normal accident, because it is due to new organizational changes.

Modeling failure behavior: The activated FPTC rules are underlined in Figure 4.25. In this scenario, financialized strategy behaves as source and while there is no failure on its input, it produces valueSubtle failure on its output. Organizational process, digitalized task, time pressure, mental state, information processing and deciding subcomponents behave as propagational and propagate valueSubtle from their inputs to their outputs and executing subcomponent transforms valueSubtle to valueCoarse. The reason is that value failure in executing function can be detected by user.

Analysis of system behavior: ValueSubtle failure mode on IP18 means that the there is failure in the provided financialized strategy. ValueSubtle propagates to organizational process, digitalized task, time pressure, mental state, information processing, deciding and executing. The failure propagation is shown by blue color.

Interpreting the results: Based on the back propagation of the results, we can explain how the rules are triggered. ValueCoarse on OP40 is because of valueSubtle on OP39 and it is because of valueSubtle on OP37. ValueSubtle on OP37 is because of valueSubtle on OP32 and it is because of valueSubtle on OP20. ValueSubtle on OP20 is because of valueSubtle on OP19 and it is because of valueSubtle on OP18. Finally, valueSubtle on OP18 is because of valueSubtle on OP17.

The results can be helpful to support hazard identification and analysis required by safety standards used in robotic and human robot collaboration.

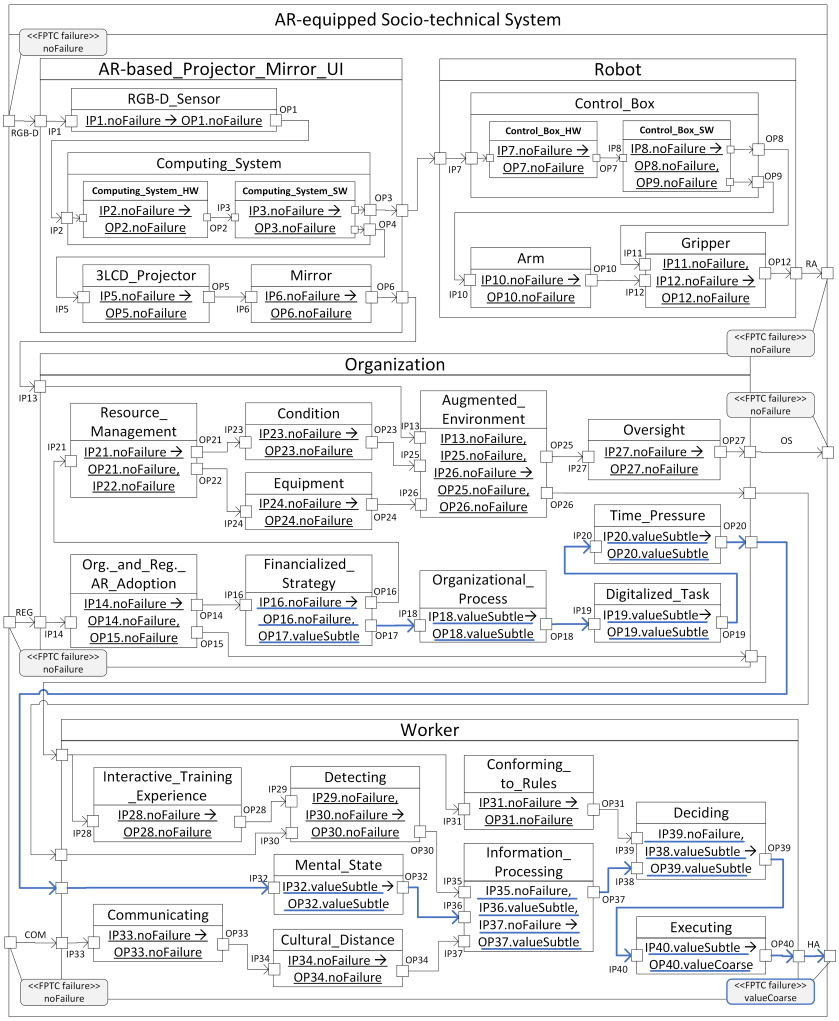


Figure 4.25: Analyzing the AR-equipped socio-technical system (Scenario 1)

In this case, unexpected movement by human is the identified hazard and the reason is improper financialized strategy leading to time pressure. System failure in this scenario would lead to sever injury since the human worker

would move his/her hands under the rocker shaft when the robot is bringing the shaft to install it. Based on the standard ISO 13849-1:2015 [57] explained in Subsection 2.4.2, severity is s2 and frequency and duration of exposure to the risk is f1 and the possibility of avoiding the risk is p1. Thus, based on Figure 2.8, required performance level is PLr = c, which is quit high.

In this case we define the following safety requirement:

- **Safety requirement**: Evaluation for financialized strategies shall be provided.

Scenario 2:

Description of the scenario: In this scenario, we assume there is failure in the augmented environment, while there is no failure in the augmented reality information provided by the projector and there is also no failure in the condition and equipment provided by the organization. However, the table used for projection of AR information has some patterns on it and it causes that the worker misread (value failure mode) the AR information shown by projector. This leads to wrong detecting, wrong information processing, wrong deciding and wrong executing by the human worker (value failure mode).

Modeling failure behavior: The activated FPTC rules are underlined in Figure 4.26. In this scenario, augmented environment behaves as source and while there is no failure on its inputs, it produces valueSubtle failure on its outputs. Oversight, detecting, information processing and deciding subcomponents behave as propagational and propagate valueSubtle from their inputs to their outputs and executing subcomponent transforms valueSubtle to valueCoarse. The reason is that value failure in executing function can be detected by user.

Analysis of system behavior: ValueSubtle failure mode on IP30 means that the detected AR information by the user is incorrect. ValueSubtle propagates to information processing, deciding, and executing. The failure propagation is shown by blue color. ValueSubtle failure mode on IP27 means that the oversight received from the organization is not correct. However, since it is not detected by managers it is propagated as valueSubtle and it is not transformed to valueCoarse.

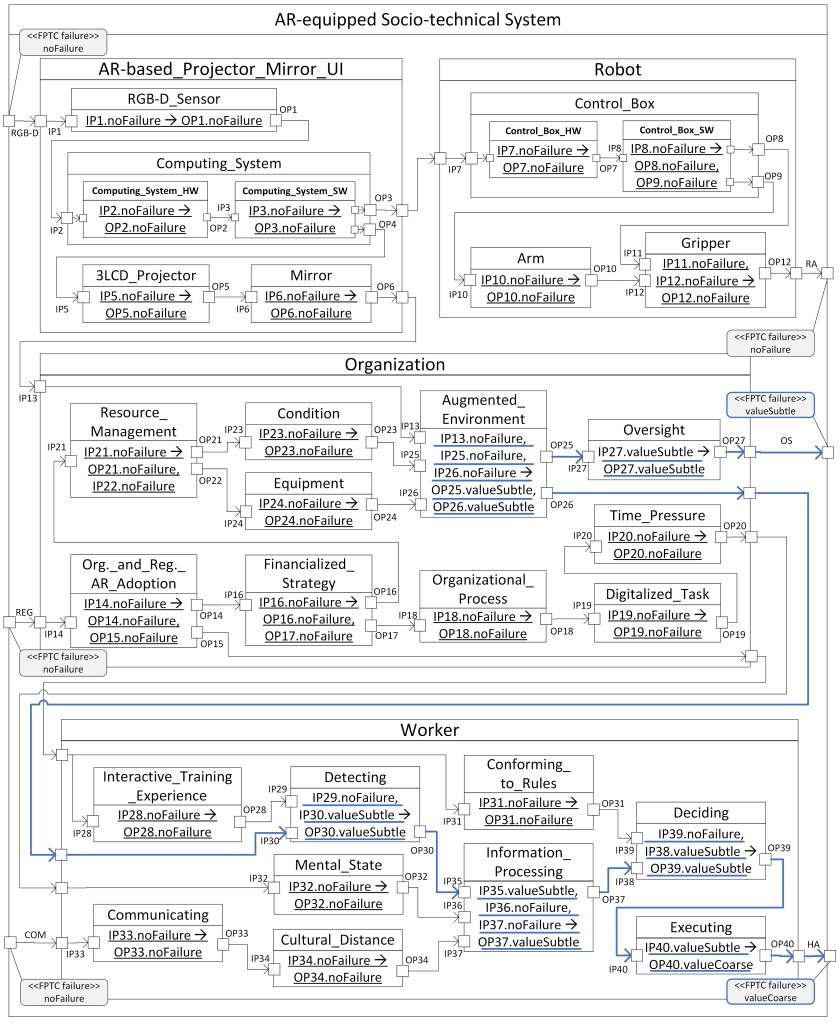


Figure 4.26: Analyzing the AR-equipped socio-technical system (Scenario 2)

Interpreting the results: Based on the back propagation of the results, we can explain how the rules are triggered. ValueCoarse on OP40 is because of valueSubtle on OP39 and it is because of valueSubtle on OP37. ValueSubtle

on OP37 is because of valueSubtle in OP30 and it is because of valueSubtle on OP26.

In this case also, unexpected movement by human (failure in human action) is the identified hazard and the reason is failure in augmented environment. Similar to the previous scenario, system failure in this scenario would lead to sever injury since the human worker may move his/her hands under the rocker shaft when the robot is bringing the shaft to install it. In this case also severity is s2 and frequency and duration of exposure to the risk is f1 and the possibility of avoiding the risk is p1. Thus, based on Figure 2.8, required performance level is PLr = c, which is quit high.

To reduce this risk, it is possible to limit the speed of the robot using mechanical safety design of the gripper. However, it would affect on system performance and efficiency. Another possibility is to provide necessary display requirements as part of safety requirements in order to prevent intervention in the augmented environment. Thus, in this case we define the following safety requirement:

- **Safety requirement**: The environment shall conform to the requirements of AR integration.

Scenario 3:

Description of the scenario: In this scenario, we assume there is failure in control box software. This can lead to failure in arm and gripper movements leading to drop of shaft (value failure mode).

Modeling failure behavior: The activated FPTC rules are underlined in Figure 4.27. In this scenario, control box software behaves as source and while there is no failure on its input, it produces valueSubtle failure on its output. Arm subcomponent behaves as propagational and propagates valueSubtle from its input to its output and gripper subcomponent transforms valueSubtle to valueCoarse. The reason is that value failure in robot movement can be detected by user.

Analysis of system behavior: ValueSubtle failure mode in IP10 means that the there is failure in the provided command from control box. ValueSubtle propagates to gripper. The failure propagation is shown by blue color.

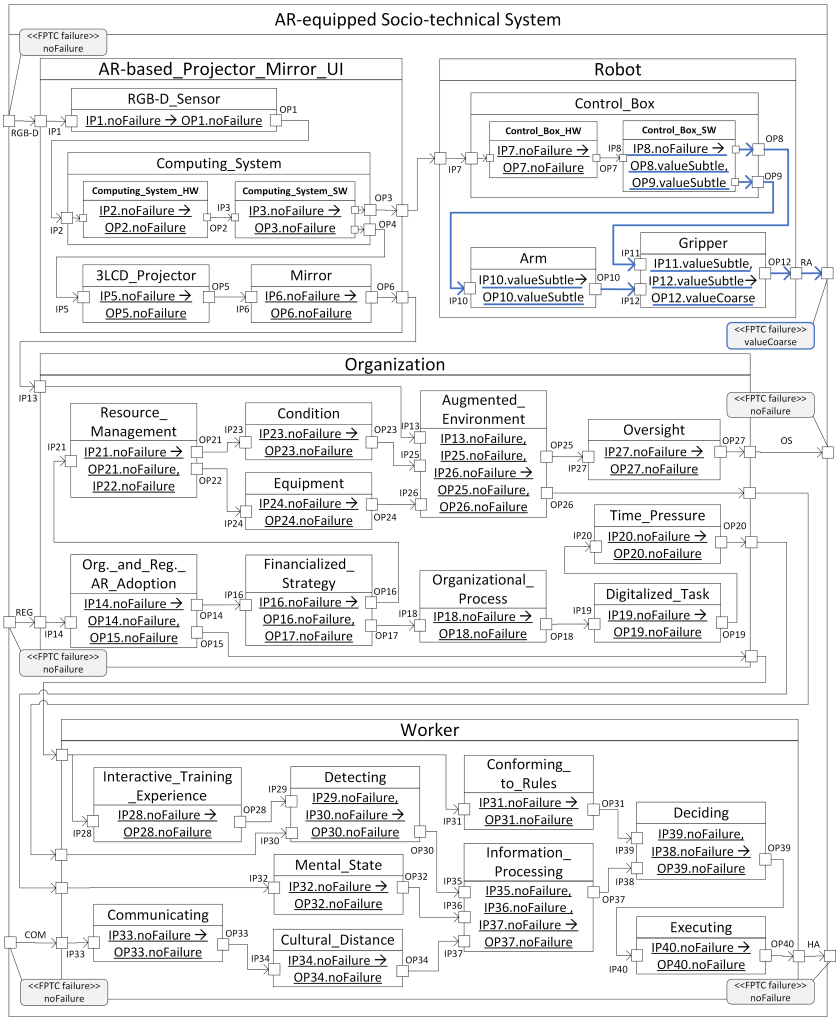


Figure 4.27: Analyzing the AR-equipped socio-technical system (Scenario 3)

Interpreting the results: Based on the back propagation of the results, we can explain how the rules are triggered. ValueCoarse on OP12 is because of valueSubtle on OP8 and OP10 and valueSubtle on OP10 is because of value-

Subtle on OP9. ValueSubtle on OP8 and OP9 is because of failure in control box software.

In this case, drop of the shaft is the identified hazard and the reason is improper provided command by control box. System failure in this scenario would lead to sever injury since the human worker's hands may be under the rocker shaft when the robot drops it. In this case severity is s2 and frequency and duration of exposure to the risk is f1 and the possibility of avoiding the risk is p2. Thus based on Figure 2.8, required performance level is PLr = d, which is high.

In this case we define the following safety requirement:

- **Safety requirement**: The computing system shall actively monitor the status of the control box.

Similarly, we can consider various other scenarios and update the system analysis based on them to investigate further risk sources, their effects and related safety requirements.

In this subsection, we applied the FRAAR framework for three example scenarios using some important aspects of socio and technical entities to illustrate how the modeling and analysis is conducted and how we can identify risk sources and related safety requirements. There is the possibility to consider more important aspects and extend the modeling and analysis. For example, in Table 4.3 and 4.4, we provide further possible risk sources in relation to socio aspects using the extended modeling elements which are integrated in the FRAAR framework. We show the risk sources in connection with effects of organizational changes or AR with gray color to be able to illustrate the extent of risk assessment extension with respect to effects of AR and organizational changes.

As it is shown in this table, there are various risk sources in relation to effects of AR and organizational changes which are identified and analyzed using the extended modeling elements.

4.5.4 Discussion on the Results and Their Validity

Discussion on the results

In this subsection, we discuss on the results and how metrics are calculated to answer the research questions to reach the goal.

Table 4.3: Identified list of dependability threats/risk sources

Identified risk sources	**Description**	**Safety requirement**
Training/experience problem	The required training is not (properly) provided for the user to perform the assembly task	Training shall be provided based on best practices
Interactive training/experience problem	The required training is not (properly) provided for the user to work with AR interface	AR-related training shall be provided based on best practices
Social presence problem	The user is fully taken by AR technology and miss the connectivity with other people and environment	The user shall receive notification through the system in case of receiving crucial communication requirement
Cultural distance problem	Communication between user and manager is affected by culture causing misinterpretation	Guidelines shall be provided for defining critical communication keywords
Physical state problem	There is injury or physical problem in the user body	Minimum level of required physical state for starting the work shall be defined
Mental state problem	There is problem in psychological state of the user	Minimum level of required psychological state for starting the work shall be defined
Deciding/ making plan problem	There is problem in deciding and making plan	Evaluation for deciding competence shall be provided
Supported deciding problem	Problem in deciding which is based on guidance provided by AR technology	Evaluation of AR notifications for supporting deciding shall be provided
Information processing problem	the user has problem in processing information	Evaluation for information processing competence shall be provided
Paying attention problem	The user has problem in paying attention during the task performance	Evaluation of AR notifications for paying attention competence shall be provided
Directed paying attention problem	There is problem in directing attention of user by AR-based UI	Evaluation of AR notifications for directed paying attention shall be provided
Identifying problem	The user has identification problem	Evaluation for identifying competence shall be provided
Perceiving problem	The user has perceiving problem	Evaluation for perceiving competence shall be provided
Surround perceiving problem	The user can not perceive surrounding environment as it is intended by AR	Evaluation of AR notifications for surround perceiving shall be provided
Sensing problem	The user has problem in sensing	Evaluation for sensing competence shall be defined
Accelerated perceiving problem	The user can not accelerate perceiving as it is intended by AR	Evaluation of AR notifications for accelerated perceiving shall be provided
Conforming to rules problem	The user has problem in conforming to rules	Evaluation for conforming to rules competence shall be provided
Executing problem	The user has problem in executing	Evaluation for executing competence shall be provided
Communicating problem	The user has problem in communicating	Evaluation for communicating competence shall be provided
Ensuring goal achievement by feedback problem	The user has problem in ensuring goal achievement by feedback	Evaluation for ensuring goal achievement by feedback competence shall be defined

Table 4.4: Identified list of dependability threats/risk sources (Cont.)

Identified risk sources	**Description**	**Safety requirement**
Resource management problem	There is problem in managing resources in the organization	Guidelines shall be provided for resource management
Organizational process problem	There is problem in daily corporate decisions	Guidelines shall be provided for organizational process
Organizational climate problem	There is problem in organization culture and policy	Guidelines shall be provided for organizational climate
Rules and regulations problem	There is problem in rules and regulations	Guidelines shall be provided for organizational rules and regulations
Oversight problem	There is problem in providing feedback for managers	Guidelines shall be provided for organizational oversight
Networked structure of organization problem	There is problem because of the networked structure of organization	Guidelines shall be provided for organizing networked structure
Supervision communication problem	There is problem in communication between the supervisors	Guidelines shall be provided for communication at supervision level
Monitoring and feedback problem	There is problem in monitoring and feedback	Guidelines shall be provided for monitoring and feedback
Organization and regulation AR adoption problem	Rules and regulations are not updated based on changes due to AR	Updates shall be provided for rules and regulations based on AR changes
Organizational industrial strategy problem	There is problem in industrial strategy defined by organization	Evaluation of organizational industrial strategy shall be provided based on best practices
Organizational financialized strategy problem	There is problem in financialized strategy defined by organization	Evaluation of organizational financialized strategy shall be provided based on best practices
Condition problem	There is problem in condition	Conditional evaluation shall be provided
Equipment problem	There is problem in equipment required for performing the task	Equipment evaluation shall be provided
Self-regulated environment problem	There is problem in self-regulated environment of the organization	Evaluation of self-regulated environment of the organization shall be provided based on best practices
Augmented environment problem	There is problem in the integration of AR and the environment	The environment shall conform to the requirements for AR integration
Time pressure problem	Time pressure is imposed by organization	Evaluation for time adequacy shall be provided
Task objectives problem	Task objectives are not (properly) defined	Guidelines shall be provided for defining task objectives
Task complexity problem	The task is too complex	Defined tasks shall be evaluated in terms of complexity
Digitalized task problem	There is a problem due to the digitalization of the task	Evaluation of digitalization shall be provided
AR guided task problem	There is a problem in the definition of the task which is guided by AR	Evaluation of definition of AR guided task shall be provided
Standardized task problem	There is a problem due to the standardization	Evaluation of standardization shall be provided

Results for the First Research Question In Subsection 4.5.3, we illustrated how the framework can be applied in robotic domain and how the standards can be used for evaluating the risk. In order to calculate the percentage of supported risk assessment steps provided by related safety standards (first metric), we show the risk assessment steps based on robotic standards explained in Subsection 2.4.2 and we show different activities of FRAAR framework which support them in Table 4.5.

Table 4.5: Supported risk assessment steps based on robotic standards by FRAAR risk assessment activities

Risk assessment step based on standard	**FRAAR risk assessment activity**
1. Risk analysis	Defining the involved entities and their important aspects, modeling their behavior and analyzing system behavior (step 1, 2, 3 and 4)
1.1. Determining the limits of the robot system	Defining the involved entities, their important aspects and their behavior (step 1, 2 and 3)
1.1.1. Defining intended use	Defining the involved entities, their important aspects and their behavior (step 1, 2 and 3)
1.1.1. Defining foreseeable misuse	Defining the involved entities, their important aspects and their behavior (step 1, 2 and 3)
1.2. Identifying the hazards and associated hazardous situations	Analyzing system behavior (step 4)
1.2.1. Considering robot related hazards	Analyzing system behavior (step 4) by considering technical hazards
1.2.2. Considering hazard related to robot system	Analyzing system behavior (step 4) by considering technical and socio hazards
1.2.3. Considering application related hazards	Analyzing system behavior (step 4) by considering technical and socio hazards
1.2.4. Identifying tasks	Defining the involved entities and their important aspects (step 1 and 2)
1.3. Estimating the risk of each hazard and hazardous situation	Analysis results from step 4
2. Risk evaluation	Analysis results from step 4
2.1. Evaluating the risk and taking decision about necessity of reducing the risk based on risk analysis results	Analysis results from step 4

As it is explained in Subsection 2.4.2, based on extended risk assessment definition provided in ISO/TS 15066:2016 [56], risk assessment contains two main activities: *risk analysis* and *risk evaluation*. The first step in risk analysis is *determining the limits of the robot system (intended use and foreseeable mis-*

use). In step 1 of the FRAAR framework shown in Figure 4.8, involved entities should be defined. Then, in step 2, important aspects of each entity should be modeled and in step 3, the behavior of each aspect is analyzed. Defining the entities, modeling their important aspects and their behavior as we illustrated in Subsection 4.5.3 can be helpful for *determining the limits containing the intended use and foreseeable misuse*. Thus, we can conclude that these activities required for risk assessment are supported by the first three steps of the FRAAR framework. The second step of risk analysis is *identifying the hazards and associated hazardous situations (considering hazards related to robot, robot system and application and identifying tasks)*. This step is also supported by the analysis results from step 4 of the FRAAR framework. Furthermore, *estimating the risk of each hazard and hazardous situation* is supported by the analysis results from step 4. In addition, as we explained in the three example scenarios in Subsection 4.5.3, we can estimate the risk of each hazard and hazardous situation. Finally, *risk evaluation and deciding about necessity of reducing the risk* is also supported by analysis results from step 4 of the FRAAR framework as it was explained for three example scenarios in Subsection 4.5.3.

As it is shown in Table 4.5, all tasks/sub-tasks defined based on standards in robotic domain are supported by FRAAR framework and it shows that 100 percent of risk assessment steps of robotic safety standards are supported using the FRAAR framework.

Results for the Second Research Question For this research question we calculate the second metric (percentage of covered typical human robot interaction failures). However, first and third metric are also in alignment with demonstrating the effectiveness of the framework in socio-technical robotic manufacturing with respect to considering effects of AR and organizational changes and support for related safety standards. In order to calculate the percentage of covered typical human robot interaction failures, we use the taxonomy proposed in [101], explained in Subsection 4.5.2. In Table 4.6, it is shown how failures are covered by the available modeling elements/failure modes/-failure behaviors in FRAAR risk assessment framework.

As it is shown in this table, 28 failures of the total 29 failures are covered by the available modeling elements, failure modes and failure behaviors in the FRAAR framework. Based on these results about 96 percent of the typical human robot interaction failures are supported by FRAAR framework, which is a generic risk assessment framework. In the following paragraphs, we explain more about details of the assignments shown in the table.

As we explained in Section 2.5, *technical failures* can be modeled using

Table 4.6: Covered typical human robot interaction failures

Typical human robot interaction failure	**Available modeling element/failure mode/-failure behaviors in FRAAR for modeling the failure**
1. Technical failures	Technical components
1.1. Software failures	Software component
1.1.1. Design failures	Equipment component
1.1.2. Communication failures	Connector
1.1.2.1. Incorrect data	Value failure mode
1.1.2.2. Bad timing	Early or late failure mode
1.1.2.3. Extra data	Commission failure mode
1.1.2.4. Missing data	Omission failure mode
1.1.3. Processing failures	Source failure behavior
1.1.3.1. Missing events	Omission failure mode
1.1.3.2. Timing and ordering	Early or late failure mode
1.1.3.3. Abnormal terminations	Commission failure mode
1.1.3.4. Incorrect logic	Value failure mode
1.2. Hardware failures	Hardware component
1.2.1. Effectors failures	Hardware component
1.2.2. Power failures	Hardware component
1.2.3. Control failures	Hardware component
1.2.4. Sensors failures	Hardware component
2. Interaction failures	Socio components
2.1. Human errors	Human components
2.1.1. Mistakes	Selecting goal component
2.1.2. Slips	Acting component
2.1.3. Lapses	Information processing component
2.1.4. Deliberate violations	Conforming to rules component
2.2. Environmental and other agents failures	Environment unit component
2.2.1. Group-level judgment	Organizational climate component
2.2.2. Working environment	Environment unit component
2.2.3. Organizational flaws	Organization and regulation unit component
2.3. Social norm violations	-

technical components and then failure behavior can be modeled by defining possible failure modes in the inputs and by defining FPTC rules for each component. Similarly, *software and hardware failures* can be modeled using *software and hardware components* and *communication failures* can be modeled using *connectors*. For example, in modeling and analysis of our selected case in Subsection 4.5.3, we show how the software and hardware components are used for modeling technical failures. Equipment component can be used for modeling *design failures*. More details about equipment component are in [38], where we have previously proposed the extensions in relation to organizational factors. We also illustrated how we can use this component in Section 4.5.3. *Incorrect data, bad timing, extra data* and *missing data* can be modeled by using *value failure mode, early/late, commission* and *omission failure modes* as explained in Section 2.1.

Processing failures can be modeled by modeling a component failure behavior as *source* as explained in Subsection 2.5.2. It shows that a technical component is producing failure and there is problem in the processing. *Missing events, timing and ordering, abnormal terminations* and *incorrect logic* can be modeled by using different failure modes in the source behavior.

Effectors failures, power failures, control failures and *sensor failures* can be modeled using hardware component and defining their behavior and possible failure modes.

Based on the definition provided in [101], *interaction failures* are failures due to uncertainties in interaction between human, environment and other agents. These failures can be modeled by *socio components* and human errors can be modeled by using human components.

For *mistakes, slips, lapses* and *deliberate violations* there are specific components named *selecting goal, acting, information processing* and *conforming to rules components*, respectively. These components can be used for modeling the assigned failures as it is completely explained in [71].

Finally, *environment and other agents failures* and *working environment failures* can be modeled using *environment unit component*, organizational flaws can be modeled using *organization and regulation unit component* and *group-level judgement* (for example failure due to effects of group-level judgements on human actions) can be modeled using *organization climate component*. There are no associated modeling element for modeling *social norm violations* (for example failure in robot behavior due to not being in compliance with social norm).

Most of the failures in the considered taxonomy are technical failures and failures related to socio aspects are not intensely investigated, while these socio

failures, in addition to effects of AR and organizational changes are considered in our extensions to a great extent.

Results for the Third Research Question In order to calculate the percentage of extension in risk assessment with respect to effects of AR and organizational changes (third metric), we use the number of identified risk sources which are in connection with AR and organization changes divided by the total number of identified possible risk sources discussed in subsection 4.5.3, Table 4.3 and 4.4. There are 16 identified risk sources in connection with AR and organizational changes in total of 41 identified possible risk sources, which shows 39 percent extension in the risk assessment with respect of effects of AR and organizational changes. From the 16 identified risk sources in connection with AR and organizational changes, 7 of them are in connection with organizational changes with the potential to result in post normal accidents. Therefore, 17 percent extension in risk assessment is provided in order to prevent post-normal accidents.

Discussion on the validity

As it is described in [97], validity of a study discusses the trustworthiness of the results and to what extent the results may be biased by subjective viewpoint of the researcher. We use three aspects of validity, which are introduced in the study containing construct validity, internal and external validity.

Construct validity This aspect refers to the extent of representation of operational measures based on research questions. We defined operational measures based on the research questions using GQM method. We considered defining operational measures in a way to be able to use data which is possible for us to collect and use it to answer the research questions. For example, we defined typical human robot interaction failure coverage as operational measure in order to measure effectiveness of capturing the essential information for assessing risk in socio-technical robotic factory. This selection was affected by considering that it was possible for us to measure coverage using a typical failure taxonomy in human robot collaboration domain. Thus, some extent of subjectivity is not avoidable, meanwhile we tried to perform it with subjectivity as low as possible.

Internal validity This aspect refers to considering different causal relations affecting an investigated factor and not missing some of them. In our case,

we considered percentage of supported risk assessment steps based on standards, percentage of human robot interaction failure coverage and percentage of extensions with respect to effects of AR and organizational changes as three distinct metrics for measuring support for standards, the extent of effectiveness of the framework and development of risk assessment with respect to effects of AR and organizational changes, respectively. We defined our goal, research questions and metrics based on GQM method in order to consider causal relations affecting our goal, which can be helpful to increase internal validity. However, we are aware of some limitations in relation to internal validity. For example, in the system modeling and designing various scenarios, we considered different assumptions, which can lead to missing some causal relations affecting on system behavior. In modeling and analyzing system behavior, we have considered simplifications and in reality, much more effort is required to investigate various causal relations and to investigate fulfillment of the assumptions.

External validity This aspect refers to possibility of generalization of the findings. We have discussed about generalization of the FRAAR risk assessment in [54] and one of the main purposes of this study is demonstrating the applicability of the framework in a new domain, which is in line with demonstrating that the framework can be used as a general framework in different domains for risk assessment of AR-equipped socio-technical systems with respect to effects of AR and organizational changes.

This contribution is presented in Paper E (see Chapter 11).

4.6 Systematic Literature Review

In order to be able to position our work, we provide a Systematic Literature Review (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/globalization and by considering evolution of safety standards and safety perspectives. 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 section, we report the results of our SLR based on development of current techniques for risk assessment of safety-critical socio-technical systems. We undertake the SLR based on the guidelines proposed by Kitchenham

and Charters [106] and we aim at identifying primary studies on risk assessment of safety-critical socio-technical systems, analyzing them and providing our interpretation on evolution of socio-technical systems' conceptualization. Then, the results are used for positioning and comparing our work.

Research Questions: By considering the goal of the SLR we formulate the 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 effects of AR 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 provided explanation in Subsection 2.1))
  - 1.10. Which safety perspective is supported? (safety I, safety II, safety III or safety engineering today explained in Subsection 2.1)
- **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 Figure 4.28.

After a thoughtful evaluation from a list of 1752 papers found in recognized online libraries, 19 primary studies are selected. The study overview of the identified 19 primary studies are presented in Table 4.7. Our framework proposed in [54] is also identified as a primary study.

After analyzing the primary studies and extracting the findings related to the research questions, we provide tables summarizing the findings related to research questions shown in Tables 4.8 - 4.12.

As it is shown in Table 4.8, 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, which are extensively tackled in our work. 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

Figure 4.28: Defined abbreviations for possible options of extracted information in relation to each research question

Table 4.7: Selected primary studies

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

Table 4.8: Summary of the reviewed primary studies in relation to the first research question

ID	Socio entities characterization	Technical aspects characterization	Socio-technical orchestration characterization	Organizational changes effects characterization	Technological changes effects characterization	AR effects characterization
[107]	PfC	PfC	PfC	NC	NC	NC
[108]	PfC	PfC	PfC	NC	NC	NC
[109]	PfC	PfC	PfC	NC	NC	NC
[110]	PfC	PfC	PfC	NC	NC	NC
[111]	SC	SC	SC	PfC	NC	NC
[112]	PfC	PfC	PfC	NC	NC	NC
[113]	PfC	PfC	PfC	NC	NC	NC
[114]	PfC	PfC	PfC	NC	NC	NC
[115]	PfC	PfC	PfC	NC	NC	NC
[116]	PfC	PfC	PfC	NC	NC	NC
[117]	PfC	PfC	PfC	PfC	PfC	NC
[118]	PfC	NC	NC	NC	NC	NC
[119]	SC	SC	PfC	NC	NC	NC
[120]	PfC	PfC	PfC	NC	NC	NC
[121]	PfC	NC	NC	NC	NC	NC
[122]	NC	SC	NC	NC	NC	NC
[123]	SC	NC	NC	NC	NC	NC
[54]	SC	SC	SC	SC	SC	SC
[124]	SC	SC	SC	NC	NC	NC

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

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 4.9, 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 4.10, in most papers the modeling formality is in

Table 4.9: Summary of the reviewed primary studies in relation to the first research question (Con.)

ID	Risk/dependability characterization	Provided steps of risk assessment process	Safety perspective
[107]	PfC	RI, RA, RE	Set
[108]	SC	RI, RA, RE	Set
[109]	PfC	RI, RA, RE	Set
[110]	PfC	RI, RA, RE	Set
[111]	PfC	RI, RA, RE	Safety III
[112]	PfC	RI, RA, RE	Set
[113]	PfC	RI, RA, RE	Safety II
[114]	PfC	RI, RA, RE	Safety III
[115]	PfC	RI, RA, RE	Safety III
[116]	PfC	RI, RA, RE	Set
[117]	PfC	RI, RA, RE	Set
[118]	PfC	RI, RA, RE	Set
[119]	PfC	RI, RA, RE	Safety II
[120]	PfC	RI, RA, RE	Safety III
[121]	PfC	RI, RA, RE	Set
[122]	SC	RI, RA, RE	Set
[123]	PfC	RI, RA, RE	Safety II
[54]	SC	RI, RA, RE	Set
[124]	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.

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

Table 4.10: Summary of the reviewed primary studies in relation to the second research question

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
[107]	IN	NEFC	Ql + Qn	Ln	BL+FL	No	No
[108]	IN	Concepts	Ql + Qn	Ln	BL+FL	Yes	No
[109]	IN	Concepts	Ql	Ln	FL	No	No
[110]	IN	NEFC	Ql + Qn	Ln	FL	Yes	Yes
[111]	IN	Concepts	Ql	NL	FL	No	No
[112]	IN	Concepts	Ql + Qn	Ln	BL	No	No
[113]	IN	Concepts	Ql + Qn	NL	BL	No	Yes
[114]	IN	Concepts	Ql	NL	FL	No	No
[115]	IN	Concepts	Ql	NL	FL	No	No
[116]	IN	Concepts	Ql + Qn	Ln	BL	No	No
[117]	IN	Concepts	Ql	Ln	FL	No	No
[118]	IN	NEFC	Ql + Qn	Ln	BL+FL	Yes	No
[119]	FM	Concepts + Semantics	Ql	Ln	FL	No	No
[120]	IN	Concepts	Ql	NL	BL+FL	No	No
[121]	IN	Concepts	Ql	Ln	BL	No	No
[122]	FM	Concepts + Semantics	Ql	Ln	FL	Yes	Yes
[123]	IN	Concepts	Ql	Ln	FL	No	No
[54]	FM	Concepts	Ql	Ln	FL	Yes	No
[124]	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

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 improvements in this context.

As it is shown in Table 4.11, 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 stan-

Table 4.11: Summary of the reviewed primary studies in relation to the third research question

ID	Application	Potential for other applications	Support for standards	Presence of scenarios
[107]	General	Yes	NM	Yes
[108]	Specific	Yes	M (IEC 61508 and IEC 61511)	Yes
[109]	General	Yes	NM	Yes
[110]	General	Yes	NM	Yes
[111]	Specific	Yes	NM	Yes
[112]	Specific	No	NM	Yes
[113]	Specific	Yes	NM	Yes
[114]	General	Yes	NM	Yes
[115]	General	Yes	M (SAE ARP-4761)	Yes
[116]	General	Yes	NM	Yes
[117]	Specific	Yes	NM	No
[118]	General	Yes	NM	Yes
[119]	General	Yes	NM	Yes
[120]	General	Yes	NM	Yes
[121]	General	Yes	NM	Yes
[122]	Specific	Yes	NM	Yes
[123]	Specific	Yes	NM	Yes
[54]	General	Yes	M (ISO 26262 and ISO/-PAS 21448-SOTIF)	Yes
[124]	Specific	Yes	M (IEC 61508, etc.)	Yes

NM: Not Mentioned. M: Mentioned

dards. Thus, it is important to explain how they can support the standards to ease their selection when practitioners need to choose a method/techniques/-models/frameworks 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 4.12, 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 for-

Table 4.12: Summary of the reviewed primary studies in relation to the fourth research question

ID	Stated challenges
[107]	1) Lack of readily available data to be used by human factor approaches that can be used in the framework
[108]	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
[109]	1) Requiring tools for evaluating quantitative analysis, 2) requiring exploration to mesh with existing safety/dependability assurance processes
[110]	1) Requiring further research for understanding the influence of human and organizational factors on safety operation
[111]	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
[112]	1) Use of limited reports from specific databases, 2) subjectivity in the reports
[113]	1) No criteria for assessment of the significance of introducing this approach
[114]	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)
[115]	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
[116]	1) Requiring further research for providing details of the proposed broad concepts
[117]	1) Lack of application, 2) lack of analysis procedure
[118]	1) Requiring further testing, 2) requiring detailed validation
[119]	1) Lack of quantitative analysis, 2) lack of tool support
[120]	Not mentioned
[121]	1) Lack of quantitative analysis, 2) lack of identification of the dynamic characteristics of systems, 3) lack of non-linear relationships characterization
[122]	1) Requiring further research for providing more applications, 2) providing automatic risk assessment, and 3) providing safeguard interpretation
[123]	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
[54]	1) Requiring further research for providing more applications, 2) providing automatic risk assessment by implementing the extensions, and 3) providing scenarios from other domains
[124]	1) Lack of formal verification for checking safety rules consistency and the safety justification

mality 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 AR-equipped 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, it is essential to consider effects of new technological and organizational changes on system behavior.

This contribution is presented in Paper D (see Chapter 10).

Chapter 5

Related Work

In this section, related work is discussed. In Section 5.1, works that address modeling of socio-technical systems are presented. In Section 5.2, works that address risk analysis in socio-technical systems are presented. In Section 5.3, works that address literature reviews on risk and safety analysis are presented. Finally, in Section 5.4, works that address case studies in safety and risk analysis are presented.

5.1 Modeling Socio-technical Systems

In this section, we discuss about works with contributions mainly in socio-technical system modeling. However, there may be other proposed contributions as well.

In [111], the authors propose a technique for modeling global software development project as a complex socio-technical system. In this method, functional components are identified and links between the components are defined. Feedback controller is used between two components to control if there is any deviation between the interpretation of the component providing the output and the component receiving the output as its input. Feedback controller implementation can not be done through mechanical device and informal communication is required. The provided modeling technique is specific for software development as a socio-technical system and can not be used for other domains. In comparison, the proposed modeling constructs in our work can be used for socio-technical systems including hardware, software and socio entities used

in various domains.

In [108], 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. This work is provided specifically for operational phase of a safety instrumented system with a focus on SIL prediction, while our work is a general framework that can be used in various domains. In addition, we also consider effects of AR and organizational changes on modeling system behavior.

In [109], 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. In comparison to our work, this approach provides limited concepts and effects of AR and organizational changes are not integrated in the modeling process.

In [123], 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. This framework is developed for railway context, but there are guidance for generic application of HFiS. In comparison to our work, in this paper there is no consideration on effects of new technologies such as augmented reality and organizational changes on hu-

man functions and organizational factors.

In [113], 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. This framework is proposed for railway domain and has a backward looking due to focus on safety occurrence modeling, while our work is a general forward looking framework which can be used in various domains.

5.2 Risk Analysis in Socio-technical Systems

In this section, we discuss about works with contributions mainly in risk analysis of socio-technical systems. However, there may be other proposed contributions as well and other terms such as hazard analysis and safety analysis may be used.

Modeling can be considered as part of risk analysis and we model systems to empower analysis techniques to do risk analysis in the systems. In risk analysis techniques for socio-technical systems, failures emanated from human and organizational factors are also considered in addition to technical failures. Human failure taxonomies provide the possible human failures while working in a socio-technical system. There are also taxonomies on organizational factors that provide the factors influencing human performance.

In some of the works risk analysis is done based on questionnaires and ratings provided by people using the system. For example, in [125], risk analysis for context-adaptive augmented reality aerodrome control towers assistance

system is done through ratings provided by aerodrome controllers using the system in a simulation environment. Criteria used for risk analysis are transparency, complexity, interference, disruptiveness, distraction potential, failure modes and trust/complacency. The results of the analysis show that this system is supportive for air traffic controllers and provides safety benefits. This study would be useful for demonstrating the effectiveness of using augmented reality in aerodrome control towers assistance systems. Instead, our approach includes modeling of the AR-equipped socio-technical systems to analyze system behavior and to find the possible risk sources and eliminating possible failures during the system development process.

A risk analysis technique for systems containing augmented reality, named Safe-AR, is proposed in [126]. Safe-AR integrates failures of AR/user interface at three levels: perception, comprehension and decision-making. Likely risks and their severity are based on reports available in literature. The proposed technique is shown on an AR left-turn assist app, which is an example from automotive domain. Human functions and failure modes in this study are limited to the provided example and a generalization is required to be used for other domains and more complicated case studies.

In [110], 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. In this study, there is no proposed formality extension and MMOS concepts are used along with human reliability analysis methods. In comparison, we proposed new concepts in order to integrate effects of AR and organizational changes on socio aspects in the risk assessment process.

In [124], 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)

[127]. 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. This work is provided specifically for autonomous train driving, while our work is a general framework that can be used in various domains. In addition, we also consider effects of AR and organizational changes on modeling system behavior.

5.3 Literature Reviews on Safety/Risk Analysis

In this section, we discuss about works with contributions mainly on reviewing previous studies in the context of risk/safety analysis. However, there may be other proposed contributions and different terms may be used such as risk assessment, hazard analysis.

A review of advances on the foundation of risk assessment and risk management is performed in [128]. 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 [129]. 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 [130] 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 integrating challenges.

A systematic literature review is provided in [131] 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 [132] 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. However, there is no systematic literature review considering conceptualization of evolution of socio-technical systems in the risk assessment process. Due to the broad 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 and this is what we tackled in the systematic literature review prepared in this thesis.

5.4 Case Studies in Safety/Risk Analysis

In this section, we discuss about works with contributions mainly as case study in the context of safety/risk analysis in socio-technical systems. However, there may be other proposed contributions as well and various terms may be used instead of safety/risk analysis, such as risk assessment, hazard analysis.

In [133], the authors provide a case study for safety analysis in aircraft ground handling services using STAMP (Systems Theoretic Accident Model and Process) causation model [134]. Based on the case study, the limitations of using this model as an organizational management theory are discussed. For example, it is discussed that behavior of people is not represented and by placing a control on behavior without knowing its driving forces, the possible contribution of workers to safety and the complexities that they face are neglected. In addition, it is recommended in this study to use complementary approaches to STAMP in order to consider social dynamics and understanding emergent behavior of systems before introducing control.

In [135], the authors provide a case study for modeling and situational awareness analysis of human-computer interaction in the aircraft cockpit. It considers the model with three modules: pilot agent, technical system and environment modules. Two scenarios with human-computer interaction are used and the results are compared with past studies to illustrate the advantages.

In [136], the authors provide a case study for modeling heating, ventilation and air-conditioning (HVAC) systems using FRAM (Functional Resonance

Analysis Method) [137]. In order to decrease the complexity of the FRAM model representation, a layered FRAM is presented in this study. Scenarios containing dynamic nature of complex socio-technical systems are considered and the results show better view of the functions and facilitation in analyzing the model.

In [138], the authors discuss the challenges of providing safety in an intelligent human robot collaborative station using the current safety standards and the need for updating and improving them. As it is explained in this paper, according to robotic safety standards, it is mandatory to have risk assessment process for all robotic applications. However, the standards do not support the collaboration in an efficient manner. Manual assembly station from a truck engine final assembly line is used as a use-case and five hazards are identified and described. For each hazards some recommendations are provided to reduce the risk. Finally, a new collaboration mode called "Deliberation in planning and acting" is suggested to include advanced control strategies and improve the current standards. For implementing the suggested mode, control system component should be added to support the deliberation and to provide an agreed plan for safe collaboration. Good understanding of the system and well received education and training is also required by the operator.

In [139], the authors propose a systematic risk assessment approach and apply it to an automated warehouse use case. Based on the proposed approach, different humans with different levels of interaction are identified and their safety requirements are provided. In addition, a list of hazards and their related scenarios are identified using HAZOP method. Finally, the hazards are analyzed, and safety requirements and recommendations are generated to be used in the next risk mitigation phase. Furthermore, a simulation setup is implemented for risk management process using a Virtual Robot Experimentation Platform (V-REP).

In comparison to the above-mentioned works, we provided two case studies from two different domains using our proposed general risk assessment framework with the integration of risks emanated from human, organization and technology containing augmented reality. In addition, effects of new organizational changes and support for safety standards are considered.

Chapter 6

Conclusion and Future Work

In this section, we first summarize our work and provide concluding remarks and then we present the future research directions.

6.1 Conclusions

The goal of our research is to strengthen risk assessment in augmented reality-equipped socio-technical systems considering post normal accidents by providing a safety-centered risk assessment framework. For achieving this goal, we focused on various kinds of dependability threats that would cause risk for post normal accidents and we proposed constructs for modeling and analyzing system behavior in order to be able to assess the related risks. We defined four subgoals (presented in detail in Section 3.3):

- **Subgoal 1**: Capturing dependability threats leading to post normal accidents in AR-equipped socio-technical systems.
- **Subgoal 2**: Integrating captured threats in risk assessment of AR-equipped socio-technical systems.
- **Subgoal 3**: Validating modeling and analysis capabilities, applicability and effectiveness of contributions for risk assessment in AR-equipped socio-technical systems.
- **Subgoal 4**: Positioning and comparing the contributions of our work.

To reach the specified subgoals, we presented set of research contributions (detailed in Chapter 4):

- **Thesis contribution 1**: A metamodel extension for capturing post normal accidents

  We reviewed post normal accident theory and we extracted influencing factors on system behavior, which would act as dependability threats leading to post normal accidents. We used the extracted influencing factors, which are related to organizational changes, in extending a metamodel for modeling socio-technical systems to empower the metamodel by the required expressive power for capturing dependability threats related to organizational changes leading to post normal accidents.

- **Thesis contribution 2**: Proposing a process for dependability analysis in AR-equipped socio-technical systems based on the proposed modeling extensions

  We proposed extension for a synergy of qualitative and quantitative dependability analysis in order to incorporate our proposed modeling extensions and enable the analysis process to be used for assessing risk in AR-equipped socio-technical systems.

- **Thesis contribution 3**: Proposing a safety-centered risk assessment framework for AR-equipped socio-technical systems integrating modeling and analysis extensions

  We proposed a safety-centered risk assessment framework for AR-equipped socio-technical systems containing modeling and analysis processes in order to incorporate our proposed extensions and strengthen the risk assessment for AR-equipped socio-technical systems with respect to effects of AR and organizational changes and provided support for safety standards.

- **Thesis contribution 4**: Applying the contributions in automotive domain

  We designed and executed a case study in automotive domain in cooperation with our industrial partner in order to validate modeling and analysis capabilities of the framework in compliance with related safety standards. We indicated how different steps support different development process activities of safety standards.

- **Thesis contribution 5**: Applying the contributions in robotic domain

  We applied our framework in robotic domain to evaluate the contributions of our work with respect to effects of AR and organizational changes and to

demonstrate the applicability and effectiveness of our framework in robotic domain in compliance with related safety standards.

- **Thesis contribution 6**: A systematic literature review on risk assessment of safety-critical socio-technical systems

  We conducted a systematic literature review based on the evolution of the conceptualization of socio-technical systems to position our work and to compare our contributions with other related works.

Figure 6.1 presents the mapping between subgoals, research contributions and included papers.

6.2 Future Work

The contributions provided in this thesis can be improved and extended in several directions. Here we present the suggested areas for future work.

- We used SafeConcert metamodel as the basis of our extensions. The reason that we decided to work on SafeConcert metamodel and to extend this metamodel is that this metamodel provides modeling elements for socio-technical systems. In addition, this metamodel is tool-supported and it is based on previous accepted works with open source implementation. One research direction is to integrate our proposed extensions in other metamodels.

- The proposed extension for extending CHESS toolset is not implemented in this toolset. One suggested area for future work is to implement the conceptual extension of SafeConcert within CHESSML, in order to enable using of the extensions by the analysis plugin of CHESS toolset.

- In this thesis, we used extended modeling elements for modeling effects of AR and organizational changes on human and organization as socio entities. One research direction is to consider effects of AR and organizational changes on technical entities and proposing required constructs for modeling technical issues due to AR and organizational changes such as problems in design, material or production process due to AR and organizational changes while producing a technical fragment.

- We used Concerto-FLA analysis technique for describing the analysis results after using our modeling extensions. The reason that we decided to

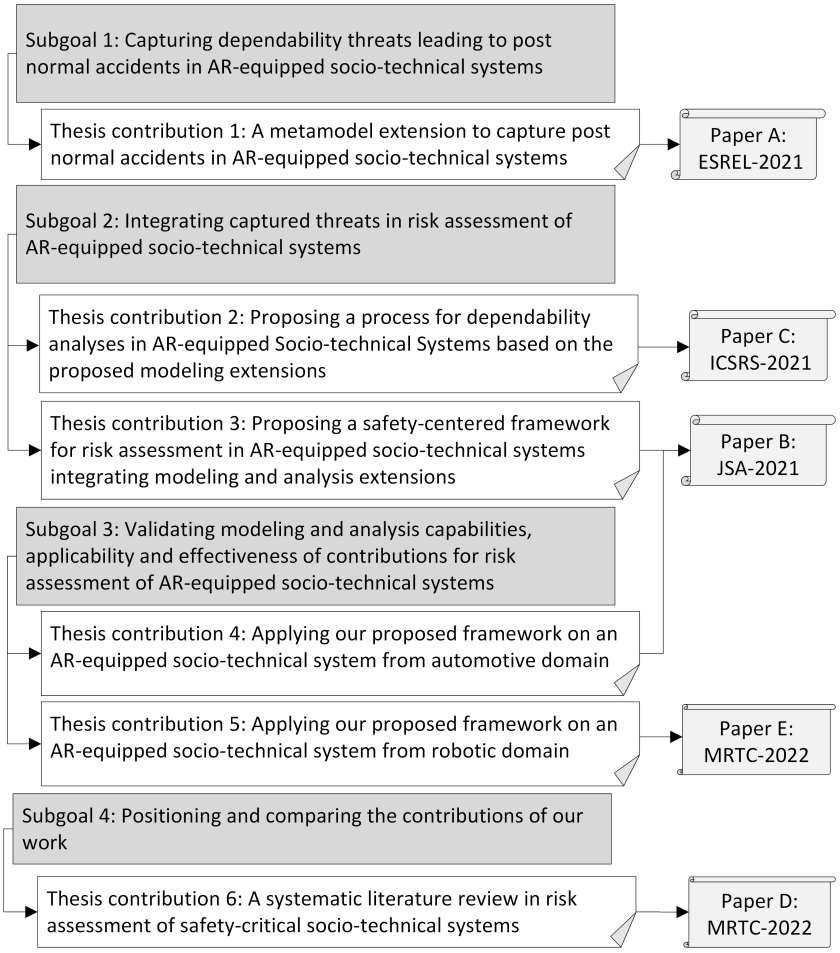


Figure 6.1: Connection between subgoals, contributions and the papers

use Concerto-FLA analysis technique is that this technique provides analysis constructs for socio-technical systems. In addition, this technique is tool-supported and it is based on previous accepted works with open source implementation. One research direction is to extend other analysis techniques based on our proposed extensions.

- Implementing the conceptual extensions of CHESSML provides the possibility for implementing the extensions on analysis, in order to have the analysis results automatically. One suggested area for future work is to extend implementation of Concerto-FLA based on extended modeling elements to provide the analysis results within the CHESS toolset.

- Probabilities used in the quantitative analysis are assumed and as we discussed, accident reports or expert opinions can be used for defining these probabilities. In the context of human and organization components, defining the probabilities are really challenging and one research direction would be to consider this challenge and to provide a solution. However, it might be more helpful and more descriptive to use qualitative analysis in socio-technical systems.

- Concerto-FLA, which is used as the basis for the analysis in our framework, is a static and linear analysis technique. In addition, ensuring consistency over time is not considered in this technique. In order to consider dynamic, non-linear relationships and in order to consider consistency over time, more extensions are required.

- We provided a theoretical comparison in our systematic literature review. However, more descriptive comparisons with widespread methods are of value to show the validity of the contributions in different levels. One future research direction is to provide a comparative study based on the best practices and guidelines.

- The results of the modeling and analysis would be affected by inconsistencies in defining the rules or by complexity of the system and components. Checking consistency and dealing with complexity is out of scope of this thesis and it can be considered as a direction for future work.

- In this thesis, we focused on identifying the dependability threats and assessing risk caused by these dependability threats. We did not provide any mitigation technique for the identified risks and this can be considered as future work to decrease risk and define risk reduction measures in order to provide risk management techniques.

- Defining various scenarios requires different meetings between involved people in the system and people familiar with the risk assessment process. It is required to have comprehensive discussions about various possible scenarios to integrate possible risk sources as much as possible and to define the related

safety requirements. one research direction for future work can be defining methods for formalizing the process of defining scenarios.

Bibliography

[1] ImmerSAFE., "Immersive Visual Technologies for Safety-critical Applications." URL: https://immersafe-itn.eu/.

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

[3] C. Perrow, *Normal accidents: Living with high risk technologies-Updated edition*. Princeton university press, 2011. DOI: https://doi.org/10.2307/j.ctt7srgf.

[4] J. Noll and S. Beecham, "Measuring global distance: A survey of distance factors and interventions," in *International Conference on Software Process Improvement and Capability Determination*, pp. 227–240, Springer, 2016. DOI: https://doi.org/10.1007/978-3-319-38980-6_17.

[5] S. Alajrami, B. Gallina, and A. Romanovsky, "Enabling GSD task allocation via cloud-based software processes," in *International Conference on Software Engineering Research, Management and Applications*, pp. 179–192, Springer, 2017. DOI: https://doi.org/10.2991/ijndc.2017.5.4.4.

[6] L. Q. Yeong, *Investigating the influence of cultural differences on systems engineering: a case study of the manned spaceflight programs of the United States and China*. PhD thesis, Massachusetts Institute of Technology, 2015.

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

[8] ISO 26262-1, "Road vehicles — Functional safety — Part 1: Vocabulary," 2018. URL: https://www.iso.org/standard/68383.html.

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

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

[11] CONCERTO D2.7, "Analysis and back-propagation of properties for multicore systems – Final Version." URL: http://www.concerto-project.org/results.

[12] A. Debiasi, F. Ihirwe, P. Pierini, S. Mazzini, and S. Tonetta, "Model-based analysis support for dependable complex systems in chess," in *the 9th International Conference on Model-Driven Engineering and Software Development - Volume 1: MODELSWARD,*, pp. 262–269, INSTICC, SciTePress, 2021. DOI: https://doi.org/10.5220/0010269702620269.

[13] S. Sheikh Bahaei and B. Gallina, "Extending SafeConcert for Modelling Augmented Reality-equipped Socio-technical Systems," in *International Conference on System Reliability and Safety (ICSRS)*, IEEE, 2019. DOI: https://doi.org/10.1109/ICSRS48664.2019.8987702.

[14] S. Sheikh Bahaei, *A Framework for Risk Assessment in Augmented Reality-equipped Socio-technical Systems*. No. 293, Mälardalen University Press Licentiate Theses, 2020. URL: http://mdh.diva-portal.org/smash/record.jsf?pid=diva2%3A1432516.

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

[16] ARTEMIS-JU-100022 CHESS, "Composition with guarantees for high-integrity embedded software components assembly." URL: http://www.chess-project.org/.

[17] T. Aven, *Foundations of risk analysis*. John Wiley & Sons, 2012. DOI: https://doi.org/10.1002/0470871245.

[18] W. W. Lowrance and J. Klerer, "Of acceptable risk: Science and the determination of safety," *Journal of The Electrochemical Society*, vol. 123, no. 11, p. 373C, 1976. DOI: https://doi.org/10.1149/1.2132690.

[19] ICH Database, "ICH harmonised tripartite guideline: Quality risk management (Q9)," 2005. URL: https://www.ich.org/page/quality-guidelines.

[20] SRA, "Glossary society for risk analysis." URL: www.sra.com/resources.

[21] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," *transactions on dependable and secure computing*, vol. 1, no. 1, pp. 11–33, 2004. DOI: https://doi.org/10.1109/TDSC.2004.2.

[22] D. J. Pumfrey, *The principled design of computer system safety analyses.* PhD thesis, University of York, 1999.

[23] B. Gallina, *PRISMA: a software product line-oriented process for the requirements engineering of flexible transaction models*. PhD thesis, University of Luxembourg, 2010. URL: http://hdl.handle.net/10993/15427.

[24] J. McDermid, "Software hazard and safety analysis," in *International Symposium on Formal Techniques in Real-Time and Fault-Tolerant Systems*, pp. 23–34, Springer, 2002. DOI: https://doi.org/10.1007/3-540-45739-9_2.

[25] ISO 12100, "safety of machinery - General principles for design - Risk assessment and risk reduction," 2010. URL: https://www.iso.org/standard/51528.html.

[26] E. Hollnagel, R. L. Wears, and 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: https://doi.org/10.13140/RG.2.1.4051.5282.

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

[28] T. Aven, "A risk science perspective on the discussion concerning Safety I, Safety II and Safety III," *Reliability Engineering & System Safety*, vol. 217, p. 108077, 2022. DOI: https://doi.org/https://doi.org/10.1016/j.ress.2021.108077.

[29] J.-C. Le Coze, "Globalization and high-risk systems," *Policy and practice in health and safety*, vol. 15, no. 1, pp. 57–81, 2017. DOI: https://doi.org/10.1080/14773996.2017.1316090.

[30] J. D. Herbsleb and D. Moitra, "Global software development," *IEEE software*, vol. 18, no. 2, pp. 16–20, 2001. DOI: http://dx.doi.org/10.1109/52.914732.

[31] D. P. T. Piamonte, J. D. Abeysekera, and K. Ohlsson, "Understanding small graphical symbols: a cross-cultural study," *International Journal of Industrial Ergonomics*, vol. 27, no. 6, pp. 399–404, 2001. DOI: https://doi.org/10.1016/S0169-8141(01)00007-5.

[32] J. Goldenberg and M. Levy, "Distance is not dead: Social interaction and geographical distance in the internet era," *arXiv:0906.3202*, 2009. DOI: https://doi.org/10.48550/ARXIV.0906.3202.

[33] J. Tang, M. Musolesi, C. Mascolo, and V. Latora, "Temporal distance metrics for social network analysis," in *the 2nd ACM workshop on online social networks*, pp. 31–36, 2009. DOI: https://doi.org/10.1145/1592665.1592674.

[34] B. F. Goldiez, N. Saptoka, and P. Aedunuthula, "Human performance assessments when using augmented reality for navigation," tech. rep., University of Central Florida Orlando Inst for Simulation and Training, 2006.

[35] L. Roitman, J. Shrager, and T. Winograd, "A comparative analysis of augmented reality technologies and their marketability in the consumer electronics segment," *Journal of Biosensors and Bioelectronics*, vol. 8, no. 01, 2017. DOI: https://doi.org/10.4172/2155-6210.1000236.

[36] D. Van Krevelen and R. Poelman, "A survey of augmented reality technologies , applications and limitations," *The International Journal of Virtual Reality*, vol. 9, no. 2, pp. 1–20, 2010. DOI: http://dx.doi.org/10.20870/IJVR.2010.9.2.2767.

[37] M. T. Phan, *Estimation of driver awareness of pedestrian for an augmented reality advanced driving assistance system*. PhD thesis, Université de Technologie de Compiègne, 2016.

[38] S. Sheikh Bahaei, B. Gallina, K. Laumann, and M. Rasmussen Skogstad, "Effect of augmented reality on faults leading to human failures in socio-technical systems," in *International Conference on System Reliability and Safety (ICSRS)*, IEEE, 2019. DOI: https://doi.org/10.1109/ICSRS48664.2019.8987586.

[39] N. Hall, C. Lowe, and R. Hirsch, "Human factors considerations for the application of augmented reality in an operational railway environment," *Procedia Manufacturing*, vol. 3, pp. 799–806, 2015. DOI: https://doi.org/10.1016/j.promfg.2015.07.333.

[40] F. Schwarz and W. Fastenmeier, "Augmented reality warnings in vehicles: Effects of modality and specificity on effectiveness," *Accident Analysis & Prevention*, vol. 101, pp. 55–66, 2017. DOI: https://doi.org/10.1016/j.aap.2017.01.019.

[41] S. Ventura, R. M. Baños, C. Botella, and N. Mohamudally, "Virtual and augmented reality: New frontiers for clinical psychology," *State of the art virtual reality and augmented reality knowhow*, vol. 10, pp. 99–118, 2018. DOI: https://doi.org/10.5772/intechopen.74344.

[42] N. Salamon, J. M. Grimm, J. M. Horack, and E. K. Newton, "Application of virtual reality for crew mental health in extended-duration space missions," *Acta Astronautica*, vol. 146, pp. 117–122, 2018. DOI: http://dx.doi.org/10.1016/j.actaastro.2018.02.034.

[43] A. Heather, "How augmented reality affects the brain," tech. rep., Neuro-Insight, 2018. URL: https://www.zappar.com/blog/how-augmented-reality-affects-brain.

[44] M. Gutiérrez *et al.*, "Augmented reality environments in learning, communicational and professional contexts in higher education," *Digital Education Review*, vol. 26, pp. 22–35, 2014. URL: https://raco.cat/index.php/DER/article/view/288340.

[45] K. Lee, "Augmented reality in education and training," *TechTrends*, vol. 56, no. 2, pp. 13–21, 2012. DOI: https://doi.org/10.1007/s11528-012-0559-3.

[46] M. R. Miller, H. Jun, F. Herrera, J. Y. Villa, G. Welch, and J. N. Bailenson, "Social interaction in augmented reality," *PloS one*, vol. 14, no. 5, 2019. DOI: https://doi.org/10.1371/journal.pone.0216290.

[47] IEC 61508, "Functional Safety of Electrical/Electronic/Programmable Electronic Safety-related Systems," 2010. URL: https://webstore.iec.ch/publication/5515.

[48] ISO 10218-1, "Robots and robotic devices – Safety requirements for industrial robots – Part 1: Robots," 2011. URL: https://www.iso.org/standard/51330.html.

[49] T. Hecht, M. Lienkamp, C. Wang, *et al.*, "Development of a human driver model during highly automated driving for the ASIL controllability classification," in *Tagung Fahrerassistenz*, 2017. URL: https://mediatum.ub.tum.de/doc/1421389/1421389.pdf.

[50] I. Sljivo, B. Gallina, J. Carlson, H. Hansson, *et al.*, "Using safety contracts to guide the integration of reusable safety elements within iso 26262," in *21st Pacific Rim International Symposium on Dependable Computing (PRDC)*, pp. 129–138, IEEE, 2015. DOI: https://doi.org/10.1109/PRDC.2015.12.

[51] ISO/PAS 21448, " Road vehicles — Safety of the intended functionality (SOTIF)," 2019. URL: https://www.iso.org/standard/70939.html.

[52] SAE, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," 2021. URL: https://www.sae.org/standards/content/j3016_202104.

[53] G. Dimitrakopoulos, L. Uden, and I. Varlamis, *The Future of Intelligent Transport Systems*. Elsevier, 2020. DOI: https://doi.org/10.1016/C2018-0-02715-2.

[54] S. Sheikh Bahaei, B. Gallina, and M. Vidović, "A case study for risk assessment in AR-equipped socio-technical systems," *Journal of Systems Architecture*, vol. 119, p. 102250, 2021. DOI: https://doi.org/10.1016/j.sysarc.2021.102250.

[55] ISO 10218-2, "Robots and robotic devices – Safety requirements for industrial robots – Part 2: Robot systems and integration," 2011. URL: https://www.iso.org/standard/41571.html.

[56] ISO/TS 15066, "Robots and robotic devices - Collaborative robots," 2016. URL: https://www.sis.se/en/produkter/manufacturing-engineering/industrial-automation-systems/industrial-robots-manipulators/isots150662016/.

[57] ISO 13849-1, "Safety of machinery – Safety-related parts of control systems – Part 1: General principles for design," 2015. URL: https://www.iso.org/standard/69883.html.

[58] L.-O. Berntsson, H. Blom, D. Chen, P. Cuenot, U. Freund, P. Frey, S. Gérard, R. Johansson, H. Lönn, M.-O. Reiser, *et al.*, "EAST-ADL2 UML2 Profile specification," 2008. URL: http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Akth%3Adiva-82888.

[59] S. Friedenthal, A. Moore, and R. Steiner, "OMG systems modeling language (OMG SysML) tutorial," in *INCOSE Intl. Symp*, vol. 9, pp. 65–67, 2006. URL: http://retis.sssup.it/ marco/files/lesson21_SysML.pdf.

[60] S. Bernardi, J. Merseguer, and D. C. Petriu, "A dependability profile within MARTE," *Software & Systems Modeling*, vol. 10, no. 3, pp. 313–336, 2011. DOI: https://doi.org/10.1007/s10270-009-0128-1.

[61] C. André, A. Cuccuru, J.-L. Dekeyser, R. De Simone, C. Dumoulin, J. Forget, T. Gautier, S. Gérard, F. Mallet, A. Radermacher, *et al.*, "MARTE: a new OMG profile RFP for the modeling and analysis of real-time embedded systems," in *DAC 2005 Workshop-UML for SoC Design*, 2005. URL: https://hal.science/hal-02466757.

[62] AMASS Open Platform, 2018. URL: https://www.polarsys.org/opencert/news/2018-12-05-download_p2_preview/.

[63] J. L. de la Vara, A. Ruiz, B. Gallina, G. Blondelle, E. Alaña, J. Herrero, F. Warg, M. Skoglund, and R. Bramberger, "The AMASS approach for assurance and certification of critical systems," in *Embedded World*, 2019. URL: http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Ari%3Adiva-38329.

[64] S. Mazzini, J. M. Favaro, S. Puri, and L. Baracchi, "CHESS: an Open Source Methodology and Toolset for the Development of Critical Systems," in *EduSymp/OSS4MDE@ MoDELS*, pp. 59–66, 2016. URL: https://ceur-ws.org/Vol-1835/paper09.pdf.

[65] K. C. Hendy, "A tool for human factors accident investigation, classification and risk management," tech. rep., Defence Research And Development Toronto (Canada), 2003. URL: https://apps.dtic.mil/sti/pdfs/ADA623696.pdf.

[66] J. Rasmussen, "Human errors. a taxonomy for describing human malfunction in industrial installations," *Journal of occupational accidents*, vol. 4, no. 2-4, pp. 311–333, 1982. DOI: https://doi.org/10.1016/0376-6349(82)90041-4.

[67] S. A. Shappell and D. A. Wiegmann, "The human factors analysis and classification system–HFACS," tech. rep., Civil Aeromedical Institute, 2000. URL: https://commons.erau.edu/publication/737/.

[68] D. Gertman, H. Blackman, J. Marble, J. Byers, C. Smith, *et al.*, "The SPAR-H human reliability analysis method," *US Nuclear Regulatory Commission*, vol. 230, 2005. URL: https://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr6883.

[69] S. Sheikh Bahaei and B. Gallina, "A Metamodel Extension to Capture Post Normal Accidents in AR-equipped Socio-technical Systems," in *European Safety and Reliability Conference (ESREL)*, Research Publishing, Singapore, 2021. DOI: https://doi.org/10.5281/zenodo.5599456.

[70] S. Shekh Bahaei and B. Gallina, "Towards assessing risk of safety-critical socio-technical systems while augmenting reality," 2019. Published as proceedings annex on the International Symposium on Model-Based Safety and Assessment (IMBSA) website, URL: http://easyconferences.eu/imbsa2019/proceedings-annex/.

[71] S. Sheikh Bahae and B. Gallina, "Augmented reality-extended humans: towards a taxonomy of failures – focus on visual technologies," in *European Safety and Reliability Conference (ESREL)*, Research Publishing, Singapore, 2019. DOI: https://doi.org/10.5281/zenodo.3601748.

[72] D. A. Norman, "Errors in human performance," tech. rep., California Univ San Diego LA JOLLA Center For Human Information Processing, 1980. URL: https://www.researchgate.net/publication/235152461_Errors_in_Human_Performance.

[73] J. Reason, *The human contribution: unsafe acts, accidents and heroic recoveries*. CRC Press, 2017. DOI: https://doi.org/10.1201/9781315239125.

[74] N. A. Stanton and P. M. Salmon, "Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems," *Safety Science*, vol. 47, no. 2, pp. 227–237, 2009. DOI: https://doi.org/10.1016/j.ssci.2008.03.006.

[75] W.-T. Fu, J. Gasper, and S.-W. Kim, "Effects of an in-car augmented reality system on improving safety of younger and older drivers," in *International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 59–66, IEEE, 2013. DOI: https://doi.org/10.1109/ISMAR.2013.6671764.

[76] M. C. Schall Jr, M. L. Rusch, J. D. Lee, J. D. Dawson, G. Thomas, N. Aksan, and M. Rizzo, "Augmented reality cues and elderly driver hazard perception," *Human factors*, vol. 55, no. 3, pp. 643–658, 2013. DOI: https://doi.org/10.1177/0018720812462029.

[77] S. Chandra and K. N. Kumar, "Exploring factors influencing organizational adoption of augmented reality in e-commerce: Empirical analysis using technology-organization-environment model," *Journal of Electronic Commerce Research*, vol. 19, no. 3, 2018. http://www.jecr.org/sites/default/files/2018vol19no3_paper3.pdf.

[78] S. Condino, M. Carbone, R. Piazza, M. Ferrari, and V. Ferrari, "Perceptual limits of optical see-through visors for augmented reality guidance of manual tasks," *IEEE transactions on bio-medical engineering*, 2019. DOI: https://doi.org/10.1109/TBME.2019.2914517.

[79] W. E. Vesely, F. F. Goldberg, N. H. Roberts, and D. F. Haasl, *Fault tree handbook*. Nuclear Regulatory Commission Washington DC, 1981. URL: https://www.nrc.gov/docs/ML1007/ML100780465.pdf.

[80] D. H. Stamatis, *Failure mode and effect analysis: FMEA from theory to execution*. Quality Press, 2003.

[81] Z. H. Qureshi, "A review of accident modelling approaches for complex socio-technical systems," in *the 12th Australian workshop on Safety critical systems and software and safety-related programmable systems*, Australian Computer Society, Inc., 2007. URL: https://apps.dtic.mil/sti/pdfs/ADA482543.pdf.

[82] M. Wallace, "Modular architectural representation and analysis of fault propagation and transformation," *Electronic Notes in Theoretical Computer Science*, vol. 141, no. 3, pp. 53–71, 2005. DOI: https://doi.org/10.1016/j.entcs.2005.02.051.

[83] X. Ge, R. F. Paige, and J. A. McDermid, "Probabilistic failure propagation and transformation analysis," in *International Conference on Computer Safety, Reliability, and Security (SafeComp)*, pp. 215–228, Springer, 2009. DOI: https://doi.org/10.1007/978-3-642-04468-7_18.

[84] Y. Papadopoulos, *Safety-directed system monitoring using safety cases*. PhD thesis, Citeseer, 2000. URL: https://www.researchgate.net/publication/2325163_Safety-Directed_-System_Monitoring_Using_Safety_Cases.

[85] B. Gallina, M. A. Javed, F. U. Muram, and S. Punnekkat, "A model-driven dependability analysis method for component-based architectures," in *38th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pp. 233–240, IEEE, 2012. DOI: https://doi.org/10.1109/SEAA.2012.35.

[86] ARTEMIS-JU-333053 CONCERTO, "Guaranteed Component Assembly with Round Trip Analysis for Energy Efficient High-integrity Multicore systems." URL: http://www.concerto-project.org.

[87] I. Šljivo, B. Gallina, J. Carlson, H. Hansson, and S. Puri, "A method to generate reusable safety case argument-fragments from compositional safety analysis," *Journal of Systems and Software*, vol. 131, pp. 570–590, 2017. DOI: https://doi.org/10.1016/j.jss.2016.07.034.

[88] L. P. Bressan, A. L. de Oliveira, L. Montecchi, and B. Gallina, "A Systematic Process for Applying the CHESS Methodology in the Creation of Certifiable Evidence," in *the 14th European Dependable Computing Conference (EDCC)*, pp. 49–56, IEEE, 2018. DOI: https://doi.org/10.1109/EDCC.2018.00019.

[89] CHESS-SBA, "CHESS State-Based Analysis," 2021. URL: https://ic.unicamp.br/ leonardo/tools.html.

[90] G. Balbo, "Introduction to Stochastic Petri Nets," in *School organized by the European Educational Forum*, pp. 84–155, Springer, 2001. DOI: https://doi.org/10.1007/3-540-44667-2_3.

[91] V. R. Basili, "Software modeling and measurement: the goal/question/metric paradigm," tech. rep., University of Maryland for Advanced Computer Studies, 1992. URL: https://dl.acm.org/doi/10.5555/137076.

[92] V. R. B. G. Caldiera and H. D. Rombach, "The goal question metric approach," *Encyclopedia of software engineering*, pp. 528–532, 1994. DOI: https://doi.org/https://doi.org/10.1002/0471028959.sof142.

[93] H. J. Holz, A. Applin, B. Haberman, D. Joyce, H. Purchase, and C. Reed, "Research methods in computing: what are they, and how should we teach them?," *ACM SIGCSE Bulletin*, vol. 38, no. 4, pp. 96–114, 2006. DOI: https://doi.org/10.1145/1189136.1189180.

[94] A. Hopkins, *Disastrous Decisions: The Human and Organisational Causes of the Gulf of Mexico Blowout*. Sydney: CCH Australi, 2012. DOI: https://doi.org/10.1093/jwelb/jws029.

[95] S. Sheikh Bahaei and B. Gallina, "Towards Qualitative and Quantitative Dependability Analyses for AR-equipped Socio-technical Systems," in *International Conference on System Reliability and Safety (ICSRS)*, IEEE, 2021. DOI: https://doi.org/10.1109/ICSRS53853.2021.9660642.

[96] D. Pavlov, I. Sosnovsky, V. Dimitrov, V. Melentyev, and D. Korzun, "Case study of using virtual and augmented reality in industrial system monitoring," in *26th Conference of Open Innovations Association (FRUCT)*, pp. 367–375, IEEE, 2020. DOI: https://doi.org/10.23919/FRUCT48808.2020.9087410.

[97] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," *Empirical software engineering*, vol. 14, no. 2, p. 131, 2009. DOI: https://doi.org/10.1007/s10664-008-9102-8.

[98] F. Ye and T. Kelly, "Component failure mitigation according to failure type," in *the 28th Annual International Computer Software and Applications Conference (COMPSAC)*, pp. 258–264, IEEE, 2004. DOI: https://doi.org/10.1109/CMPSAC.2004.1342841.

[99] C. Becker, J. C. Brewer, and L. Yount, "Safety of the Intended Functionality of Lane-Centering and Lane-Changing Maneuvers of a Generic Level 3 Highway Chauffeur System," tech. rep., United States. National Highway Traffic Safety Administration, 2020. URL: https://rosap.ntl.bts.gov/view/dot/53628/dot_53628_DS1.pdf.

[100] A. Hietanen, R. Pieters, M. Lanz, J. Latokartano, and J.-K. Kämäräinen, "AR-based interaction for human-robot collaborative manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 63, p. 101891, 2020. DOI: https://doi.org/10.1016/j.rcim.2019.101891, license: https://creativecommons.org/licenses/by-nc-nd/4.0/.

[101] S. Honig and T. Oron-Gilad, "Understanding and resolving failures in human-robot interaction: Literature review and model development," *Frontiers in Psychology*, vol. 9, 2018. DOI: https://doi.org/10.3389/fpsyg.2018.00861.

[102] "Universal robots." URL: https://www.universal-robots.com/cb3/.

[103] "RG2-Grippe." URL: https://onrobot.com/en/products/rg2-gripper.

[104] "Flaticon database of free icons." URL: https://www.flaticon.com/. Accessed: 2022-09-05.

[105] "Vecteezy resources of photography, videos and vector illustrations." URL: https://www.vecteezy.com/. Accessed: 2022-09-05.

[106] B. Kitchenham and 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.

[107] P. C. Cacciabue, "Human error risk management for engineering systems: a methodology for design, safety assessment, accident investigation and training," *Reliability Engineering & System Safety*, vol. 83, no. 2, pp. 229–240, 2004. DOI: https://doi.org/https://doi.org/10.1016/j.ress.2003.09.013.

[108] M. Schönbeck, M. Rausand, and J. Rouvroye, "Human and organisational factors in the operational phase of safety instrumented systems: A new approach," *Safety science*, vol. 48, no. 3, pp. 310–318, 2010. DOI: https://doi.org/https://doi.org/10.1016/j.ssci.2009.11.005.

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

[110] M. Farcasiu and I. Prisecaru, "MMOSA–a new approach of the human and organizational factor analysis in PSA," *Reliability Engineering & System Safety*, vol. 123, pp. 91–98, 2014. DOI: https://doi.org/https://doi.org/10.1016/j.ress.2013.10.004.

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

[112] A. Mazaheri, J. Montewka, J. Nisula, and P. Kujala, "Usability of accident and incident reports for evidence-based risk modeling–A case study on ship grounding report," *Safety science*, vol. 76, pp. 202–214, 2015. DOI: https://doi.org/https://doi.org/10.1016/j.ssci.2015.02.019.

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

[114] A. Dehghan Nejad, R. Gholamnia, and A. Alibabaee, "A new framework to model and analyze organizational aspect of safety control structure," *International Journal of System Assurance Engineering and Management*, vol. 8, no. 2, pp. 1008–1025, 2017. DOI: https://doi.org/https://doi.org/10.1007/s13198-016-0561-9.

[115] C. Leong, T. Kelly, and R. Alexander, "Incorporating epistemic uncertainty into the safety assurance of socio-technical systems," *Electronic Proceedings in Theoretical Computer Science*, vol. 259, pp. 56–71, 2017. DOI: https://doi.org/https://doi.org/10.4204%2Feptcs.259.7.

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

[117] P.-c. Li, L. Zhang, L.-c. Dai, X.-f. Li, and Y. Jiang, "A new organization-oriented technique of human error analysis

in digital NPPs: Model and classification framework," *Annals of Nuclear Energy*, vol. 120, pp. 48–61, 2018. DOI: https://doi.org/https://doi.org/10.1016/j.anucene.2018.05.021.

[118] E. Zarei, M. Yazdi, R. Abbassi, and F. Khan, "A hybrid model for human factor analysis in process accidents: FBN-HFACS," *Journal of loss prevention in the process industries*, vol. 57, pp. 142–155, 2019. DOI: https://doi.org/https://doi.org/10.1016/j.jlp.2018.11.015.

[119] A. Lališ, R. Patriarca, J. Ahmad, G. Di Gravio, and B. Kostov, "Functional modeling in safety by means of foundational ontologies," *Transportation research procedia*, vol. 43, pp. 290–299, 2019. DOI: https://doi.org/https://doi.org/10.1016/j.trpro.2019.12.044.

[120] B. Accou and G. Reniers, "Developing a method to improve safety management systems based on accident investigations: The SAfety FRactal ANalysis," *Safety science*, vol. 115, pp. 285–293, 2019. DOI: https://doi.org/https://doi.org/10.1016/j.ssci.2019.02.016.

[121] G. Fu, X. Xie, Q. Jia, Z. Li, P. Chen, and 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*, vol. 134, pp. 47–82, 2020. DOI: https://doi.org/https://doi.org/10.1016/j.psep.2019.11.027.

[122] J. I. Single, J. Schmidt, and J. Denecke, "Ontology-based computer aid for the automation of HAZOP studies," *Journal of Loss Prevention in the Process Industries*, vol. 68, p. 104321, 2020. DOI: https://doi.org/https://doi.org/10.1016/j.jlp.2020.104321.

[123] B. Ryan, D. Golightly, L. Pickup, S. Reinartz, S. Atkinson, and N. Dadashi, "Human functions in safety-developing a framework of goals, human functions and safety relevant activities for railway sociotechnical systems," *Safety Science*, vol. 140, p. 105279, 2021. DOI: https://doi.org/10.1016/j.ssci.2021.105279.

[124] N. Chouchani, S. Debbech, and M. Perin, "Model-based safety engineering for autonomous train map," *Journal of Systems and Software*, vol. 183, p. 111082, 2022. DOI: https://doi.org/https://doi.org/10.1016/j.jss.2021.111082.

[125] H. Gürlük, O. Gluchshenko, M. Finke, L. Christoffels, and L. Tyburzy, "Assessment of risks and benefits of context-adaptive augmented reality for aerodrome control towers," in *Digital Avionics Systems Conference (DASC)*, pp. 1–10, IEEE, 2018. DOI: https://doi.org/10.1109/DASC.2018.8569859.

[126] R. R. Lutz, "Safe-AR: Reducing risk while augmenting reality," in *29th International Symposium on Software Reliability Engineering (ISSRE)*, pp. 70–75, IEEE, 2018. DOI: https://doi.org/10.1109/ISSRE.2018.00018.

[127] S. Debbech, S. C. Dutilleul, and P. Bon, "An Ontological Approach to Support Dysfunctional Analysis for Railway Systems Design," *Journal of Universal Computer Science*, vol. 26, no. 5, pp. 549–582, 2020. DOI: https://doi.org/https://doi.org/10.3897/jucs.2020.030.

[128] T. Aven, "Risk assessment and risk management: Review of recent advances on their foundation," *European Journal of Operational Research*, vol. 253, no. 1, pp. 1–13, 2016. DOI: https://doi.org/https://doi.org/10.1016/j.ejor.2015.12.023.

[129] H. J. Pasman, W. J. Rogers, and 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*, vol. 55, pp. 80–106, 2018. DOI: https://doi.org/https://doi.org/10.1016/j.jlp.2018.05.018.

[130] A. Adriaensen, W. Decré, and 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*, vol. 5, no. 4, 2019. DOI: https://doi.org/10.3390/safety5040065.

[131] R. Sadeghi and F. Goerlandt, "The State of the Practice in Validation of Model-Based Safety Analysis in Socio-Technical Systems: An Empirical Study," *Safety*, vol. 7, no. 4, 2021. DOI: https://doi.org/https://doi.org/10.3390/safety7040072.

[132] N. Berx, W. Decré, I. Morag, P. Chemweno, and L. Pintelon, "Identification and classification of risk factors for human-robot collaboration from a system-wide perspective," *Computers*

& Industrial Engineering, vol. 163, p. 107827, 2022. DOI: https://doi.org/https://doi.org/10.1016/j.cie.2021.107827.

[133] D. Passenier, A. Sharpanskykh, and R. J. de Boer, "When to STAMP? A case study in aircraft ground handling services," *Procedia Engineering*, vol. 128, pp. 35–43, 2015. DOI: https://doi.org/https://doi.org/10.1016/j.proeng.2015.11.502.

[134] N. Leveson, "A new accident model for engineering safer systems," *Safety Science*, vol. 42, no. 4, pp. 237–270, 2004. DOI: https://doi.org/https://doi.org/10.1016/S0925-7535(03)00047-X.

[135] X. Zhang, Y. Sun, Y. Zhang, and S. Su, "Multi-agent modelling and situational awareness analysis of human-computer interaction in the aircraft cockpit: A case study," *Simulation Modelling Practice and Theory*, vol. 111, p. 102355, 2021. DOI: https://doi.org/https://doi.org/10.1016/j.simpat.2021.102355.

[136] I. T. de Souza, A. C. Rosa, A. C. J. Evangelista, V. W. Tam, and A. Haddad, "Modelling the work-as-done in the building maintenance using a layered FRAM: A case study on HVAC maintenance," *Journal of Cleaner Production*, vol. 320, p. 128895, 2021. DOI: https://doi.org/https://doi.org/10.1016/S0925-7535(03)00047-X.

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

[138] A. Hanna, K. Bengtsson, P.-L. Götvall, and M. Ekström, "Towards safe human robot collaboration-risk assessment of intelligent automation," in *the 25th International Conference on Emerging Technologies and Factory Automation (ETFA)*, vol. 1, pp. 424–431, IEEE, 2020. DOI: https://doi.org/https://doi.org/10.1109/ETFA46521.2020.9212127.

[139] R. Inam, K. Raizer, A. Hata, R. Souza, E. Forsman, E. Cao, and S. Wang, "Risk assessment for human-robot collaboration in an automated warehouse scenario," in *23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, vol. 1, pp. 743–751, IEEE, 2018. DOI: https://doi.org/https://doi.org/10.1109/ETFA.2018.8502466.

II

Included Papers

Chapter 7

Paper A: A Metamodel Extension to Capture Post Normal Accidents in AR-equipped Socio-technical Systems

Soheila Sheikh Bahaei, Barbara Gallina.
In Proceedings of the 31th European Safety and Reliability Conference (ESREL-2021), Research Publishing, Singapore, September 2021.

Abstract

In the past twenty to thirty years, organizations have extremely changed and these changes in addition to technological changes such as use of augmented reality (AR) introduce new system risks. Post normal accidents theory describes that organizations are more globalized and digitalized and are formed as networks of organizations, which would lead to post normal accidents such as network failure accident. In addition, it states that strategies and organizational structures are more financialized and networked respectively and technology and task are more digitalized and standardized. These organizational factors affect also on human performance. Organization and human are considered as the socio parts of socio-technical systems. Metamodels should provide the modeling elements required for modeling human and organizational factors in new AR-equipped socio-technical systems. Current metamodels do not consider factors that would lead to post normal accidents. In this paper, we elaborate the theory of post normal accidents and we extract the influencing factors leading to post normal accidents. We also consider global distance including geographical, temporal and cultural distances, as an influencing factor on human performance. Then, we use the extracted influencing factors for extending modeling elements in our previously proposed conceptual metamodel for modeling AR-equipped socio-technical systems. Our proposed extended metalmodel can be used by analysis techniques in order to perform risk assessment for AR-equipped socio-technical systems.

7.1 Introduction

Significant changes in organizations over the past two to three decades besides utilizing new technologies such as augmented reality (AR) would act as new causes of accidents. The theory of post normal accident [1], which is an extension of normal accident theory [2], has highlighted the important changes of organizations over the last two to three decades. Based on this theory, technology and task are more digitalized and standardized in comparison to 1980s, which were more automated. Organizational structures are more networked (externalized, horizontal) in comparison to 1980s, which were more integrated (internalized, vertical). In addition, Organizational strategies are more financialized in comparison to 1980s, which were only industrial. Furthermore, environments are more globalized and self-regulated, while they were national and state regulated during 1980s. Effect of these changes on human performance is not negligible and thus it is crucial to investigate it, since both human and organization take part as the socio entities of socio-technical systems. Global distance metric [3] is also a new metric capturing a new influencing factor on human. It is defined as distances in geographical, temporal and cultural features of people working in an organization [4]. It is now well established that this metric affects on human performance [5]. There is a need to address these factors originated in recent changes, which would be the reasons for new types of accidents called post normal accidents.

In order to perform risk assessment, which plays a key role in different phases of product development in system engineering, modeling the system plays a vital role. UML (Unified Modeling Language)-based metamodels [6] are the most widely used groups of metamodels and have been extensively used for defining means required for modeling the involved system. SafeConcert [7] is a metamodel proposed for modeling socio-technical systems. This metamodel is implemented by CHESS ML (CHESS Modeling Language) [8], which is a UML-based modeling language in CHESS framework [9]. Effect of new technologies such as Augmented Reality (AR) on human and organizational factors have been explored in several studies [10] [11]. There is no detailed investigation into the effects of organizational changes leading to post normal accidents on modeling. The objectives of this research are investigating the effects of the new organizational changes on modeling and updating available metamodels to enable capturing post normal accidents in AR-equipped socio-technical systems. In order to do that, we extract the new influencing factors on human performance based on post normal accident theory and global distance metric, and we integrate these factors in the previously proposed con-

ceptual metamodel for modeling AR-equipped socio-technical systems. In addition, this research provides a potential usage of the extended metamodel on an example from petroleum domain.

The rest of the paper is organized as follows. In Section 7.2, we provide essential background information. In Section 7.3, we propose a metamodel extension, based on post normal accident theory and global distance metric, on our previously proposed conceptual metamodel. In Section 7.4, we discuss about the strength and limitation of the proposed extension. Finally, in Section 7.5, we present some concluding remarks and discuss about future work.

7.2 Background

In this section, we provide essential background information about modeling AR-equipped socio-technical systems, post normal accident theory and global distance metric.

7.2.1 Modeling AR-equipped Socio-technical Systems

There are different metamodels used for modeling various types of systems. SafeConcert [7] is a metamodel, which proposes constructs for modeling socio-technical systems. It is implemented within CHESS ML/CHESS Toolset, which is integrated in the AMASS platform [12]. AMASS platform is the first open-source platform that supports engineering and certification processes of safety-critical systems. Main elements in this metamodel are components, ports, and connectors. These elements are used for modeling main entities of a socio-technical system. Failure modes and failure behaviors are also used for modeling behaviors of system elements.

Main entities of a socio-technical system are software and hardware, which are the technical entities and human and organization, which are the socio entities. Each of these entities are modeled as components and their relations are modeled through connectors. Components can contain sub-components. Sub-entities in technical entities are modeled as sub-components, while in socio entities different aspects are modeled as sub-components. For example, in an organization, examples of different aspects are process management and resource management. In a human, examples of different aspects are human characteristics such as sensing and executing.

In [13], extensions are proposed for this metamodel in order to incorporate AR related factors. As it is shown in Figure 7.1, AR-equipped socio-technical

system is a system which has augmented reality technology in addition to usual socio and technical entities. This technology affects on human and organization. Human using augmented reality would have extended capabilities, which are required to be modeled in order to consider their failure behavior while doing risk assessment. For example, with the use of augmented reality a person can sense surrounding environment, thus surround sensing is an AR-extended characteristic for human.

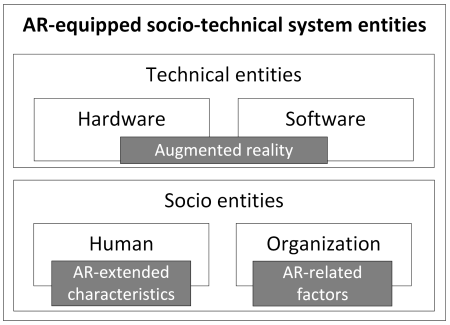


Figure 7.1: AR-equipped socio-technical system entities

As it is shown in Figure 7.2, entities, their characteristics and their relations are modeled using components, sub-components and connectors. Subcomponents of human and organization are selected based on SafeConcert human and organization modeling elements and AR-related modeling extensions. The factors with gray color are the conceptual extensions. Organizational factors are based on several state-of-the-art taxonomies such as Rasmussen [14], HFACS [15], SERA [16] and SPAR-H [17] and AR-related factors are added based on studies and experiments on AR such as [18] and [19].

7.2.2 Post Normal Accident Theory

Post normal accident theory [1], which is an extension for normal accident theory [2], is proposed by Jean-Christophe Le Coze. Perrow's normal accident theory argues that in tightly complex systems accidents are unavoidable or normal. Four analytical categories are also argued by Perrow to provide strong understanding of the situations which happen in organizations. These four categories are technology and task, structure, goal (later updated to strategy by Jean-Christophe Le Coze) and environment. Post normal accident theory argues that because of advent of new notions such as *globalization*, an update or

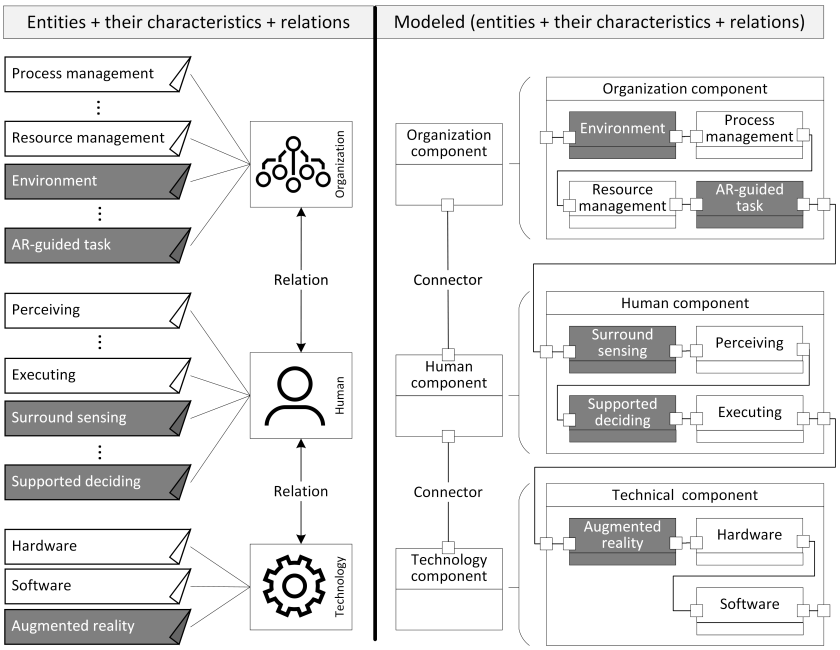


Figure 7.2: AR-equipped socio-technical system Modeling

adaptation for normal accident theory is required. In this theory, goal category is updated to strategy and features of the four categories (environment, strategy, structure, technology and task) are compared during 1980s and 2010s (Shown in Figure 7.3). Post normal accident theory, illustrates implications of trends such as *digitalization*, *standardization*, *financialization* and *self-regulation* on these four layers.

It discusses that environment was national and state regulated during the time normal accident theory was proposed (1980s). However, it is more *globalized* and *self-regulated* during 2010s. Based on definition provided in [20], *globalization* referes to “extended financial environment and greater exposure, worldwide competition, work and labour flexibility, incentives to breakdown vertical structures to gain flexibility through novel and expanding ICT networked infrastructure, normalized practices and dependence on a growing service activity (e.g. consulting)”. *Self-regulation* refers to ”industry regulating itself through the production of its own standards and internal control”.

Analytical categories	Features based on NA (1980s)	Features based on Post NA (2010s)
Environment	*National & state regulated*	*Globalized & self-regulated*
Strategy	*Industrial*	*Financialized and Industrial*
Structure	*Integrated (internalized, vertical)*	*Networked (externalized, horizontal)*
Technology and task	*Automated*	*Digitalized & standardized*

Figure 7.3: Post normal accident theory [1]

Strategy was more industrial during 1980s, while it is more *financialized* and industrial during 2010s. *Financialization* refers to "increasing the influence of financial actors (e.g. hedge funds) in companies' managerial decision-making processes".

Structure was more integrated during 1980s, while it is more networked during 2010s.

Finally technology and task were more automated during 1980s, while they are more *digitalized* and *standardized* during 2010s. *Digitalization* refers to "the progressive replacement or extension of human activities by a combination of ICT systems and machines (or robots) which can perform an increasingly wide range of manual and cognitive tasks more and more independently". *Standardization* refers to "widespread management principles promoted by outsourcing and self-regulation, consulting firms and certification schemes for global markets".

As it is discussed in [20], recent changes introduce new safety challenges and besides their provided progress, they would be source of harm. It is also stated in this study that looking into new categories of system risks is required as a complementary perspective for the study.

7.2.3 Global Distance Metric

Global distance metric [3] has been suggested by Noll and Beecham, for global distance measurement between distributed sites on Global Software Development (GSD) [21]. Geographic, temporal and cultural distances are considered and quantified in this metric. For example, for organization buildings in dif-

ferent countries a higher impact value is considered in comparison to buildings in the same region or in the same campus. Similarly, for temporal and cultural distance different impact values are considered. It is also discussed in this study that global distance would obstruct the communication among people in distributed teams.

In [22], an evaluation is designed to test cultural difference in understanding graphical symbols such as icons used in technological devices. US and Swedish subjects are evaluated and the results show that culture influences on their certainties for graphical symbol understanding. In [23], empirical evidence is provided showing that geographical proximity influences on social interactions and these effects even have increased by IT revolution. In [24], it is discussed that temporal distance influences on information diffusion processes in social and technological networks.

Based on these studies, global distance can be considered as an influencing factor on human performance. For example, a safety manager would live in a country with a culture that human safety is not so critical, while for another safety manager, it is highly critical based on the culture of the country he is living in. Thus, there would be some misunderstanding in discussions between these two people, if they work in two different buildings of a same organization located in different countries.

7.3 Proposed Extended Metamodel

In this section, first, based on the post normal accident theory and global distance metric, discussed in Subsection 7.2.2 and 7.2.3, we extract the factors influencing on human performance leading to accident. Then, we extend the organization and human modeling elements in our previously proposed conceptual metamodel, which was briefly explained in Subsection 7.2.1.

7.3.1 Extracted Influencing Factors

Influencing factors are selected, if they have the potential to influence on human performance leading to accidents. Definitions and safety effects discussed in [20] and [4] are used for identifying these factors. We explained about the definitions in Subsection 7.2.2 and 7.2.3. In this subsection, we extract and categorize these influencing factors. Safety effects are also provided based on [20].

Extracted influencing factors on human performance are divided into two

groups. The first group is organizational factors and the second group is human factors. These two groups are as follows:

1. Group 1 (organizational factors)
   - **Globalized environment**: It may cause complex interactions between different entities. These implications may affect on human performance and would lead to an accident.
   - **Self-regulated environment**: It may cause missing of independent oversights by states that may affect on human performance and would lead to an accident.
   - **Organizational strategy**:
     - **financialized strategy**: It may cause pressure for returning the investment and shifting power to financial actors. These implications may affect on human performance and would lead to an accident.
     - **Industrial strategy**: It may cause changes in industrial relations that may affect on human performance and would lead to an accident.
   - **Organizational structure**:
     - **Networked structure**: It may cause increase in complexity of interactions across organizations and other entities of the system that may affect on human performance and would lead to an accident
   - **Digitalized task**: It may cause complexity in human and machine interactions and development of new information structures. These changes may affect on human performance and would lead to an accident.
   - **Standardized task**: It may cause change in practices that may affect on human performance and would lead to an accident.
2. Group 2 (human factors)
   - **Global distance**:
     - **Geographic distance**: It may cause difficulties in managing physical places that may affect on human performance and would lead to an accident.
     - **Temporal distance**: It may cause difficulties in time management that may affect on human performance and would lead to an accident.
     - **Cultural distance**: It may cause difficulties in communications that may affect on human performance and would lead to an accident.

7.3.2 Extended Modeling Elements

The first group of factors explained in Subsection 7.3.1 can be used for extending organization modeling elements and the second group can be used for extending human modeling elements.

Based on the provided definitions for each of the extracted influencing factors and based on the three categories of organization modeling elements proposed in [10], we add new modeling elements to the categories, shown in Figure 7.4. The components with dotted line border are AR-extended components, which were proposed in our previous extension. Extended modeling elements in this paper are shown with gray color and our previous categorization of meta classes are shown with white color.

For example, time pressure is an organizational modeling element using to model scenarios that time pressure would influence on human performance and would lead to system failure or an accident. AR guided task refers to a task that AR is used for guiding the operator for doing the task. If this task is not defined correctly, it would influence on human performance leading to system failure. Standardized task is an extended modeling element proposed in this paper based on post normal accident theory. Standardization would influence on human performance and would lead to an accident.

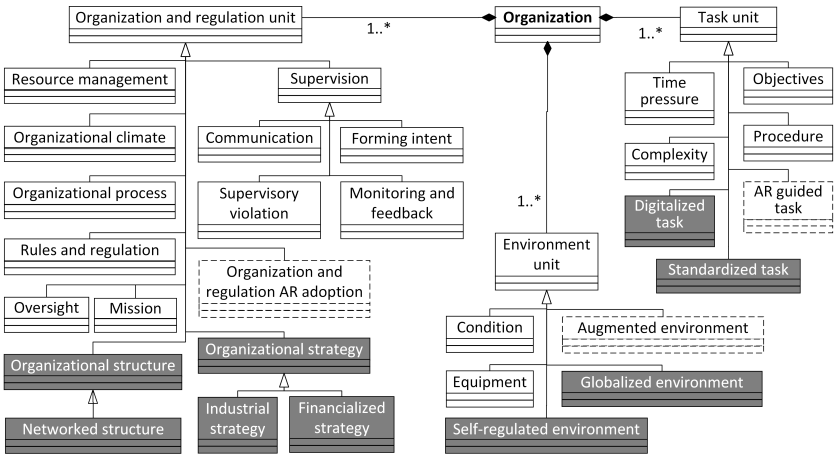


Figure 7.4: Extended organization modeling elements

We also use global distance metric for extending human modeling elements, shown in Figure 7.5. The components with dotted line border are AR-

extended components, which were proposed in our previous extension. Extended modeling elements in this paper are shown with gray color. For example, social modeling element is a human modeling element. This modeling element can be used for modeling scenarios that problem in communication between people would lead to misunderstanding and failure in human performance. Thus, it would lead to an accident. Social presence modeling element can be used for modeling scenarios that using AR would decrease social presence, meaning that people miss their communication because of AR. Thus, it would influence on human performance and it would lead to an accident. Global distance is the extended modeling element proposed in this paper. This modeling element can be used for modeling scenarios that for example cultural distance between people causes misunderstanding. Thus, it would influence on human performance and it would lead to an accident.

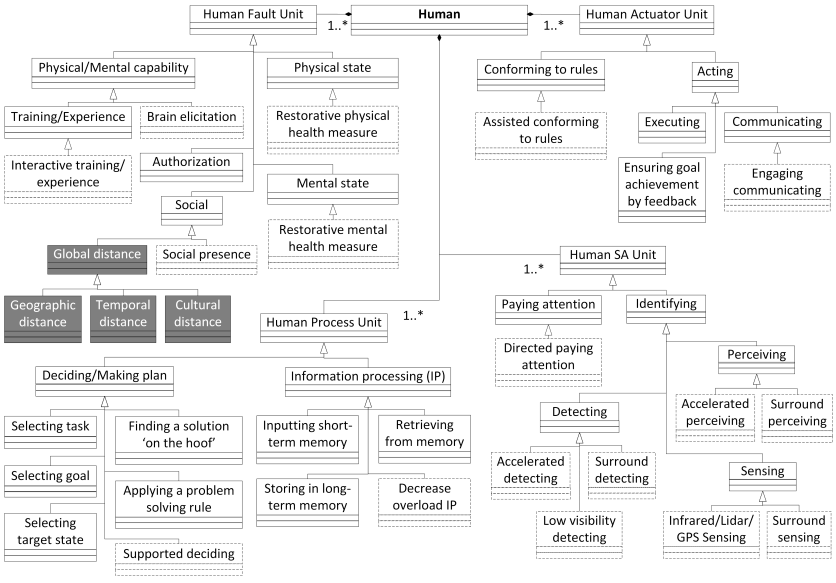


Figure 7.5: Extended human modeling elements

7.3.3 Potential Usage on an Example

British Petroleum (BP) is one of the biggest multinational companies in the world. A series of accidents between 2005 and 2010 in multinational BP in different branches occurred. We use this example to show our extension contribution in modeling conditions leading to these accidents.

Based on the analysis of these accidents using commission reports and social concepts for interpretation [25], potentials for these accidents are as follows:

- Networked structure of BP
- Lack of appropriate learning from experience
- Fault in control authority
- Strategies of CEO of the company

In Figure 7.6, we show a modeling example using our extended modeling elements. The modeling elements representing three factors leading to accident in BP, as examples, are shown using networked structure, experience and organizational strategy components. Two of these three used modeling elements are the modeling elements extended in this paper. These modeling elements, which are based on the factors explained in post normal theory as factors leading to post normal accidents are shown in gray. We show three scenarios and in each of them failure in one of the three components has contributed to accident. For example, in the first scenario (S1), output of networked structure produces a failure and the other three provide correct service, which means no failure in their outputs. Final output of the system, which is shown by OP13 produces failure because of the failure in networked structure component. In the second scenario (S2), the reason for failure in the output of the system is failure in OP4, which is output of organizational strategy component. In the third scenario (S3), the reason for failure in the output of the system is failure in OP10, which is output of the experience component. Similarly, different scenarios can be modeled and discussed using different representation means proposed in our conceptual metamodel.

Another interpretation is proposed by Jean-Christophe Le Coze in [1], in the context of globalization. In this interpretation, the author explains how deregulation, externalization, standardization, digitalization and financialization have contributed to the accidents in BP. Since our extension contains the representation means required for modeling these concepts, different scenarios

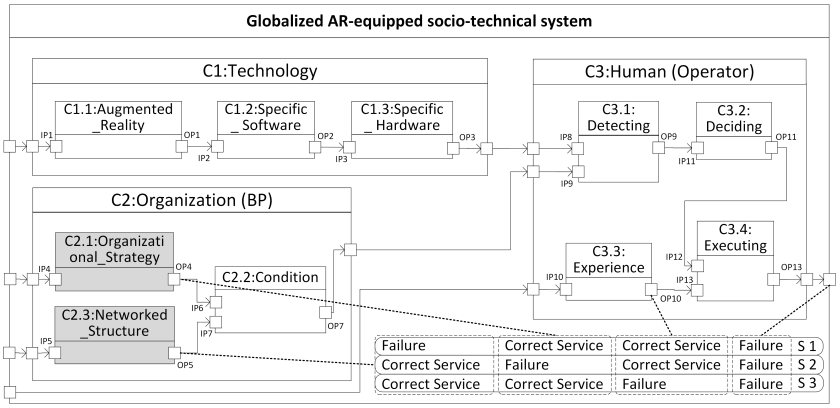


Figure 7.6: Globalized AR-equipped socio-technical system modeling

using these modeling elements can be considered and discussed during modeling and risk assessment to improve system design.

Managing multinationals is a big challenge for companies like BP. Considering technological factors and organizational factors in our previous metamodel were not enough for describing such events. We show in this example that the new proposed modeling elements can be helpful for modeling recent factors such as networked structure of an organization in order to incorporate their effect while performing risk assessment.

7.4 Discussion

In this section, we discuss about the strength and limitation of our proposed extension.

The strength of our proposed extension is provision of means in modeling process for incorporating features of new AR-equipped socio-technical systems based on an accepted theory. As it is stated in [20], it is important to investigate effects of dynamics on system risks and it is important to identify root causes of accidents in the new globalized systems to prevent post normal accidents. In this study, we took the initial step towards investigating these concepts and we updated modeling elements that can be used for modeling process as the fundamental process of risk assessment. Safety analysts can model different scenarios considering effect of globalization by discussing about the root

causes of accidents based on the updated modeling elements. Next step is to incorporate these concepts in the analysis process.

The limitation of our work is that we could not provide a complete evaluation, since there is no analysis technique to be used for globalized systems. Thus, it is required to update these techniques to be able to provide analysis results on an example. However, we provided a potential usage of our extension on an example from petroleum domain. This example can be extended and further research is required to define metrics for evaluating the success of the proposed extensions in modeling and analyzing the new scenarios.

7.5 Conclusion and Future Work

New socio-technical systems containing new technologies such as augmented reality encompass contemporary organizational changes. These changes bring up new system risks, which should be considered while performing risk assessment. In this paper, we elicited new organizational and influencing factors on human performance based on normal accident theory and global distance metric. Then, we used these elicited factors for updating our previously proposed conceptual metamodel for AR-equipped socio-technical systems. There is abundant room for further progress in determining the updates for analysis techniques and providing the full process for risk assessment.

As future work, we aim at using the extended metalmodel on an industrial case study to illustrate the contribution of the extended modeling elements. In addition, we plan to use this extended metamodel for extending analysis techniques such as Concerto-FLA [26], which is an implemented technique in CHESS Toolset [27].

Bibliography

[1] J.-C. Le Coze, *Post Normal Accident: Revisiting Perrow's Classic*. CRC Press, 2020.

[2] C. Perrow, *Normal accidents: Living with high risk technologies-Updated edition*. Princeton university press, 2011.

[3] J. Noll and S. Beecham, "Measuring global distance: A survey of distance factors and interventions," in *International Conference on Software Process Improvement and Capability Determination*, pp. 227–240, Springer, 2016.

[4] S. Alajrami, B. Gallina, and A. Romanovsky, "Enabling GSD task allocation via cloud-based software processes," in *International Conference on Software Engineering Research, Management and Applications*, pp. 179–192, Springer, 2017.

[5] L. Q. Yeong, *Investigating the influence of cultural differences on systems engineering: a case study of the manned spaceflight programs of the United States and China*. PhD thesis, Massachusetts Institute of Technology, 2015.

[6] G. Booch, J. Rumbaugh, and I. Jackobson, "UML: Unified Modeling Language," 1997.

[7] L. Montecchi and B. Gallina, "SafeConcert: A Metamodel for a concerted safety modeling of socio-technical systems," in *International Symposium on Model-Based Safety and Assessment (IMBSA)*, pp. 129–144, Springer, 2017.

[8] CONCERTO D2.7, "Analysis and back-propagation of properties for multicore systems – Final Version, http://www.concerto-project.org/results," 2016.

[9] A. Debiasi, F. Ihirwe, P. Pierini, S. Mazzini, and S. Tonetta, "Model-based analysis support for dependable complex systems in chess," in *Proceedings of the 9th International Conference on Model-Driven Engineering and Software Development - Volume 1: MODELSWARD,*, pp. 262–269, INSTICC, SciTePress, 2021.

[10] S. Sheikh Bahaei, B. Gallina, K. Laumann, and M. Rasmussen Skogstad, "Effect of augmented reality on faults leading to human failures in socio-technical systems," in *International Conference on System Reliability and Safety (ICSRS)*, IEEE, 2019.

[11] S. Sheikh Bahaei, *A Framework for Risk Assessment in Augmented Reality-equipped Socio-technical Systems*. Mälardalen University Press Licentiate Theses, 2020.

[12] J. L. de la Vara, E. P. Corredor, A. R. Lopez, and B. Gallina, "The amass tool platform: An innovative solution for assurance and certification of cyber-physical systems," in *26th International Working Conference on Requirements Engineering: Foundation for Software Quality*, March 2020.

[13] S. Sheikh Bahaei and B. Gallina, "Extending safeconcert for modelling augmented reality-equipped socio-technical systems," in *International Conference on System Reliability and Safety (ICSRS)*, IEEE, 2019.

[14] J. Rasmussen, "Human errors. a taxonomy for describing human malfunction in industrial installations," *Journal of occupational accidents*, vol. 4, no. 2-4, 1982.

[15] S. A. Shappell and D. A. Wiegmann, "The human factors analysis and classification system–HFACS," tech. rep., Civil Aeromedical Institute, 2000.

[16] K. C. Hendy, "A tool for human factors accident investigation, classification and risk management," tech. rep., Defence Research And Development Toronto (Canada), 2003.

[17] D. Gertman, H. Blackman, J. Marble, J. Byers, C. Smith, *et al.*, "The SPAR-H human reliability analysis method," *US Nuclear Regulatory Commission*, vol. 230, 2005.

[18] M. Gutiérrez *et al.*, "Augmented reality environments in learning, communicational and professional contexts in higher education.," *Digital Education Review*, vol. 26, pp. 22–35, 2014.

[19] K. Lee, "Augmented reality in education and training," *TechTrends*, vol. 56, no. 2, pp. 13–21, 2012.

[20] J.-C. Le Coze, "Globalization and high-risk systems," *Policy and practice in health and safety*, vol. 15, no. 1, pp. 57–81, 2017.

[21] J. D. Herbsleb and D. Moitra, "Global software development," *IEEE software*, vol. 18, no. 2, pp. 16–20, 2001.

[22] D. P. T. Piamonte, J. D. Abeysekera, and K. Ohlsson, "Understanding small graphical symbols: a cross-cultural study," *International Journal of Industrial Ergonomics*, vol. 27, no. 6, pp. 399–404, 2001.

[23] J. Goldenberg and M. Levy, "Distance is not dead: Social interaction and geographical distance in the internet era," *arXiv preprint arXiv:0906.3202*, 2009.

[24] J. Tang, M. Musolesi, C. Mascolo, and V. Latora, "Temporal distance metrics for social network analysis," in *Proceedings of the 2nd ACM workshop on Online social networks*, pp. 31–36, 2009.

[25] A. Hopkins, *Disastrous Decisions: The Human and Organisational Causes of the Gulf of Mexico Blowout*. Sydney: CCH Australi, 2012.

[26] B. Gallina, E. Sefer, and A. Refsdal, "Towards safety risk assessment of socio-technical systems via failure logic analysis," in *International Symposium on Software Reliability Engineering Workshops (ISSRE)*, pp. 287–292, IEEE, 2014.

[27] A. Cicchetti, F. Ciccozzi, S. Mazzini, S. Puri, M. Panunzio, A. Zovi, and T. Vardanega, "CHESS: a model-driven engineering tool environment for aiding the development of complex industrial systems," in *Proceedings of the 27th IEEE/ACM International Conference on Automated Software Engineering*, pp. 362–365, ACM, 2012.

Chapter 8

Paper B: A Case Study for Risk Assessment in AR-equipped Socio-technical Systems

Soheila Sheikh Bahaei, Barbara Gallina and Marko Vidović.
Journal of Systems Architecture (JSA-2021), Elsevier, October 2021.

Abstract

Augmented Reality (AR) technologies are used as human-machine interface within various types of safety-critical systems. Several studies have shown that AR improves human performance. However, the introduction of AR might introduce risks due to new types of dependability threats. In order to avoid unreasonable risk, it is required to detect new types of dependability threats (faults, errors, failures). In our previous work, we have designed extensions for the SafeConcert metamodel (a metamodel for modeling socio-technical systems) to capture AR-related dependability threats (focusing on faults and failures). Despite the availability of various modeling techniques, there has been no detailed investigation of providing an integrated framework for risk assessment in AR-equipped socio-technical systems. Hence, in this paper, we provide an integrated framework based on our previously proposed extensions. In addition, in cooperation with our industrial partners, active in the automotive domain, we design and execute a case study. We aim at verifying the modeling and analysis capabilities of our framework and finding out if the proposed extensions are helpful in capturing system risks caused by new AR-related dependability threats. Our conducted qualitative analysis is based on the Concerto-FLA analysis technique, which is included in the CHESS toolset and targets socio-technical systems.

8.1 Introduction

Several studies have shown that Augmented Reality (AR) technology contributes to human performance [1]. The combination of the AR technology and humans constitute an AR-equipped socio-technical system. We focus on AR-equipped socio-technical systems because of the context of the ImmerSAFE project [2] and also due to the increased AR applications. AR technology superimposes virtual and computer generated information on the reality of the user [3]. The information can be visual, auditory, etc., for enhancing human capabilities [4]. An example of visual augmented reality is using navigational information superimposed on the windshield of a car for driver guidance.

In some cases the inclusion of the AR technology might undermine user reaction. For example, it can increase cognitive-processing load [4] and it would lead to new risks. Thus, exploiting AR in socio-technical systems demands risk assessment to make sure that it is not harmful for people and the environment. To assess risk of socio-technical systems equipped with augmented reality, it is required to identify new uncertainties, threats and their propagation.

According to ISO 26262 [5], the automotive standard for functional safety, risk assessment is a "method to identify and categorize hazardous events of items and to specify safety goals and ASILs (Automotive Safety Integrity Level) related to the prevention or mitigation of the associated hazards in order to avoid unreasonable risk". The focus in this standard is on risks emanated from malfunctions of electrical and/or electronic (E/E) system. In contrast, ISO/-PAS 21448:2019, defined as safety of the intended functionality (SOTIF) [6], considers risks emanated from other types of hazardous behavior related to intended functionality or performance limitation of the system. The reason for hazardous behavior in many instances is a triggering event. For example, lack of attention while driving an automated vehicle (triggering event) would lead to incorrect decision (hazardous behavior) causing system risk. For analyzing SOTIF related hazards qualitative analysis is used and ASIL is not determined [6].

In our previous works [7, 8], in order to identify AR-related dependability threats, we have proposed two taxonomies. Based on these taxonomies, we extended SafeConcert to investigate additional socio aspects and AR-related dependability threats in system architecture modeling and analysis [9, 10]. So far, an integrated framework for risk assessment of AR-equipped socio-technical systems has not been proposed. Current frameworks do not contain modeling and analysis constructs for modeling and analyzing several social aspects, AR-extended human functions and AR-related organizational factors. In addition,

there has been little investigation about how effective current modeling and analysis techniques are for industrial systems containing new technologies and if it is possible to capture risk caused by augmented reality-related dependability threats.

In this paper, we build on our previous work and provide an integrated framework for risk assessment of AR-equipped socio-technical systems. In addition, we use an industrial case study for verifying the framework in capturing risks caused by AR-related dependability threats. More specifically, in this paper, we build on our previously proposed conceptual extensions on SafeConcert metamodel [11]. SafeConcert metamodel [11] is part of the modeling language included in the CHESS framework [12] for modeling socio-technical systems. Extended metamodel provides modeling and analysis capabilities, which can be used for assessing risk of AR-equipped socio-technical systems. Concerto-FLA [13] analysis technique is also used in our framework. Concerto-FLA is an analysis technique for socio-technical systems and it uses FPTC (Fault Propagation and Transformation Calculus) [14] syntax. In addition, we provide a case study based on SEooC (Safety element out of context) concept of ISO 26262 standard. SEooC concept refers to elements that are not developed in the context of a particular vehicle. Based on this concept, assumptions should be defined for the context in which a component is going to be used [15]. Finally, we provide threats to validity and limitations and benefits of the extensions. The results of our work can support modeling items and analyzing the behavior of AR-equipped socio-technical systems in compliance with ISO 26262 and SOTIF safety standards, which can be used by stakeholders, including designers and developers.

The rest of the paper is organized as follows. In Section 8.2, we provide essential background information. In Section 8.3, we provide an integrated framework for assessing risk of AR-equipped socio-technical systems. In Section 8.4, we design and conduct the case study to verify modeling and analysis capabilities of the proposed framework and we discuss about lessons learnt based on limitations and benefits of our research. In Section 8.5, we discuss about threats to validity in relation to our research. In Section 8.6, we provide a discussion about the contribution of our proposed extensions in determining ISO 26262 controllability and other applications. In Section 8.7, we extensively discuss related works. Finally, in Section 8.8, we present some concluding remarks and sketch future work.

8.2 Background

This section provides essential background information onto which our work is based. First, CHESS framework is introduced. Then, SafeConcert metamodel and AR-related modeling extensions are presented. FPTC syntax and Concerto-FLA analysis technique are also explained. Finally, ISO 26262, SOTIF, SEooC and SAE automation levels are presented.

8.2.1 CHESS Framework

CHESS framework [12] provides a methodology, a language and a toolset for developing high-integrity systems.

The CHESS methodology, which is component-based and model-driven, is based on an incremental and iterative process. Based on this methodology, components are defined incrementally with functional and also extra-functional properties, such as dependability information [16]. Then, developers can use a set of analysis techniques and back propagate the results iteratively.

CHESS-ML (CHESS Modeling Language) [17] is based on UML and provides the modeling elements required for modeling high-integrity systems.

CHESS toolset includes a set of plugins for code generation and provides various analysis capabilities. For example, Concerto-FLA (Failure Logic Analysis) [13] is a plugin related to analysis. In Concerto-FLA, component-based model of the system is provided and dependability information is used for decorating components. Then, analysis results can be back propagated to the system model. In this paper, we use Concerto-FLA as the analysis technique.

8.2.2 SafeConcert and its Extension of AR

SafeConcert [11] is a metamodel for modeling socio and technical entities in socio-technical systems. In this metamodel, which is part of the CHESS-ML modeling language [17], technical (i.e., software, hardware) or socio entities can be modeled as components/composite components in component-based systems representing socio-technical systems. SERA taxonomy [18] is used for modeling human and organization, which are the socio entities of the system. In this metamodel human sub-components are modeled based on twelve categories of human failures including failures in perception, decision, response, etc.

In [9], we extended the human modeling elements based on AREXTax, which is an AR-extended human function taxonomy [7]. This taxonomy is

obtained by harmonizing about six state-of-the-art human failure taxonomies (Norman [19], Reason [20], Rasmussen [21], HFACS (Human Factor Analysis and Classification System) [22], SERA (Systematic Error and Risk Analysis) [18], Driving [23]) and then extending the taxonomy based on various studies and experiments on augmented reality. These extended modeling elements are divided to four categories, shown in Figure 8.1. Three of these categories are human functions including human process unit, human SA (situational awareness) unit, and human actuator unit. The one other category is human fault unit, which is related to human internal influencing factors affecting on human functions. We explain these modeling elements in the next two paragraphs. In the first paragraph we explain modeling elements related to human functions and in the second paragraph we explain modeling elements related to human fault unit and also other fault categories. Extended modeling elements are shown with white color and AR-stemmed modeling elements are shown with dotted line border.

The extended modeling elements in human process unit, human SA unit, and human actuator unit enable modeling of AR-extended human functions. For example, detection failure, which represents a failure in *detecting* human function, is a human failure introduced by several human failure taxonomies such as Reason [20] and Rasmussen[21] taxonomies. Based on experiments and studies on augmented reality including [24] and [25], *detecting* function would be extended to *surround detecting* while using AR (surrounding information would be augmented on real world view of the user by AR). Thus, *surround detecting* can be considered as an extended sub-component of human component; in other words *surround detecting* is an extended modeling element proposed for analysis of AR-equipped socio-technical systems.

In [10], we extended organization and human modeling elements based on AREFTax, which is a fault taxonomy including AR-caused faults [8]. This taxonomy is obtained by harmonizing about five state-of-the-art fault taxonomies (Rasmussen [21], HFACS [22], SERA [18], Driving [23] and SPAR-H (Standardized Plant Analysis Risk Human Reliability Analysis)[27]) and then extending the taxonomy based on various studies and experiments on augmented reality.

In [26], we proposed more specifications for organization and human modeling elements by considering digitalization, globalization and networked structure of organizations. More specifically, we extended the human and organization modeling elements based on post normal accident theory [28] and global distance metric [29]. The extended modeling elements are helpful to prevent post normal accidents and to include global distance metric while assessing

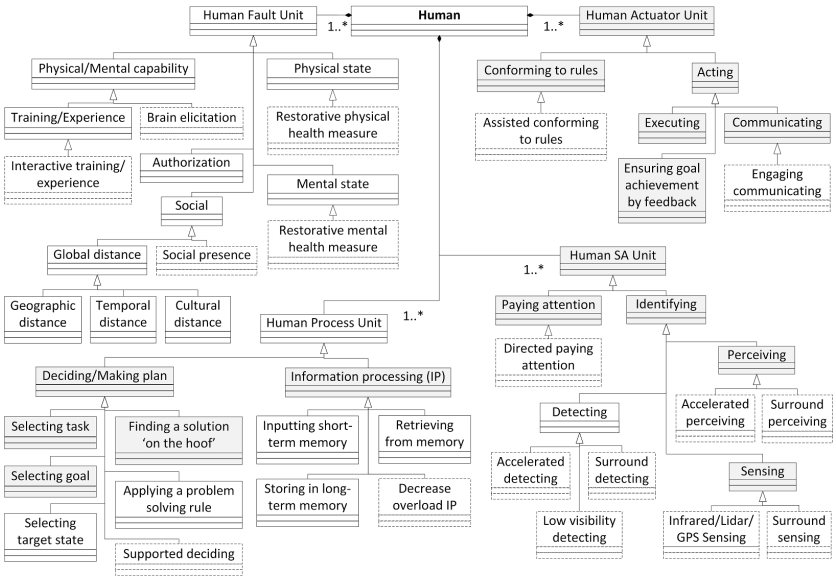


Figure 8.1: Extended modeling elements for human components [26].

risk in new AR-equipped socio-technical systems. These extended modeling elements are shown in Figure 8.2 and human fault unit of Figure 8.1. Extended modeling elements are shown with white color and AR-stemmed modeling elements are shown with dotted line border. These extended modeling elements enable modeling of various faults leading to human failures including AR-caused faults. Faults would be caused by human, environment, organization, etc. Human related faults are categorized as human fault unit of Figure 8.1 and non-human faults are categorized as three categories of organizational factors including organization and regulation unit, environment unit and task unit. For example, failure in *physical state* of a human is a human internal fault leading to human failure. This is shown as human modeling element in human fault unit category shown in Figure 8.1. Another example is *condition*, which is a non-human factor and it is categorized as extended modeling elements for organization components shown in Figure 8.2. One example of the AR-extended modeling elements is *social presence* shown in Figure 8.1. Based on studies on augmented reality [30], using AR would decrease social presence and failure in *social presence* can be considered as fault leading to human failure.

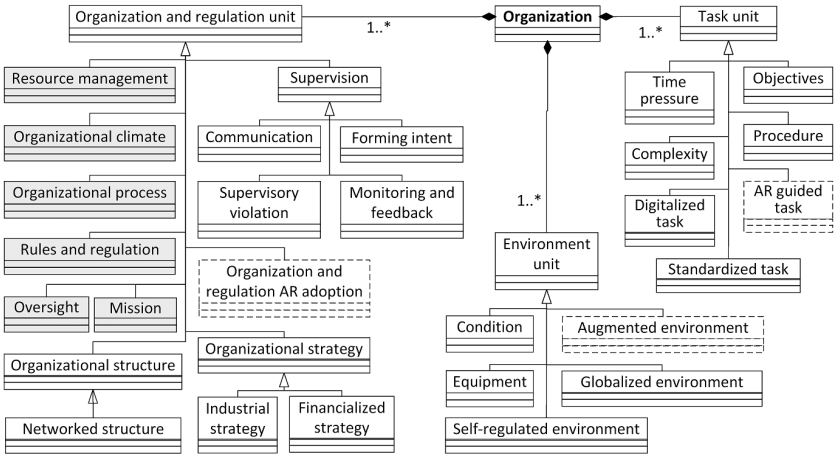


Figure 8.2: Extended modeling elements for organization components [26].

8.2.3 The FPTC Syntax

FPTC syntax was proposed as part of FPTC analysis technique [14]. FPTC rules are set of logical expressions that relate output failure modes to combinations of input failure modes in each individual component [31].

Components' behavior can be classified as source (if component generates a failure), sink (if component is able to detect and correct input failure), propagational (if component propagates failures received in its input to its output) and transformational (if component transforms the type of failure received in its input to another type in its output).

FPTC syntax for modeling failure behavior at component and connector level is as follows:

behavior = expression+

expression = LHS '→' RHS

LHS = portname'.' bL | portname

'.' bL (',' portname '.' bL) +

RHS = portname'.' bR | portname

'.' bR (',' portname '.' bR) +

failure = 'early' | 'late' | 'commission' | 'omission' |

'valueSubtle' | 'valueCoarse'

bL = 'wildcard' | bR

bR = 'noFailure' | failure

Early and late failures refer to provided function at a wrong time (early or late). Commission failures refer to provided function at a time which is not expected and omission failures refer to not provided function at a time which is expected. Value failures refer to wrong value after computations, which would be valueSubtle (user can not detect it) or valueCoarse (user can detect it).

Wildcard in an input port shows that the output behavior is the same regardless of the failure mode on this input port. NoFailure in an input port shows normal behavior.

Based on this syntax, "IP1.noFailure → OP1.omission" shows a source behavior and should be read as follows: if the component receives noFailure (normal behavior) on its input port IP1, it generates omission on its output port OP1.

8.2.4 Concerto-FLA Analysis Technique

Concerto-FLA [13], which extends FPTC [14], is a model-based analysis technique that provides the possibility for analyzing failure behavior of humans and organizations in addition to technical entities by using SERA [18] classification of socio-failures. As we recalled in Subsection 8.2.1, this technique is provided as a plugin within the CHESS toolset and allows users to define component-based architectural models composed of hardware, software, human and organization. This technique includes five main steps.

1. Modeling architectural elements including software, hardware, human, organization, connectors, interfaces and etc.
2. Modeling failure behavior at component and connector level by using FPTC syntactical rules. Concerto-FLA has adopted the FPTC syntax for modeling failure behavior at component and connector level (explained in Subsection 8.2.3).
3. Modeling failure modes at system level by injection of inputs.
4. Performing qualitative analysis through automatic calculation of the failure propagations. This step is similar to FPTC technique that system architecture is considered as a token-passing network and set of possible failures that would be propagated along a connection is called tokenset (default value for each tokenset is noFailure, which means normal behavior). In order to obtain system behavior, maximal tokenset is calculated for each connection through a fixed-point calculation.
5. Interpreting the results at system level. Based on the interpretation it will be decided to do the re-design or not.

8.2.5 ISO 26262, SOTIF, SEooC and SAE

ISO 26262 [5] is the standard for functional safety. ISO 26262 provides the requirements and set of activities that should be performed during the lifecycle phases such as development, production, operation, service and decommissioning. ISO 26262 addresses functional safety and specifies risk assessment for risks due to malfunctioning behavior of the items. If the risk is because of intended functionality or performance limitation of a system, it is addressed in ISO/PAS 21448-SOTIF [6]. In ISO 26262, ASIL (Automotive Safety Integrity Level) is determined and used for applying the requirements to avoid unreasonable residual risk. ASIL specifies item's necessary safety requirements to

achieve an acceptable residual risk. Residual risks are remaining risks after using safety measures. An ASIL value is one of four levels (A-D) and it is determined based on three factors: severity, exposure and controllability. The severity factor indicates class of severity in case of hazard occurrence and it is classified from 0 to 3 (shown by S0-S3). S3 shows the category with the highest severity and it is related to situations with life threating injuries. The exposure factor indicates class of probability of exposure with respect to operational situations and it is classified from 0 to 4 (shown by E0-E4). E4 shows the category with the highest probability of exposure (with time in use more than 10%). The controllability factor indicates the class of driver controllability and it is classified from 0 to 3 (shown by C0-C3). C3 shows the category with the highest controllability (more than 99% of drivers can control). ASIL classification based on these three factors is shown in Figure 8.3. QM (quality management) shows that no safety requirement is necessary. ASIL value A shows the lowest safety requirements and ASIL value D shows the highest safety requirements.

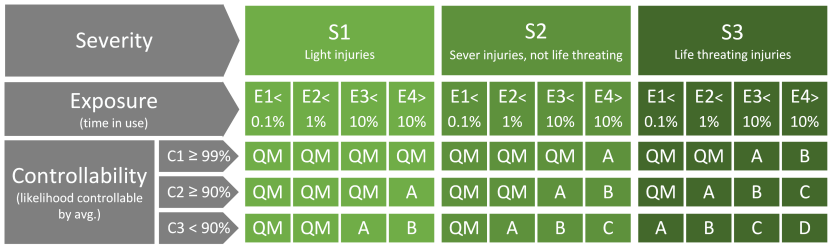


Figure 8.3: ASIL classification [32]).

Safety element out of context (SEooC), introduced by ISO 26262, refers to an element that is not defined in the context of a special vehicle, but it can be used to make an item, which implements functions at vehicle level. SEooC is based on ISO 26262 safety process and information regarding system context such as interactions and dependencies on the elements in the environment should be assumed [33].

The SEooC development contains 4 main steps:

1. (a) Definition of the SEooC scope: assumptions related to the scope, functionalities and external interfaces of the SEooC should be defined.
   (b) Definition of the assumptions on safety requirements for the SEooC: assumptions related to item definition, safety goals of the item and

functional safety requirements related to SEooC functionality, which are required for defining technical safety requirements of the SEooC should be defined.

2. Development of SEooC: based on the assumed functional safety requirements, technical safety requirements are derived and then SEooC is developed based on ISO 26262 standard.

3. Providing work products: work products are documents that show the fulfilled functional safety requirements and assumptions on the context of SEooC.

4. Integration of the SEooC into the item: safety goals and functional safety requirements defined in item development should match with assumed functional safety requirements for the SEooC. In case of a SEooC assumption mismatch, change management activity based on ISO 26262 standard should be conducted.

The process required for improving the intended functionality to ensure safety includes eight activities. Possible interactions between these activities and ISO 26262 activities and SEooC are shown in Figure 8.4.

Safety process of the ISO 26262 standard starts with *concept phase* containing *item definition*, *hazard analysis and risk assessment* and *functional safety concept* [33]. An *item* implements a vehicle level function. In *item definition* the main objective is defining items. Defining items requires defining the dependencies and interactions with environment. Then, related hazards are identified and functional safety requirements are obtained. In SEooC, assumptions related to system context are the main output of the *concept phase*. *Functional safety concept* includes providing functional safety requirements. Output provided by *Functional safety concept* is used by *technical safety concept*. *Technical safety concept* includes technical safety requirements and system design. Then, *hardware and software development* is done based on *technical safety concept*. *HW/SW development* is based on assumptions provided in concept phase. Next steps in the process are *verification test*, *validation test* and *functional safety assessment*. In SEooC, these steps require establishing validity of assumptions.

SOTIF process starts with *functional and system specification*, which includes functional description and considerations on system design and architecture. Then, potential hazardous events should be identified. If the harm is possible for the identified potentially hazardous events, then analysis of their

triggering events should be conducted. Functional modification is the next activity for avoiding the hazards or for reducing the resulting risk. Next activities are verification and validation strategy specification and then in verification and validation activities arguments are provided to illustrate that the residual risk is below acceptable level by testing on various known and unknown scenarios. Finally, evaluation on residual risk should be performed based on the verification and validation results and specified criteria.

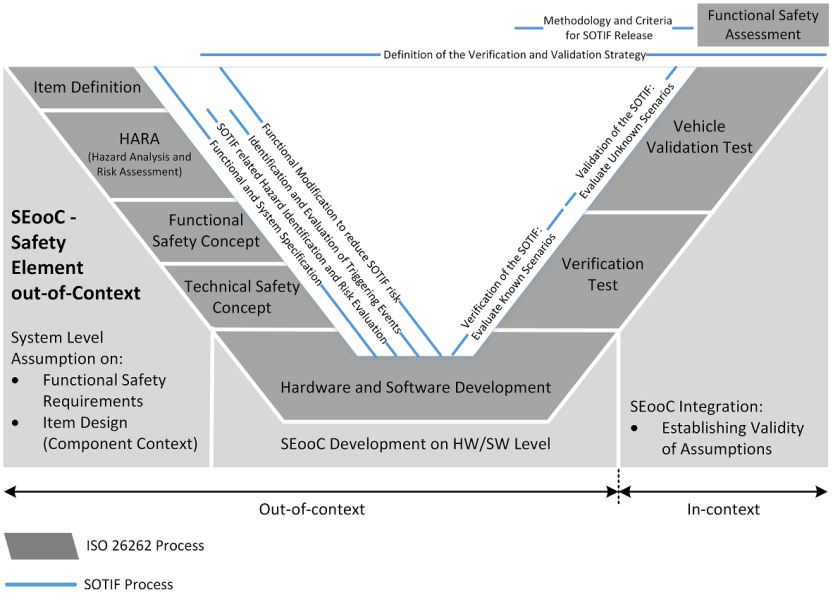


Figure 8.4: Alignment of SOTIF activities to ISO 26262 activities and SEooC (adapted from [33] and [6])

Based on the taxonomy and definitions related to driving automation systems for on-road motor vehicles performing part or the entire dynamic driving task (DDT) on a sustained basis, there are six levels of driving automation. SAE level 0 refers to no driving automation and SAE level 5 refers to full driving automation [34]. These levels with description and example are shown in Figure 8.5. Assessing human factor in driver-vehicle interface is not only important on lower SAE levels, but also on higher levels because of the importance of safe transition between automated and non-automated vehicle op-

eration [35]. In order to improve safety, various scenarios of driver/vehicle interaction should be considered.

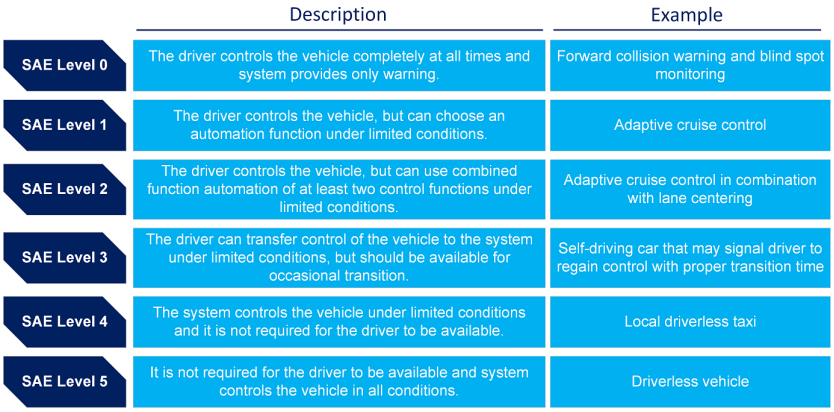


Figure 8.5: SAE levels of driving automation

8.3 An Integrated Framework for Assessing Risk of AR-equipped Socio-technical Systems

Our provided framework for assessing risk of AR-equipped socio-technical systems is based on our previously proposed modeling extensions and the Concerto-FLA analysis technique [13]. We name this framework FRAAR (Framework for Risk Assessment in AR-equipped socio-technical systems).

Our previously proposed modeling extensions on SafeConcert was recalled in Subsection 8.2.2. Concerto-FLA analysis technique [13] is also recalled in Subsection 8.2.4. Essentially, the added value with respect to SafeConcert/ Concerto-FLA is the availability of modeling and analysis capabilities for modeling and analyzing various socio aspects, AR-extended human functions and AR-related influencing factors on human functions.

We use V-model structure to illustrate methodology of the provided framework. Different steps of the methodology are shown in Figure 8.6.

As it is shown in Figure 8.6, there are four main steps.

In the first step, we need to answer to the question of what are the involved entities in the system. Since we model the system as a component-based sys-

tem, defining involved entities determines the composite components. In an AR-equipped socio-technical system, involved entities include technical (including AR) and socio entities.

In the second step, we need to identify important aspects of each entity. These important aspects are used to determine sub-components of each composite component. In this step, our proposed taxonomies and extended modeling elements explained in Subsection 8.2.2 can be helpful to have a list of important aspects. Based on scenario and the selected case study, required sub-components can be selected. For example, paying attention can be considered as an important aspect of a human driving a car. Not paying attention would lead to failure in deciding, which is a hazardous behavior that would lead to system risk.

Third step is to model the behavior of each sub-component, which should be done based on analysis of each sub-component individually. FPTC syntax explained in Subsection 8.2.3, can be used for modeling the behavior of each sub-component.

Finally, last step is analyzing system behavior, which provides system behavior based on the provided model. We do this step based on Concerto-FLA analysis technique explained in Subsection 8.2.4.

Based on the analysis results there would be feedback for changing the system design in order to decrease risk. This feedback can be suggestions for safety requirements or functional modifications.

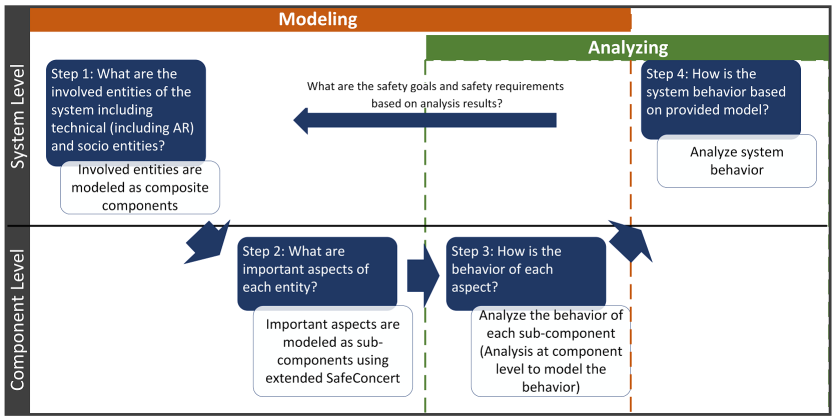


Figure 8.6: Methodology of the provided framework for assessing risk of AR-equipped socio-technical systems

Proposed risk assessment activities support several ISO 26262 and SOTIF development process activities, shown in Table 8.1. Defining involved entities in step 1 and important aspects of each entity in step 2 supports *Item definition* activity of ISO 26262 standard and *functional and system specification* of SOTIF standard. In step 1 and 2 of our proposed activities, components and sub-components are defined, which can support provision of items and functional specification. System model including all components and sub-components support provision of system specification. Provided component-based model in step 1 and 2 of our proposed framework can be used as work products expected by the standards.

Modeling important aspects of each entity, analyzing their behavior and analyzing system behavior supports *hazard analysis and risk assessment (HARA)* of ISO 26262 standard and *SOTIF related hazard identification and risk evaluation* and also *identification and evaluation of triggering events* of SOTIF standard. In *hazard analysis and risk assessment* of ISO 26262, the aims are to identify the hazards and formulating safety goals. Step 2, 3 and 4 of our proposed activities support hazard identification by modeling failure propagation and by providing analysis results of different scenarios. These results support formulating safety goals to avoid unaccepted risks. In *SOTIF related hazard identification and risk evaluation*, the aims are identifying and evaluating SOTIF related hazards and their consequences. Modeling and analyzing activities in step 2, 3 and 4 provide the support for identification and evaluation of SOTIF related hazards and their consequences. For example, failing to pay attention leads to deciding incorrectly, which is a SOTIF related hazard and it leads to executing incorrectly. The modeling elements, used in step 2 and 3, provide the possibility to model and analyze paying attention, deciding and executing functions. Analysis in step 4 also provides the consequences at system level. Provided model in step two, three and analysis results in step four can be used as work products expected by the standards.

Analyzing system behavior in step 4 also supports defining functional and technical safety requirements, which are used in *functional and technical safety concept* of ISO 26262 standard and it also supports *functional modification to reduce SOTIF risk* of SOTIF standard. In addition, analysis results are based on considering various scenarios, which support *verification test* in ISO 26262 and *verification of the SOTIF*. Required work products for verification test in ISO 26262 and SOTIF standards can be prepared based on analysis results in step four of our proposed framework.

Table 8.1: Risk assessment activities of our provided framework and supported ISO 26262 and SOTIF development process activities.

The proposed activity	ISO 26262 activity	SOTIF activity
Defining involved entities and important aspects of each entity (Step1 and 2)	Item definition	Functional and system specification
Defining important aspects of each entity, analyzing its behavior and system behavior (Step2, 3 and 4)	HARA	SOTIF related hazard identification and risk evaluation and Identification and evaluation of triggering events
Analyzing system behavior (Step 4)	Functional safety concept	Functional modification to reduce SOTIF risk
Analyzing system behavior (Step 4)	Technical safety concept	Functional modification to reduce SOTIF risk
Analyzing system behavior (Step 4)	Verification test	Verification of the SOTIF

8.4 Case Study Design and Execution

In this section, we design a case study to present the modeling and analyzing capabilities of the proposed framework that can be used to qualitatively analyze the risks for AR-equipped socio-technical systems.

8.4.1 Objectives

Our objectives include presenting the modeling capabilities and analysis capabilities of our proposed framework containing AR-related extensions. In other words, we aim at estimating the effectiveness of the provided framework in predicting risk caused by new AR-related dependability threats. In order to do that, we use an industrial case study from automotive domain to evaluate the proposed extensions. Analysis results can be used for defining related safety requirements

8.4.2 Research Methodology

We use case study research methodology based on [36]. The steps carried out for the presented research are presented in Figure 8.7. In the first step, objectives and the structure of the research are discussed.

In the second step, we asked Xylon Company for a case study in the context of augmented reality socio-technical systems. Surround view system as a case

study was suggested by this company and a meeting was organized to decide about the collaboration. We also discussed about system description.

In the third step, system architecture was provided based on information provided by the company and it was reviewed in several iterations for improvement.

In the fourth step, analysis of the case study was provided based on Concerto-FLA analysis technique and it was reviewed in iterations for improvement.

In the fifth step, a discussion about results and lessons learnt was provided. Then, the results are reviewed and a discussion about validity of the work is provided.

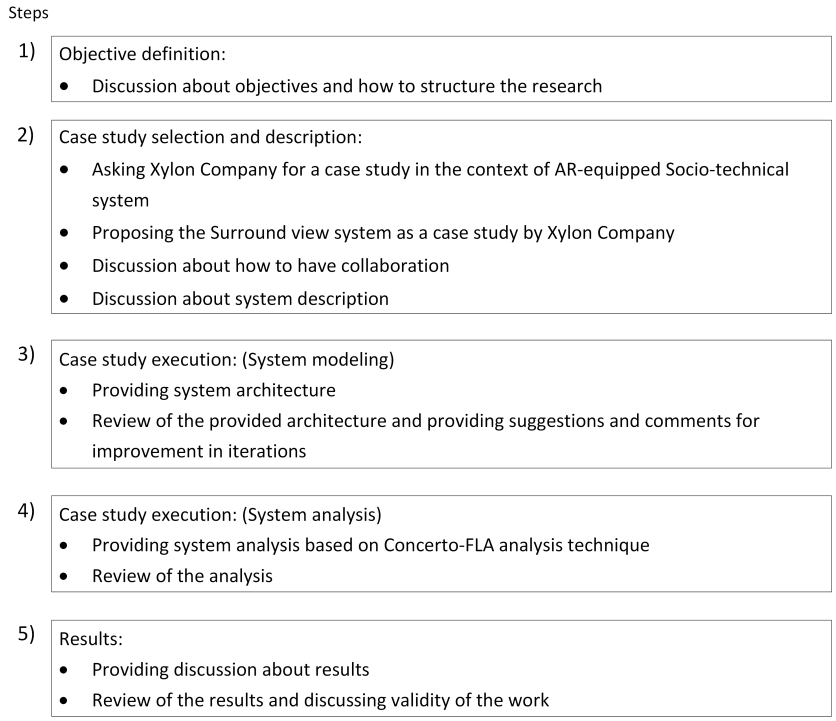


Figure 8.7: Steps taken for the carried out research

8.4.3 Case Study Selection and Description

The case study is conducted in collaboration with Xylon, an electronic company providing intellectual property in the fields of embedded graphics, video, image processing and networking.

In this study, we select as case study subject a socio-technical system containing the following entities:

- Road transport organization (socio entity): representing the organization responsible for providing transport rules and regulations, proper road conditions and etc.
- Driver (socio entity): representing a human who is expected to drive a vehicle and park it safely by utilizing augmented reality technology used in the surround view system of the vehicle.
- Vehicle (technical entity): representing vehicle containing surround view system (a SEooC with the potential for using in vehicles with high levels of driving automation. However, currently it is used at driving automation level 0. It includes augmented reality technology to empower drivers).

Surround view systems are used to assist drivers to park more safely by providing a 3D video from the surrounding environment of the car. In Figure 8.8, it is illustrated how the 3D video is shown to the driver. As it is shown in Figure 8.8, driver can have a top view of the car while driving. This top view is obtained by compounding 4 views captured by 4 cameras mounted around the car and by changing point of view. It is like there is a flying camera visualizing vehicle's surrounding, which is called virtual flying camera feature. A picture of a virtual car is also augmented to the video to show the position of the car. Navigation information and parking lines also can be annotated to the video by visual AR technology. The current surround view system is a SEooC of driving automation level 0. However, Xylon plan to develop automated driving system features in higher levels for the future versions of the system.

Assumptions on the scope of the SEooC are:

- The system can be connected to the rest of the vehicle in order to obtain speed information. In case of drawing parking path lines, steering wheel angle and information from gearbox would also be obtained to determine reverse driving.

Assumptions on functional requirements of the SEooC are:

Figure 8.8: Sample images from 3D videos provided in surround view system

- The system is enabled either at low speed or it can be activated manually by the driver.
- The system is disabled either when moving above some speed threshold or it can be deactivated by driver.

Assumptions on the functional safety requirements allocated to the SEooC are:

- The system does not activate the function at high vehicle speed automatically.
- The system does not deactivate the functionality at low speed automatically.

8.4.4 Case Study Execution: System Modeling

This subsection reports on how we model the described system in Subsection 8.4.3 using our proposed framework.

Subsection 8.4.3 provides the required information for the first step of the risk assessment process defined in Figure 8.6, which is identifying the entities for defining composite components. Based on the selected case study explained in Subsection 8.4.3, organization, driver and vehicle containing an automotive surround view system are three composite components of this system. In this subsection, we provide information for the second and third steps of risk assessment process.

Important aspects of each entity are modeled as sub-components of each composite component. For socio entities, the important aspects are selected

from extended modeling elements explained in Subsection 8.2.2 and for vehicle, which is a technical entity the important aspects are based on system description.

- Important aspects of road transport organization (selected from Figure 8.2):
  - *Organization and regulation AR adoption*: it refers to upgrading rules and regulations of road transport organization based on AR technology.
  - *Condition*: it refers to road condition.
  - *Monitoring and feedback*: it refers to the monitoring task and feedback provided by organization.
- Important aspects of driver (selected from Figure 8.1):
  - *Surround detecting*: it is an AR-extended function, because driver can detect surround environment through AR technology.
  - *Supported deciding*: it is an AR-extended function, because driver can decide with the support of AR technology.
  - *Executing*: it is human executing function.
  - *Interactive experience*: it is an AR-caused factor, because AR provides interactive ways for enhancing user experience.
  - *Social presence*: it is an AR-caused factor, because AR may decrease social presence and lead to human failure.
- Important aspects of vehicle containing surround view system (selected based on system description received from Xylon Company):
  - A set of speed sensors: each sensor is a hardware for providing speed of the vehicle based on its movement.
  - A set of cameras: each camera is a hardware for providing raw data for a video receiver. Usually there are four cameras that can be attached to four sides of the car.
  - Switch: switch is a hardware for receiving on/off command from driver. It is also possible to send on/off command automatically based on driving requirement.
  - Peripheral controller: peripheral controller includes hardware and driver for receiving user inputs such as on/off command and speed and for sending them to user application implementation.

- A set of video receivers: each video receiver includes a hardware and a driver. Its hardware is used for transforming raw data to AXI-stream based on the command from its driver implementation.
- Video storing unit: video storing unit includes a hardware and a driver. Its hardware is used for receiving AXI-stream and storing it to the memory by means of DDR memory controller based on the command received form its driver.
- DDR controller: DDR controller is a hardware for accessing DDR memory, which stores video in DDR memory and provides general memory access to all system IPs.
- Video processing IP: Video processing IP includes hardware and driver for reading prepared data structures and video from memory, for processing video accordingly and finally for storing the processed video to memory through DDR controller. The prepared data is stored to memory by video processing IP driver based on the data structures received from memory.
- Display controller: Display controller includes hardware and driver for reading memory where processed video is stored and for converting it in the format appropriate for driving displays.
- Processing unit: processing unit includes hardware and software, which its software contains all the software and drivers of all other IPs. The software also contains user application implementation and video processing engine implementation. User application implementation receives inputs from peripheral unit and controls operation of all IPs by means of their software drivers. Video processing engine implementation prepares data structures to be stored in DDR memory through DDR controller.

Figure 8.9 provides an overview of the integration of the human, organization and some of vehicle important aspects.

In Figure 8.10, we show how this AR-equipped socio-technical system is modeled using SafeConcert AR-extended modeling elements. Driver is composed of five sub-components. Driver has four inputs and two of its inputs are from system inputs with the names human detection input (HDI) and human communication input (HCI). Two other inputs are from organization and surround view system. We consider *interactive experience* and *social presence* as two sub-components of human component, which are influencing factors on human functions. *Interactive experience* affects on *supported deciding* and is affected by *surround detecting*. *Social presence* affects on human *executing*. Driver output, which is output of the system is human action shown by HA.

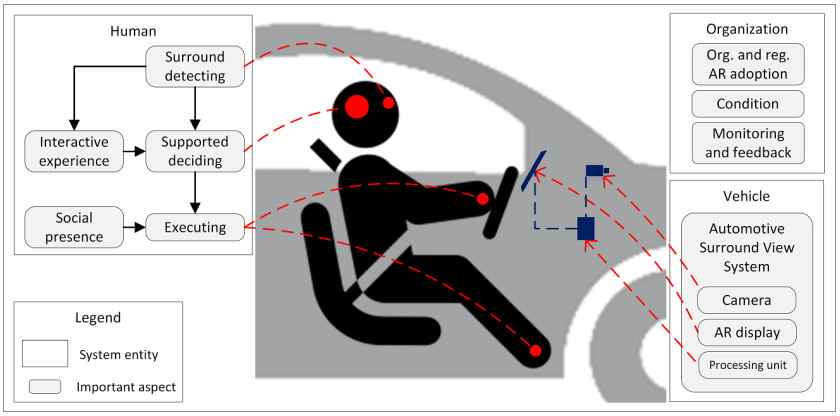


Figure 8.9: Integration of the human, organization and vehicle effective aspects

Organization and regulation AR adoption, *condition* and *monitoring and feedback* are three sub-components of organization composite component. Organization component receives input from system, which represents influences from regulation authorities on the organization (REG).

Vehicle is also modeled with three inputs including user command shown by CMD, vehicle movement shown by VMV and camera input shown by CAM. Green color is used to show the extended modeling elements used in this system.

8.4.5 Case Study Execution: System Analysis

This subsection reports on the analysis of the system using AR-related extensions, which refers to the last step of the risk assessment process defined in Figure 8.6. We follow the five steps of Concerto-FLA analysis technique explained in subsection 8.2.4 for system analysis.

1. First step is provided in Figure 8.10. We explained how the system is modeled in Subsection 8.4.4.

2. Second step is shown by providing FPTC rules, which are used for linking possible failure modes in the input of each component to the possible failure modes in the output. ”IP.variable → OP.variable” shows propagational behavior of the component, which means that any failure mode in its input is

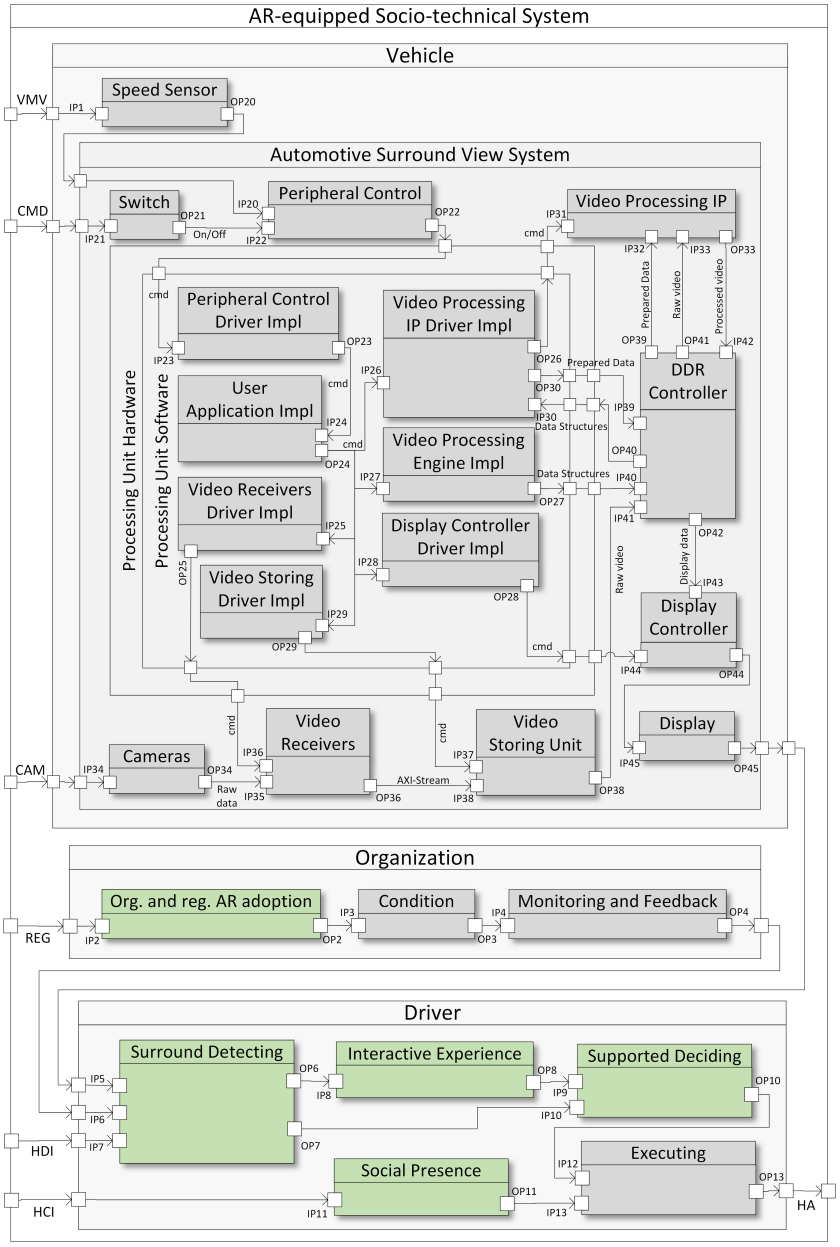


Figure 8.10: AR-equipped socio-technical system modeling

propagated to its output. FPTC rules of modeled sub-components are shown in Figures 8.11-8.14. There is one box for each component. The left part of the box shows the name of the component. The right part of the box shows possible failure modes in the input (up left), possible failure modes in the output (up right) and FPTC rules (bottom). Based on dependability-related terminology in literature such as [21], [37] and [22], we consider omission, commission, etc. as failure modes. However, these are named failures in FPTC terminology.

In this paragraph, we explain how the possible failure modes at input and output are identified/defined in Figures 8.11-8.14. For example, the camera takes in input a raw image. Based on the definitions of failure modes in Subsection 8.2.3, omission and valueSubtle are the possible failure modes for the case of camera. The reason for having omission as a possible failure mode at input is the possibility of an occlusion in front of the camera, which prevents receiving raw image as input. The reason for having valueSubtle as possible failure mode at input is the possibility of intervene, which leads to receiving input not in the expected range. For example, when image is blurred because of foggy weather. Possible failure modes at output can be obtained by considering the possible input failure modes in the FPTC rules. Defining the FPTC rules are explained in the next paragraph.

In this paragraph, we explain how the FPTC rules are defined in Figures 8.11-8.14. FPTC rules show how the component behaves. For example, the camera would not produce any failure, but if the input image is not in the expected range, then the output would not be in the expected range either. Moreover, if the input is not provided when expected, then the output would not be provided when expected. Thus, the camera propagates possible input failure modes to the output and it does not behave as source, sink and transformational (explained in Subsection 8.2.3).

In scenarios, we may change some components' failure behavior to source based on assumptions related to that scenario. For example, if we assume that an AR-related component is producing failure, then we need to change its failure behavior to source and update its FPTC rules.

3. Third step is to consider failure modes in inputs of the system to calculate failure propagation. In this case study, we inject noFailure to four inputs of the system, because we aim at analyzing system for scenarios that failure is originating from our modeled system and more specifically from our AR-related part of the system.

4. Fourth step is calculating the failure propagations. We consider three scenarios and show the analysis results in Figures 8.15 - 8.17.

5. Last step is back propagation of results. Interpretation of the back-propagated results can be used to make decision about design change or defining safety barrier, if it is required.

Scenario 1:

- **Description of the scenario:** In this scenario, we assume that failure in the system is emanated from the technical part of the system. We assume video processing IP produces processed video incorrectly. For example, we assume that the expected visual mark for parking lot striping is assigned on an incorrect position (value failure mode). As a consequence, the driver cannot detect the surround environment correctly and decides and executes incorrectly (value failure mode).

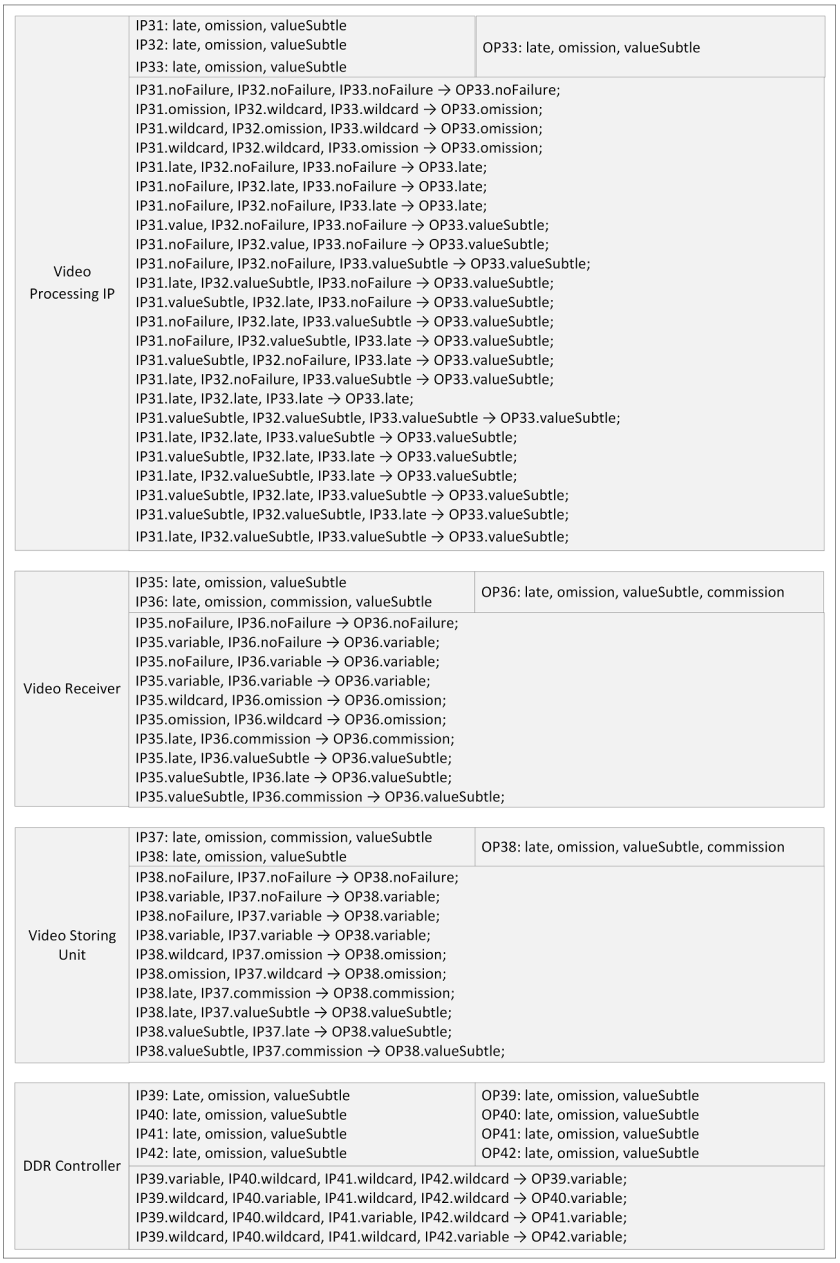
- **Modeling failure behavior:** We show the failure propagation with underlined FPTC rules, which are the rules that are activated, shown in Figure 8.15. In this scenario, video processing IP behaves as source and while its inputs are noFailure, it produces valueSubtle failure mode in its output. This activated rule is shown on its sub-component. DDR controller, display controller and display sub-components behave as propagational and propagate valueSubtle from inputs to outputs.

- **Analysis of system behavior:** ValueSubtle failure mode in IP5 means that displayed information on the display is not correct. ValueSubtle propagates to *surround detecting*, *interactive experience* and *supported deciding* and it transforms to valueCoarse in *executing*. The reason for this transformation is that if there is value failure mode in *executing* function, it can be detected by user, which means valueSubtle transforms to valueCoarse. We show the failure propagation by blue color of the underlined FPTC rules.

- **Interpreting the results:** Based on back propagation of the results, we can explain how the rules have been triggered. ValueCoarse on OP13 is because of valueSubtle on IP12. ValueSubtle on IP12 is because of valueSubtle on OP10 and we continue this back propagation to reach a component originating the failure, which is component with inputs IP31, IP32 and IP33. This component is video processing IP.

Name of the component	Possible failure modes at input	Possible failure modes at output
	FPTC rules	
Camera	IP34: omission, valueSubtle	OP34: omission, valueSubtle
	IP34.variable → OP34.variable;	
Speed Sensor	IP1: omission, valueSubtle	OP20: omission, valueSubtle
	IP1.variable → OP20.variable;	
Switch	IP21: late, omission, commission	OP21: late, commission, omission
	IP21.variable → OP21.variable;	
Peripheral Control	IP20: late, omission, valueSubtle IP22: late, omission, commission, valueSubtle	OP22: late, omission, commission, valueSubtle
	IP20.noFailure, IP22.noFailure → OP22.noFailure; IP20.variable, IP22.noFailure → OP22.variable; IP20.noFailure, IP22.variable → OP22.variable; IP20.variable, IP22.variable → OP22.variable; IP20.wildcard, IP22.omission → OP22.omission; IP20.omission, IP22.wildcard → OP22.omission; IP20.late, IP22.commission → OP22.commission; IP20.late, IP22.valueSubtle → OP22.valueSubtle; IP20.valueSubtle, IP22.late → OP22.valueSubtle; IP20.valueSubtle, IP22.commission → OP22.valueSubtle;	
Peripheral Control Driver Imp	IP23: late, omission, commission, valueSubtle	OP23: late, omission, commission, valueSubtle
	IP23.variable → OP23.variable;	
User Application Imp	IP24: late, omission, commission, valueSubtle	OP24: late, omission, commission, valueSubtle
	IP24.variable → OP24.variable;	
Video Receiver Driver Imp	IP25: late, omission, valueSubtle, commission	OP25: late, omission, commission, valueSubtle
	IP25.variable → OP25.variable;	
Video Processing Engine Imp	IP27: late, omission, valueSubtle	OP27: late, omission, valueSubtle
	IP27.variable → OP27.variable;	
Display Controller Driver Imp	IP28: late, omission, valueSubtle, commission	OP28: late, omission, commission, valueSubtle
	IP28.variable → OP28.variable;	
Video Storing Driver Imp	IP29: late, omission, valueSubtle, commission	OP29: late, omission, commission, valueSubtle
	IP29.variable → OP29.variable;	
Video Processing IP Driver Imp	IP26: late, omission, commission IP30: late, omission, valueSubtle	OP26: late, omission, commission, valueSubtle OP30: late, omission, valueSubtle
	IP26.noFailure, IP30.noFailure → OP26.noFailure, OP30.noFailure; IP26.variable, IP30.variable → OP26.variable, OP30.variable; IP30.valueSubtle, IP26.late → OP30.valueSubtle, OP26.late; IP30.wildcard, IP26.omission → OP26.omission, OP30.omission; IP30.omission, IP26.wildcard → OP30.valueSubtle, OP26.valueSubtle; IP30.late, IP26.commission → OP30.commission, OP26.valueSubtle; IP30.valueSubtle, IP26.commission → OP30.commission, OP26.valueSubtle;	

Figure 8.11: Modeling failure behavior of components



Video Processing IP	IP31: late, omission, valueSubtle IP32: late, omission, valueSubtle IP33: late, omission, valueSubtle	OP33: late, omission, valueSubtle
	IP31.noFailure, IP32.noFailure, IP33.noFailure → OP33.noFailure; IP31.omission, IP32.wildcard, IP33.wildcard → OP33.omission; IP31.wildcard, IP32.omission, IP33.wildcard → OP33.omission; IP31.wildcard, IP32.wildcard, IP33.omission → OP33.omission; IP31.late, IP32.noFailure, IP33.noFailure → OP33.late; IP31.noFailure, IP32.late, IP33.noFailure → OP33.late; IP31.noFailure, IP32.noFailure, IP33.late → OP33.late; IP31.value, IP32.noFailure, IP33.noFailure → OP33.valueSubtle; IP31.noFailure, IP32.value, IP33.noFailure → OP33.valueSubtle; IP31.noFailure, IP32.noFailure, IP33.valueSubtle → OP33.valueSubtle; IP31.late, IP32.valueSubtle, IP33.noFailure → OP33.valueSubtle; IP31.valueSubtle, IP32.late, IP33.noFailure → OP33.valueSubtle; IP31.noFailure, IP32.late, IP33.valueSubtle → OP33.valueSubtle; IP31.noFailure, IP32.valueSubtle, IP33.late → OP33.valueSubtle; IP31.valueSubtle, IP32.noFailure, IP33.late → OP33.valueSubtle; IP31.late, IP32.noFailure, IP33.valueSubtle → OP33.valueSubtle; IP31.late, IP32.late, IP33.late → OP33.late; IP31.valueSubtle, IP32.valueSubtle, IP33.valueSubtle → OP33.valueSubtle; IP31.late, IP32.late, IP33.valueSubtle → OP33.valueSubtle; IP31.valueSubtle, IP32.late, IP33.late → OP33.valueSubtle; IP31.late, IP32.valueSubtle, IP33.late → OP33.valueSubtle; IP31.valueSubtle, IP32.late, IP33.valueSubtle → OP33.valueSubtle; IP31.valueSubtle, IP32.valueSubtle, IP33.late → OP33.valueSubtle; IP31.late, IP32.valueSubtle, IP33.valueSubtle → OP33.valueSubtle;	
Video Receiver	IP35: late, omission, valueSubtle IP36: late, omission, commission, valueSubtle	OP36: late, omission, valueSubtle, commission
	IP35.noFailure, IP36.noFailure → OP36.noFailure; IP35.variable, IP36.noFailure → OP36.variable; IP35.noFailure, IP36.variable → OP36.variable; IP35.variable, IP36.variable → OP36.variable; IP35.wildcard, IP36.omission → OP36.omission; IP35.omission, IP36.wildcard → OP36.omission; IP35.late, IP36.commission → OP36.commission; IP35.late, IP36.valueSubtle → OP36.valueSubtle; IP35.valueSubtle, IP36.late → OP36.valueSubtle; IP35.valueSubtle, IP36.commission → OP36.valueSubtle;	
Video Storing Unit	IP37: late, omission, commission, valueSubtle IP38: late, omission, valueSubtle	OP38: late, omission, valueSubtle, commission
	IP38.noFailure, IP37.noFailure → OP38.noFailure; IP38.variable, IP37.noFailure → OP38.variable; IP38.noFailure, IP37.variable → OP38.variable; IP38.variable, IP37.variable → OP38.variable; IP38.wildcard, IP37.omission → OP38.omission; IP38.omission, IP37.wildcard → OP38.omission; IP38.late, IP37.commission → OP38.commission; IP38.late, IP37.valueSubtle → OP38.valueSubtle; IP38.valueSubtle, IP37.late → OP38.valueSubtle; IP38.valueSubtle, IP37.commission → OP38.valueSubtle;	
DDR Controller	IP39: Late, omission, valueSubtle IP40: late, omission, valueSubtle IP41: late, omission, valueSubtle IP42: late, omission, valueSubtle	OP39: late, omission, valueSubtle OP40: late, omission, valueSubtle OP41: late, omission, valueSubtle OP42: late, omission, valueSubtle
	IP39.variable, IP40.wildcard, IP41.wildcard, IP42.wildcard → OP39.variable; IP39.wildcard, IP40.variable, IP41.wildcard, IP42.wildcard → OP40.variable; IP39.wildcard, IP40.wildcard, IP41.variable, IP42.wildcard → OP41.variable; IP39.wildcard, IP40.wildcard, IP41.wildcard, IP42.variable → OP42.variable;	

Figure 8.12: Modeling failure behavior of components(Cont.)

Display Controller	IP43: late, omission, valueSubtle IP44: late, omission, commission, valueSubtle	OP44: late, omission, valueSubtle
	IP43.noFailure, IP44.noFailure → OP44.noFailure; IP43.variable, IP44.noFailure → OP44.variable; IP43.noFailure, IP44.variable → OP44.variable; IP43.variable, IP44.variable → OP44.variable; IP43.wildcard, IP44.omission → OP44.omission; IP43.omission, IP44.wildcard → OP44.omission; IP43.late, IP44.commission → OP44.commission; IP43.late, IP44.valueSubtle → OP44.valueSubtle; IP43.valueSubtle, IP44.late → OP44.valueSubtle; IP43.valueSubtle, IP44.commission → OP44.valueSubtle;	
Display	IP45: late, omission, commission, valueSubtle	OP45: late, omission, commission, valueSubtle
	IP45.variable → OP45.variable;	
Org. and Reg. AR Adoption	IP2: late, omission, valueSubtle, valueCoarse	OP2: late, omission, valueSubtle, valueCoarse
	IP2.variable → OP2.variable;	
Condition	IP3: late, omission, valueSubtle, valueCoarse	OP3: late, omission, valueSubtle, valueCoarse
	IP3.variable → OP3.variable;	
Monitoring and Feedback	IP4: late, omission, valueSubtle, valueCoarse	OP4: late, omission, valueSubtle, valueCoarse
	IP4.variable → OP4.variable;	
Surround Detecting	IP5: late, omission, valueSubtle IP6: late, omission, valueSubtle IP7: omission, valueSubtle, late	OP6: late, omission, valueSubtle OP7: late, omission, valueSubtle
	IP5.noFailure, IP6.noFailure, IP7.noFailure → OP6.noFailure, OP7.noFailure; IP5.omission, IP6.wildcard, IP7.wildcard → OP6.omission, OP7.omission; IP5.wildcard, IP6.omission, IP7.wildcard → OP6.omission, OP7.omission; IP5.wildcard, IP6.wildcard, IP7.omission → OP6.omission, OP7.omission; IP5.late, IP6.noFailure, IP7.noFailure → OP6.late, OP7.late; IP5.noFailure, IP6.late, IP7.noFailure → OP6.late, OP7.late; IP5.noFailure, IP6.noFailure, IP7.late → OP6.late, OP7.late; IP5.valueSubtle, IP6.noFailure, IP7.noFailure → OP6.valueSubtle, OP7.valueSubtle; IP5.noFailure, IP6.valueSubtle, IP7.noFailure → OP6.valueSubtle, OP7.valueSubtle; IP5.noFailure, IP6.noFailure, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.valueSubtle, IP7.noFailure → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.late, IP7.noFailure → OP6.valueSubtle, OP7.valueSubtle; IP5.noFailure, IP6.late, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.noFailure, IP6.valueSubtle, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.noFailure, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.noFailure, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.late, IP7.late → OP6.late, OP7.late; IP5.valueSubtle, IP6.valueSubtle, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.late, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.late, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.valueSubtle, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.late, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle; IP5.valueSubtle, IP6.valueSubtle, IP7.late → OP6.valueSubtle, OP7.valueSubtle; IP5.late, IP6.valueSubtle, IP7.valueSubtle → OP6.valueSubtle, OP7.valueSubtle;	
Interactive Experience	IP8: late, omission, valueSubtle	OP8: late, omission, valueSubtle
	IP8.variable → OP8.variable;	

Figure 8.13: Modeling failure behavior of components(Cont.)

Supported Deciding	IP9: late, omission, valueSubtle IP10: late, omission, valueSubtle	OP10: late, omission, valueSubtle
	IP9.noFailure, IP10.noFailure → OP10.noFailure; IP9.variable, IP10.noFailure → OP10.variable; IP9.noFailure, IP10.variable → OP10.variable; IP9.variable, IP10.variable → OP10.variable; IP9.wildcard, IP10.omission → OP10.omission; IP9.omission, IP10.wildcard → OP10.omission; IP9.late, IP10.valueSubtle → OP10.valueSubtle; IP9.valueSubtle, IP10.late → OP10.valueSubtle;	
Social Presence	IP11: late, omission, valueSubtle	OP11: late, omission, valueSubtle
	IP11.variable → OP11.variable;	
Executing	IP12: late, omission, valueSubtle IP13: late, omission, valueSubtle	OP13: late, omission, valueCoarse
	IP12.noFailure, IP13.noFailure → OP13.noFailure; IP12.late, IP13.noFailure → OP13.late; IP12.noFailure, IP13.late → OP13.late; IP12.late, IP13.late → OP13.late; IP12.valueSubtle, IP13.noFailure → OP13.valueCoarse; IP12.noFailure, IP13.valueSubtle → OP13.valueCoarse; IP12.valueSubtle, IP13.valueSubtlev → OP13.valueCoarse; IP12.wildcard, IP13.omission → OP13.omission; IP12.omission, IP13.wildcard → OP13.omission; IP12.late, IP13.valueSubtle → OP13.valueCoarse; IP12.valueSubtle, IP13.late → OP13.valueCoarse;	

Figure 8.14: Modeling failure behavior of components(Cont.)

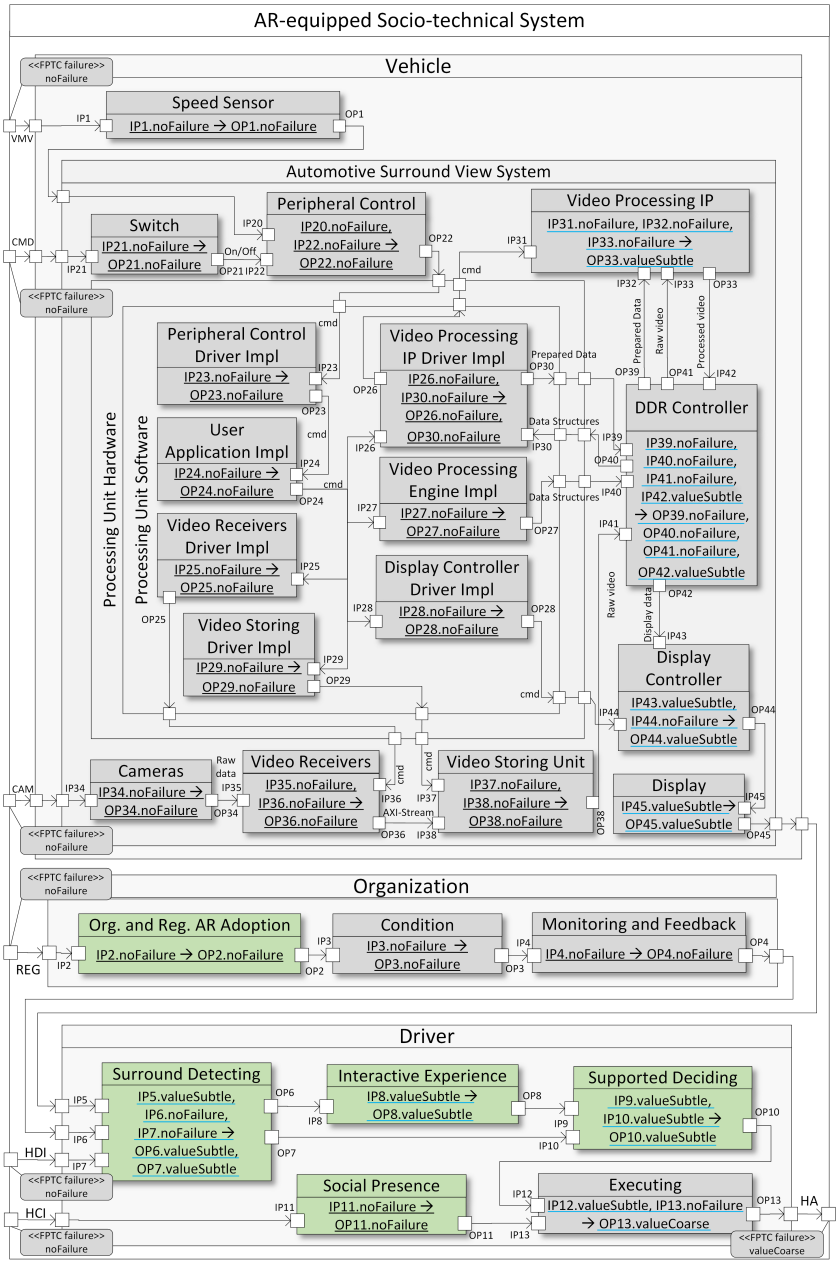


Figure 8.15: Analyzing AR-equipped socio-technical system (Scenario1)

The analysis results can be helpful in hazard identification and categorization. Since the reason for system failure is a technical component, functional safety is addressed by ISO 26262.

In this case, unintended displayed information is the identified hazard and the reason is failure in video processing IP. System failure in this scenario would lead to light accident and light injuries. The reason is that the speed is not usually high while parking the car. Based on the explanation in Subsection 8.2.5 and Figure 8.3, severity in this case is S1. Class of exposure is E4, because probability of exposure is more than 10%. It means that it is more than 10 percent probable that a driver be exposed to parking situation while driving a car. Finally, class of controllability is C1 or normally controllable. It means that more than 90% of drivers can control this situation. Therefore, ASIL level for this case is A, based on Figure 8.3. If we aim at overcoming risks with ASIL level A, then we should define safety goal, functional and technical safety requirements in order to overcome this risk. For example, for this scenario safety goal, functional safety requirement and technical safety requirement can be defined as follows to prevent failure in processing unit IP:

- **Safety goal**: The driver shall be notified, if there is failure in processing.
- **Safety requirement**:
  - **Functional safety requirement**: A monitoring component should be used to check the processing actively.
  - **Technical safety requirement**: Monitoring function should check the processing output every 10ms.

After interpreting the results and providing safety requirements, system design would be updated. Then, failure behavior can also be updated and failure propagation analysis can be repeated for another iteration.

Scenario 2:

- **Description of the scenario:** In this scenario, we assume that the technical part of the system works without failure, but driver doesn't have interactive experience. For example, it is the first time driver is working with systems containing AR and he/she can not understand the meaning of AR notations. Therefore, driver would decide and execute incorrectly.
- **Modeling failure behavior:** We show the failure propagation with underlined FPTC rules, which are the rules that are activated, shown in Figure 8.16.

Surround view sub-components behave as propagational and propagate no-Failure from inputs to outputs. *Interactive experience* behaves as source and while its input is noFailure, it has omission failure mode in its output. This activated rule is shown on this component.

- **Analysis of system behavior:** Omission failure mode in *interactive experience* transforms to valueSubtle in *supported deciding*, because lack of interactive experience causes wrong decision and in *executing*, it transforms to valueCoarse. Similar to the first scenario, the reason for this transformation is that if there is value failure mode in *executing* function, it can be detected by user, which means valueSubtle transforms to valueCoarse.
- **Interpreting the results:** Based on back propagation of the results, we can explain how the rules have been triggered. ValueCoarse on OP13 is because of valueSubtle on IP12. ValueSubtle on IP12 is because of valueSubtle on OP10 and we continue to IP8, which is related to *interactive experience* component.

  In this scenario, we considered failure in AR-related part of the system and since it refers to limitation in intended functionality (SOTIF related hazards), we do not determine ASIL level. If the expected severity and controllability of the scenario is higher than S0 and C0 respectively, we need to consider SOTIF safety process [38]. As we explained in the previous scenario, severity and controllability are higher than S0 and C0. Lack of interactive experience leads to system failure and incorrect deciding is the identified hazard. Safety goal and safety requirement can be defined as follows. Since the failure is not emanated from technical part of the system, we do not need to specify technical safety requirement:

  – **Safety goal**: Interactive experience shall be provided for the driver.
  – **Safety requirement**: The Company should provide a training video for all drivers at the first time of using the system.

  After applying the requirements the behavior of this component would change from source to other types and analysis can be repeated.

  It is not possible to detect risk originated from failure in interactive experience, without using the proposed representation constructs, because using these representation constructs or modeling elements provide the possibility to analyze their failure propagation and provides the possibility to analyze effect of these failures on system behavior. Then based on analysis results decision about design change or fault mitigation mechanisms would be taken.

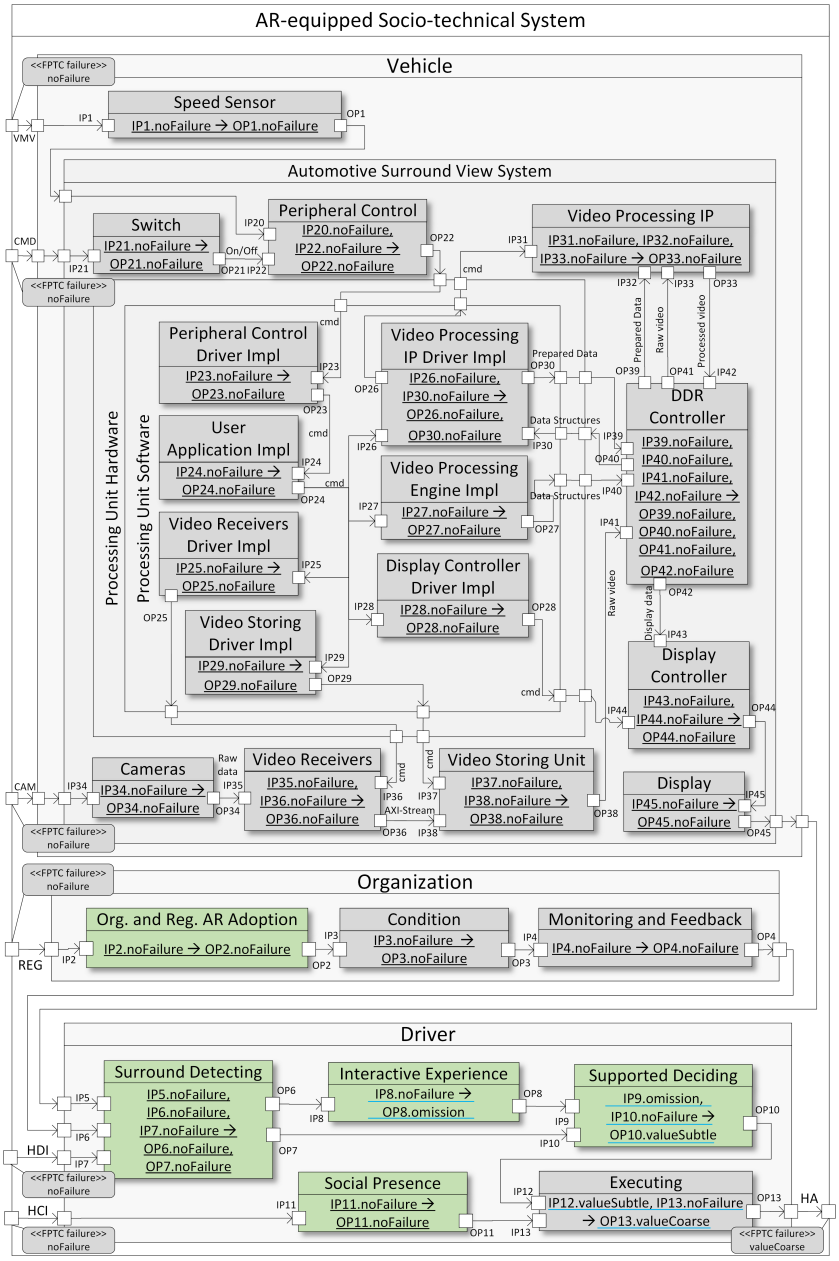


Figure 8.16: Analyzing AR-equipped socio-technical system (Scenario2)

Scenario 3:

- **Description of the scenario:** In this scenario, we assume that road transport organization has not updated rules and regulations based on AR technology, which is a limitation in intended functionality. For example, parking lot striping is not updated to be used by AR applications and it affects on road condition, but *monitoring and feedback* component detect this problem and provide a feedback to driver. This feedback would be a visual text alarm showing that there is a problem in AR information. Therefore, driver will not depend on shown result and try to decide and execute correctly.
- **Modeling failure behavior:** We show the failure propagation with underlined FPTC rules, which are the rules that are activated, shown in Figure 8.17. Similar to the previous scenario, surround view sub-components behave as propagational and propagate noFailure from inputs to output. *Organization and regulation AR adoption* behaves as source and while its input is noFailure, it has omission failure mode in its output. This activated rule is shown on this component. *Monitoring and feedback* component behaves as sink and while its input is omission, it has noFailure in its output.
- **Analysis of system behavior:** Omission failure mode propagates from *organization and regulation AR adoption* to *condition* and *monitoring and feedback*. In *monitoring and feedback* it will transform to noFailure. Then, noFailure is propagated from *surround detecting* to *interactive experience*, *supported deciding* and *executing*.
- **Interpreting the results:** In this scenario, system output is provided without failure. Thus, there is no hazard and no safety requirement is required.

8.4.6 Compliance with ISO 26262 and SOTIF

Based on the explanation in Subsection 8.2.5, the first step of ISO 26262 safety process is *item definition* and the first step of SOTIF safety process is *functional and system specification*. In Figure 8.10, we defined the components which are used for modeling items, their interactions and dependencies. We also specified system and functions through entities specification.

The second step of ISO 26262 is *hazard analysis and risk assessment* and second step of SOTIF is *SOTIF related hazard identification and risk evaluation*. Based on the interpreted results of each scenario, hazards are identified (if there are) and categorized based on ASIL level, if they are emanated from technical failures, otherwise they are evaluated qualitatively. If the hazard is

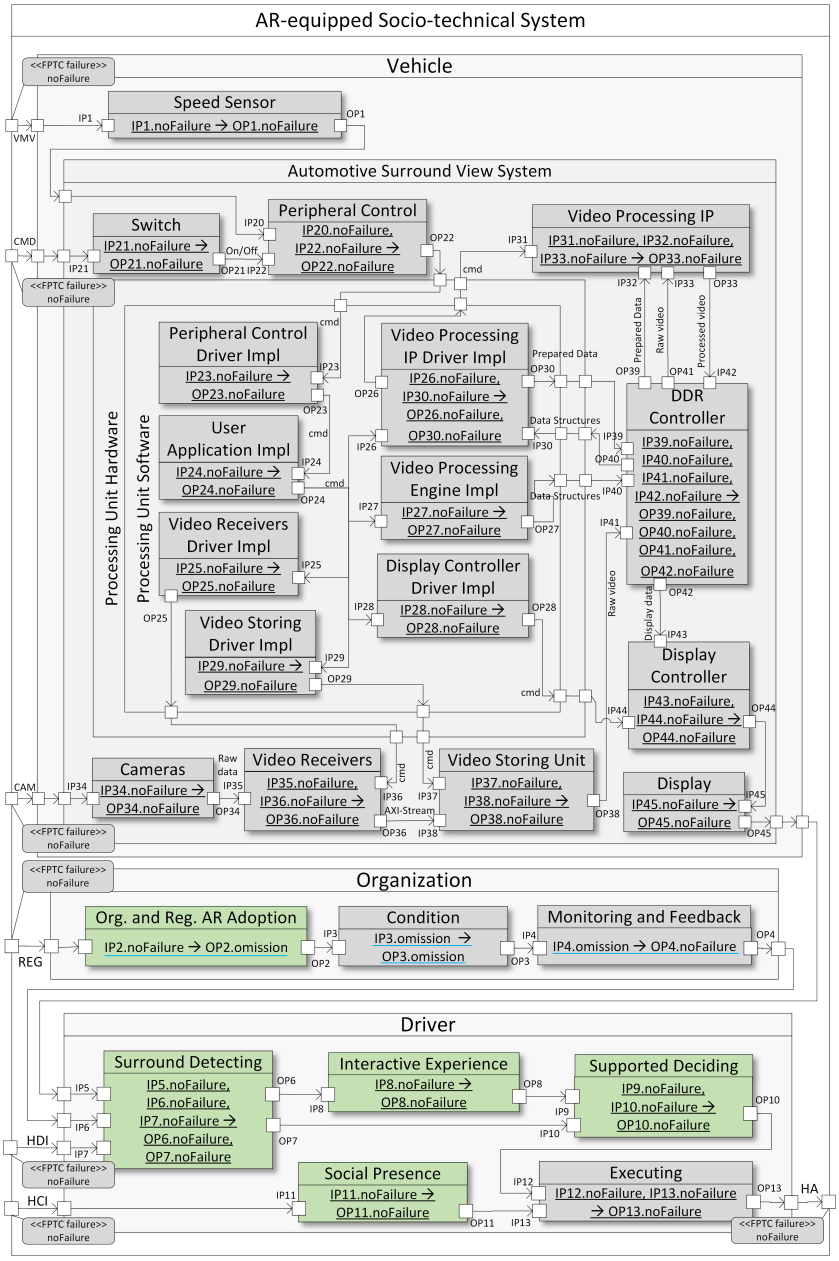


Figure 8.17: Analyzing AR-equipped socio-technical system (Scenario3)

emanated from a technical fault, functional safety is addressed by ISO 26262, otherwise it is addressed by SOTIF.

The third step of SOTIF is *identification and evaluation of triggering events*. Sub-components in Figure 8.10 are the identified potential triggering events and failure behavior of each of these sub-components in Figures 8.11-8.14 are evaluation of the triggering events.

The third and fourth steps of ISO 26262 are *functional and technical safety concept* and fourth step of SOTIF is *functional modification to reduce SOTIF risk*. The aim in the two steps of ISO 26262 is to define functional and technical safety requirements. Defining functional and technical safety requirements should be based on the analysis results as explained in the first scenario. Functional modification is also provided based on the analysis results as explained in the second scenario.

Finally, *verification test* of ISO 26262 and SOTIF includes considering several scenarios and verifying system functioning. This step is supported by the provided analysis results.

8.4.7 Lessons Learnt

In this subsection, we present the lessons learnt while conducting the case study. As it is shown in the second and third scenarios in Subsection 8.4.5, in an AR-equipped socio-technical system, there are system failures which are not caused by technical entities of the system and new AR-related dependability threats are the reason for these system failures. These new AR-caused dependability threats are related to intended functionality of socio entities of the system. Our proposed framework provides the required means to take into account these new AR-related dependability threats. We can consider these extensions from two perspectives:

- **Augmented reality concepts coverage**: from a coverage point of view, as shown in Subsection 8.4.4, modeling capabilities obtained by our proposed framework, allow architects and safety managers to model augmented reality effects on socio-technical systems by using modeling elements related to AR-extended human functions as well as modeling elements related to AR-caused faults leading to human failures. For example, in the second scenario, failure in *interactive experience* is considered as an AR-related dependability threat and its modeling element provides representation mean for taking into account AR effect as an AR-caused fault leading to human failures. In the third scenario, failure in updating rules and regulations based

on AR technology is considered as AR-related dependability threat and its modeling element provides representation mean for taking into account AR effects. It is also shown in Subsection 8.4.5 that analysis capabilities allow architects and safety managers to have at disposal means to reveal effect of AR-related dependability threats on system behavior. It is done by analyzing failure propagation that might be effective in emerging risks within an AR-equipped socio-technical system.

- **Expressiveness**: Expressiveness refers to the power of a modeling language to express or describe all things required for a given purpose [39]. Set of symbols or possible statements that can be described by modeling languages can be used for measuring expressiveness. Statement means "a syntactic expression and its meaning". As it is explained in Subsection 8.2.2, the extensions on human modeling elements used to extend the modeling language is based on an AR-extended human function taxonomy (AREXTax [7]). This taxonomy is obtained by extracting human functions from about six state-of-the-art human failure taxonomies (Norman [19], Reason [20], Rasmussen [21], HFACS [22], SERA [18], Driving [23]). This taxonomy is also extended based on various studies and experiments on augmented reality. In addition, the extension for extending organization modeling elements is based on a fault taxonomy (AREFTax [8]) containing AR-caused faults leading to human failures. This taxonomy is gained by harmonizing about five state-of-the-art fault taxonomies (Rasmussen [21], HFACS [22], SERA [18], Driving [23] and SPAR-H [27]). The taxonomy is also extended based on various studies and experiments on augmented reality. According to the basis of the extensions and as it is also shown in Subsection 8.4.4, the extensions increase power of modeling language to express new AR-caused risks.

We used Concerto-FLA analysis technique as the basis of the analysis in order to disclose the advantages of the proposed AR-related extensions included in our proposed framework at analysis level. Concerto-FLA uses FPTC syntax for the modeling failure behavior of each component or sub-component, which includes defining FPTC rules for a component/sub-component in isolation. It is possible to define FPTC rules for the AR-extended modeling elements characterizing their behavior. In addition, as known, modeling the failure behavior can be challenging, because the number of FPTC rules grows exponentially with the increase of the cardinality of the input ports. It is important to consider possible failure modes for each input in a component/sub-component and skip the others. It is not conspicuous in small and academic examples, but it is

really challenging when we use an industrial case study.

8.5 Threats to Validity

In this section, we discuss threats to validity in relation to our research based on best practices available in the scientific literature [36]. Validity of a study denotes to what extent the results can be trusted.

External validity refers to possibility of generalizing the findings. We provided a case study with three scenarios from automotive domain, but the proposed framework is not limited to specific scenarios and specific domain and the baseline for the included extensions, which are AREXTax and AREFTax taxonomies are attained from taxonomies in various domains. Thus, there is the possibility of generalizing the findings for automotive domain in general and also for other domains.

Construct validity refers to the quality of choices and measurements. In our case, we used SafeConcert, which is an accepted work, as the basis of our work. Proposed extensions are also based on state-of-the-art taxonomies (Norman [19], Reason [20], Rasmussen [21], HFACS [22], SERA [18], Driving [23] and SPAR-H [27] taxonomies) and studies and experiments for the new technologies. The modeling and analysis process is done based on standardized process to increase the repeatability of the work. However, it can not be guaranteed that different people have same answer using our proposed framework, because it depends on the analyzer skills and ability for modeling and analysis.

In this paper we used a realistic and sufficiently complex case at a level that can be found in industry to verify our proposed framework including AR-related extensions. Although we were not allowed to access confidential information related to their customers, we have been able to model system architecture and failure behavior of system components using SafeConcert metamodel, its AR-extensions and FPTC rules.

In this case study, we illustrated the modeling and analysis capabilities of our proposed framework including AR-related extensions through three different scenarios with different assumptions about the AR-related components' failure behavior. We have not shown that the modeling elements are complete for modeling all possible scenarios. Instead, we have focused on the provided elements to check if they are able to capture new system failure behaviors.

The benefit of using our proposed extensions for a particular case depends on the ability to choose the best elements and the ability to establish failure

behavior of the component related to that element. Still, this case provides evidence for the applicability and usefulness of our proposed framework. Further investigations are required to provide more beneficial results on limitations of modeling and analysis applications.

8.6 Discussion

Statistical information is used for determining exposure, severity and controllability of ASIL value of systems with SAE-levels 0-2. It would be possible to use the same statistical information for determining exposure and severity in AR-equipped systems with higher levels of automation, but controllability is a factor, which is affected by augmented reality used in higher levels of automation. Thus, it is required to model system and include effect of augmented reality on the model to be able to involve AR effect in specifying controllability factor of ISO 26262. For providing automated driving safety, Responsibility Sensitive Safety (RSS) standard [40] can be helpful. This standard provides formalization for safe decisions by self-driving cars in cases where machine learning mechanisms are used [41].

Surround view system can be mounted on vehicles with higher levels of automation (for example level 1-3) alongside more advanced systems for providing driver assistance functionalities. In these cases, driver is not supervising the car and controllability factor should be defined by modeling system as an AR-equipped socio-technical system. In [32], a controllability classification is proposed based on human takeover time and analysis of human driver models. The value of human action times, based on studies in literature, are used for predicting mean takeover times. Since classification of controllability according to ISO 26262 requires description of percentiles, normal distribution is assumed for each action time. Normal distribution can be obtained by mean value and its standard deviation. Based on the reaction times and distributions, it is possible to calculate controllability of the situation. The proposed modeling extensions included in our proposed framework provide the possibility to model effect of augmented reality on human and effect of augmented reality on influencing factors on human functions. Thus, mean takeover time and as a consequence controllability can be updated while using augmented reality by using the proposed extensions on humans and influencing factors modeling.

The generated model using our proposed framework and analysis results can be used to provide safety case for AR-equipped industrial products. Safety case contains arguments based on evidences to demonstrate that the system is

acceptably safe to work on a given environment. However, it is required to provide also some documentation explaining the results and showing how the safety requirements are achieved. Goal Structuring Notation (GSN) [42] can be used for SOTIF argumentation [43].

Extended human modeling elements can be used for modeling integration of human aspects with interactive systems in system testing. For example, MIODMIT architecture [44] is a generic architecture for interactive systems. As it is discussed in [45], human aspects should be considered and integrated while testing. Using extended modeling elements for modeling different aspects of human as a user of an interactive system would be of value for the system testing.

8.7 Related Work

A comparative study about architecture-based risk analysis techniques is provided in [46]. Specifically, in this work, authors compare: the modeling capabilities, process and tool support of various techniques. Traditional methods such as Fault Tree Analysis (FTA) [47] and Failure Modes and Effects Analysis (FMEA) [48] are manual analysis techniques. In comparison, there are also model-driven techniques, which provide the analysis results (semi-)automatically based on the system architecture and annotated failure behavior information. Model-driven techniques such as Failure Propagation and Transformation Notation (FPTN) [49], FPTC, Hierarchically Performed Hazard Origin and Propagation Studies (HIP HOPS) [50] and techniques using Architecture Analysis and Description Language (AADL) and its technical error annex [51] are considered in this study. All these techniques consider risks emanated from technical parts. Human and organization are not considered as part of the system that would introduce risk.

A framework for construction safety management and visualization system (SMVS) is proposed in [52]. This framework includes a safety management process, which includes planning, education and inspection phases. A prototype system is also developed and tested. The results shown that this framework improves risk identification and communication between managers and workers in construction sites. Augmented reality is used for improving the safety management process. In comparison to our work, in this paper the proposed framework is specific to construction domain. AR is also used for safety management process improvement, but it is not considered as part of the system, which is going to be evaluated. Thus, risks emanated from AR and

AR-related factors are not included in the process.

In risk analysis techniques for socio-technical systems, failures emanated from human and organizational factors are also considered in addition to technical failures. Human failure taxonomies provide the possible human failures while working in a socio-technical system. There are also taxonomies on organizational factors that provide the factors influencing human performance. In [13], Concerto-FLA analysis technique is proposed based on SERA taxonomy including human failures and organizational factors. Human reliability quantification techniques can be used for quantifying human error probability and providing quantitative risk assessment. Expert judgment and analysis of accident reports can be used for determining likelihood. However, error likelihood estimation usually has low accuracy. We also do not aim at using quantitative assessment, because based on SOTIF standard, SOTIF related hazards require qualitative analysis.

A risk analysis technique for systems containing augmented reality, named Safe-AR, is proposed in [53]. Safe-AR integrates failures of AR/user interface at three levels: perception, comprehension and decision-making. Likely risks and their severity are based on reports available in literature. The proposed technique is shown on an AR left-turn assist app, which is an example from automotive domain. Human functions and failure modes in this study are limited to the provided example and a generalization is required to be used for other domains and more complicated case studies.

A framework for risk management in financial services is provided in [54]. The paper focuses on risk management from a human centered perspective. In comparison to our work, this paper is specific to financial domain and it does not provide a general framework. The proposed framework does not include modeling and analysis constructs to be used for risk assessment. The required activities in different steps for assessing risk are not defined specifically.

Human Functions in Safety (HFiS) framework is proposed in [55]. This framework focuses on the role of human in system safety in socio-technical systems. Organizational factors are also considered in this framework. The output from applying the framework is a description of safety related activities through human functions, organizational goals and contextual factors. It is developed for railway context, but there are guidance for generic application of HFiS. In comparison to our work, in this paper there is no consideration on effects of new technologies such as augmented reality on human functions and organizational factors.

In comparison to the above-mentioned works, our framework provides more general risk assessment technique with the integration of risk emanated from

human, organization and technology (augmented reality). In addition, effects of augmented reality on human functions and organizational factors are considered in our framework. We highlight the features provided by our framework and pre-existing related work in Figure 8.18.

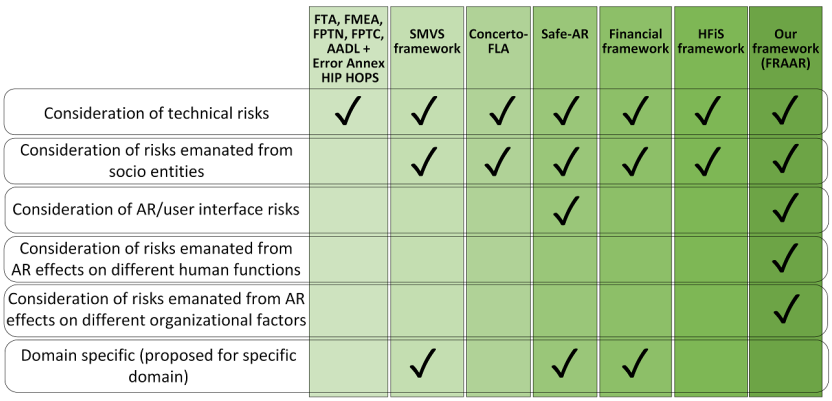


Figure 8.18: Comparative analysis summary

8.8 Conclusion and Future Work

In this paper, we provided a framework for assessing risk of AR-equipped socio-technical systems. This framework provides the possibility to detect faults and failures leading to system risk and provides the possibility to model and analyze system behavior. In addition, we conducted a case study to illustrate how our proposed framework can be used for predicting risk caused by new AR-related dependability threats. The predicted risk can then be used as a basis for developing e.g., the safety concept in compliance with ISO 26262 and SOTIF related work products.

The framework includes extensions for modeling and analyzing AR effects on human functioning and AR effects on faults leading to human failures. We showed the analysis results by providing three scenarios. In two of the scenarios, the failure was emanated from the AR-related faults. We provided failure propagation manually and we showed that in some scenarios there would be no failure in technical entities of the system, but risk would be identified caused by non-technical AR-related faults. By implementing our proposed conceptual

extensions for CHESS toolset, failure propagation calculation can be provided automatically to be used for AR-equipped socio-technical systems.

Our proposed framework supports ISO 26262 and SOTIF development process activities and can be used for providing expected work products by these safety standards. In addition, we discussed that the modeling capabilities within our proposed framework is helpful for determining ISO 26262 controllability. ISO 26262 controllability requires to be updated in order to be used for AR-equipped socio-technical systems, especially in higher levels of driving automation.

Further research is required to show the potential benefits of the proposed framework. Specifically, we intend to conduct case studies where there are scenarios with higher safety-criticality. In addition, having two or more teams composed of three or four experienced analysts would help to have more advanced scenarios including more complicated propagation of failures. In future, we also plan to evaluate a safety-critical socio-technical system within the rail industry, the passing of a stop signal (signal passed at danger; SPAD) [56], to verify if the results are transferable to the rail domain.

Bibliography

[1] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators & Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.

[2] "ImmerSAFE - Immersive Visual Technologies for Safety-critical Applications," 2021. https://immersafe-itn.eu/.

[3] B. F. Goldiez, N. Saptoka, and P. Aedunuthula, "Human performance assessments when using augmented reality for navigation," tech. rep., University of Central Florida Orlando Inst for Simulation and Training, 2006.

[4] D. Van Krevelen and R. Poelman, "A survey of augmented reality technologies, applications and limitations," *The International Journal of Virtual Reality (IJVR)*, vol. 9, no. 2, pp. 1–20, 2010.

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

[6] International Organization for Standardization (ISO), " ISO/PAS 21448: Road vehicles — Safety of the intended functionality (SOTIF)," 2019. https://www.iso.org/standard/70939.html.

[7] S. Sheikh Bahaei and B. Gallina, "Augmented reality-extended humans: towards a taxonomy of failures – focus on visual technologies," in *European Safety and Reliability Conference (ESREL)*, Research Publishing, Singapore, 2019.

[8] S. Sheikh Bahaei, B. Gallina, K. Laumann, and M. Rasmussen Skogstad, "Effect of augmented reality on faults leading to human failures in socio-technical systems," in *International Conference on System Reliability and Safety (ICSRS)*, IEEE, 2019.

[9] S. Sheikh Bahaei and B. Gallina, "Towards Assessing Risk of Reality Augmented Safety-critical Socio-technical Systems." Published as proceedings annex on the International Symposium on Model-Based Safety and Assessment (IMBSA) website http://easyconferences.eu/imbsa2019/proceedings-annex/, 2019.

[10] S. Sheikh Bahaei and B. Gallina, "Extending safeconcert for modelling augmented reality-equipped socio-technical systems," in *International Conference on System Reliability and Safety (ICSRS)*, IEEE, 2019.

[11] L. Montecchi and B. Gallina, "SafeConcert: A metamodel for a concerted safety modeling of socio-technical systems," in *International Symposium on Model-Based Safety and Assessment (IMBSA)*, pp. 129–144, Springer, 2017.

[12] S. Mazzini, J. M. Favaro, S. Puri, and L. Baracchi, "CHESS: an Open Source Methodology and Toolset for the Development of Critical Systems.," in *Join Proceedings of EduSymp and OSS4MDE*, pp. 59–66, 2016.

[13] B. Gallina, E. Sefer, and A. Refsdal, "Towards safety risk assessment of socio-technical systems via failure logic analysis," in *International Symposium on Software Reliability Engineering Workshops (ISSRE)*, pp. 287–292, IEEE, 2014.

[14] M. Wallace, "Modular architectural representation and analysis of fault propagation and transformation," *Electronic Notes in Theoretical Computer Science*, vol. 141, no. 3, pp. 53–71, 2005. Proceedings of the Second International Workshop on Formal Foundations of Embedded Software and Component-based Software Architectures (FESCA).

[15] A. Ruiz, A. Melzi, and T. Kelly, "Systematic application of ISO 26262 on a SEooC: support by applying a systematic reuse approach," in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 393–396, IEEE, 2015.

[16] L. P. Bressan, A. L. de Oliveira, L. Montecchi, and B. Gallina, "A Systematic Process for Applying the CHESS Methodology in the Creation of Certifiable Evidence," in *14th European Dependable Computing Conference (EDCC)*, pp. 49–56, IEEE, 2018.

[17] "CONCERTO D2.7 – Analysis and back-propagation of properties for multicore systems – Final Version," 2016. http://www.concerto-project.org/results.

[18] K. C. Hendy, "A tool for human factors accident investigation, classification and risk management," tech. rep., Defence Research And Development Toronto (Canada), 2003.

[19] D. A. Norman, "Errors in human performance," tech. rep., California Univ San Diego LA JOLLA Center For Human Information Processing, 1980.

[20] J. Reason, *The human contribution: unsafe acts, accidents and heroic recoveries*. CRC Press, 2017.

[21] J. Rasmussen, "Human errors. a taxonomy for describing human malfunction in industrial installations," *Journal of occupational accidents*, vol. 4, no. 2-4, pp. 311–333, 1982.

[22] S. A. Shappell and D. A. Wiegmann, "The human factors analysis and classification system–HFACS," tech. rep., Civil Aeromedical Institute, 2000.

[23] N. A. Stanton and P. M. Salmon, "Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems," *Safety Science*, vol. 47, no. 2, pp. 227–237, 2009.

[24] W.-T. Fu, J. Gasper, and S.-W. Kim, "Effects of an in-car augmented reality system on improving safety of younger and older drivers," in *International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 59–66, IEEE, 2013.

[25] M. C. Schall Jr, M. L. Rusch, J. D. Lee, J. D. Dawson, G. Thomas, N. Aksan, and M. Rizzo, "Augmented reality cues and elderly driver hazard perception," *Human factors*, vol. 55, no. 3, pp. 643–658, 2013.

[26] S. Sheikh Bahaei and B. Gallina, "A metamodel extension to capture post normal accidents in ar-equipped socio-technical systems," in *European Safety and Reliability Conference (ESREL)*, Research Publishing, Singapore, 2021.

[27] D. Gertman, H. Blackman, J. Marble, J. Byers, C. Smith, *et al.*, "The SPAR-H human reliability analysis method," *US Nuclear Regulatory Commission*, vol. 230, 2005.

[28] J.-C. Le Coze, *Post Normal Accident: Revisiting Perrow's Classic*. CRC Press, 2020.

[29] J. Noll and S. Beecham, "Measuring global distance: A survey of distance factors and interventions," in *International Conference on Software Process Improvement and Capability Determination*, pp. 227–240, Springer, 2016.

[30] M. R. Miller, H. Jun, F. Herrera, J. Y. Villa, G. Welch, and J. N. Bailenson, "Social interaction in augmented reality," *PloS one*, vol. 14, no. 5, p. e0216290, 2019.

[31] I. Šljivo, B. Gallina, J. Carlson, H. Hansson, and S. Puri, "A method to generate reusable safety case argument-fragments from compositional safety analysis," *Journal of Systems and Software*, vol. 131, pp. 570–590, 2017.

[32] T. Hecht, M. Lienkamp, C. Wang, *et al.*, "Development of a human driver model during highly automated driving for the ASIL controllability classification," in *8. Tagung Fahrerassistenz*, 2017.

[33] I. Sljivo, B. Gallina, J. Carlson, H. Hansson, *et al.*, "Using safety contracts to guide the integration of reusable safety elements within iso 26262," in *21st Pacific Rim International Symposium on Dependable Computing (PRDC)*, pp. 129–138, IEEE, 2015.

[34] "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," 2021. https://www.sae.org/standards/content/j3016_202104.

[35] G. Dimitrakopoulos, L. Uden, and I. Varlamis, *The Future of Intelligent Transport Systems*. Elsevier, 2020.

[36] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," *Empirical software engineering*, vol. 14, no. 2, p. 131, 2009.

[37] F. Ye and T. Kelly, "Component failure mitigation according to failure type," in *Proceedings of the 28th Annual International Computer Software and Applications Conference, 2004. COMPSAC 2004.*, pp. 258–264, IEEE, 2004.

[38] C. Becker, J. C. Brewer, and L. Yount, "Safety of the intended functionality of lane-centering and lane-changing maneuvers of a generic level 3 highway chauffeur system," tech. rep., United States. National Highway Traffic Safety Administration, 2020.

[39] S. Patig, "Measuring expressiveness in conceptual modeling," in *International Conference on Advanced Information Systems Engineering (CAiSE)*, pp. 127–141, Springer, 2004.

[40] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "On a formal model of safe and scalable self-driving cars," 2017. https://arxiv.org/pdf/1708.06374.pdf.

[41] Y. Zhang, G. Lintern, L. Gao, and Z. Zhang, "A study on functional safety, sotif and rss from the perspective of human-automation interaction," tech. rep., SAE Technical Paper, 2021.

[42] "Goal Structuring Notation Community Standard," 2018. https://scsc.uk/SCSC-141B.

[43] International Organization for Standardization (ISO), " ISO/DIS 21448: Road vehicles — Safety of the intended functionality (SOTIF)," 2021. https://www.iso.org/standard/77490.html.

[44] G. Cockton and A. Woolrych, "Understanding inspection methods: lessons from an assessment of heuristic evaluation," in *People and computers XV—Interaction without frontiers*, pp. 171–191, Springer, 2001.

[45] A. Canny, E. Bouzekri, C. Martinie, and P. Palanque, "Rationalizing the need of architecture-driven testing of interactive systems," in *International Conference on Human-Centred Software Engineering (HCSE)*, pp. 164–186, Springer, 2018.

[46] L. Grunske and J. Han, "A comparative study into architecture-based safety evaluation methodologies using aadl's error annex and failure propagation models," in *High Assurance Systems Engineering Symposium (HASE)*, pp. 283–292, IEEE, 2008.

[47] D. F. Haasl, N. H. Roberts, W. E. Vesely, and F. F. Goldberg, "Fault tree handbook," tech. rep., Nuclear Regulatory Commission, 1981.

[48] D. H. Stamatis, *Failure mode and effect analysis: FMEA from theory to execution*. Quality Press, 2003.

[49] P. Fenelon and J. A. McDermid, "New directions in software safety: Causal modelling as an aid to integration," in *Workshop on Safety Case Construction, York (March 1994)*, Citeseer, 1992.

[50] Y. Papadopoulos and J. McDermid, *Safety-directed system monitoring using safety cases*. PhD thesis, Citeseer, 2000.

[51] P. Feiler and A. Rugina, "Dependability modeling with the architecture analysis & design language (aadl)," tech. rep., Carnegie-Mellon Univ Pittsburgh PA Software Engineering INST, 2007.

[52] C.-S. Park and H.-J. Kim, "A framework for construction safety management and visualization system," *Automation in Construction*, vol. 33, pp. 95–103, 2013.

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

[54] J. Organ and L. Stapleton, "A socio-technical systems framework for risk management in financial services: Some empirical evidence from a case study of the irish banking crisis," *IFAC-PapersOnLine*, vol. 52, no. 25, pp. 148–153, 2019.

[55] B. Ryan, D. Golightly, L. Pickup, S. Reinartz, S. Atkinson, and N. Dadashi, "Human functions in safety-developing a framework of goals, human functions and safety relevant activities for railway socio-technical systems," *Safety Science*, vol. 140, p. 105279, 2021.

[56] A. Naweed, J. Trigg, S. Cloete, P. Allan, and T. Bentley, "Throwing good money after SPAD? Exploring the cost of signal passed at danger (SPAD) incidents to australasian rail organisations," *Safety science*, vol. 109, pp. 157–164, 2018.

Chapter 9

Paper C: Towards Qualitative and Quantitative Dependability Analysis for AR- equipped Socio-technical Systems

Soheila Sheikh Bahaei and Barbara Gallina.
In Proceedings of the 5th International Conference on System Reliability and Safety (ICSRS-2021), IEEE, November 2021.

Abstract

Augmented Reality technologies are becoming essential components in various socio-technical systems. New kinds of risks, however, may emerge if the concertation between AR, other technical components and socio-components is not properly designed. To do that, it is necessary to extend techniques for risk assessment to capture such new risks. This may require the extension of modelling languages and analysis techniques. In the literature, modeling languages have been already extended by including specific language constructs for socio aspects in relation to the AR-impact. No satisfying contribution is available regarding analysis techniques. Hence, to contribute to filling the gap, in this paper, we propose an extension of previously existing analysis techniques. Specifically, we build on top of the synergy of qualitative and quantitative dependability analysis techniques and we extend it with the capability of benefiting from AR-related modelled aspects. In addition, we apply our proposed extension to an illustrative example. Finally, we provide discussion and sketch future work.

9.1 Introduction

Analyzing Augmented Reality (AR)-equipped socio-technical systems requires contemplating effect of various AR-related aspects on system behavior. Socio-technical systems are systems including socio entities such as human and organization and technical entities [1] such as augmented reality. Augmented reality technology is a technology that superimposes virtual content on the real environment of the user [2]. Augmented reality affects on human performance and it also affects on influencing factors on human performance. In order to automatize system analysis, model-based techniques [3] are used, which contain modeling system architecture and system behavior.

In [4], an extension to the Architecture Analysis and Design Language (AADL) [5] is proposed to enable modeling various types of faults and intertwining them into the system model to be used for analysis. It contains language extensions for modeling technical faults. In [6], a conceptual framework, called WAx (Work-As-x) is presented for the analysis of cyber-socio-technical systems. It contains concepts, and a language to develop information-driven model for understanding system functioning. It encompasses effect of digitalization on systems and organizations.

In [7], human entity is modeled by characterizing its behavior through human functions. In this study, effect of augmented reality on human functions are also considered. Human functions are provided based on state-of-the-art human failure taxonomies and they are extended by AR-extended human functions based on studies and experiments on AR. In [8], effect of augmented reality on influencing factors on human functions are considered and a taxonomy of faults leading to human failures based on state-of-the-art taxonomies are proposed. Then, this taxonomy is extended by considering AR-related factors causing human failures based on studies and experiments on augmented reality. SafeConcert metamodel [9], which is a metamodel for modeling socio-technical systems is extended in [10] to enable modeling of AR-equipped socio-technical systems. In [11], new concepts are proposed for modeling effect of digitalization, globalization and networked structure of organizations while performing risk assessment in AR-equipped socio-technical systems.

Currently, there is no analysis technique considering AR-related risks to be used for analyzing AR-equipped socio-technical systems' behavior. In this paper, we aim at proposing an extension for a synergy of qualitative and quantitative dependability analysis technique. The extension is based on AR-related modeling extensions and Concerto-FLA analysis plugin [1], which is Eclipse

plugin for dependability analysis and risk assessment implemented in CHESS project [12]. We use Concerto-FLA analysis technique, because it includes constructs for including socio-technical systems' concepts. We also use AR-related modeling extensions to include AR-related concepts. Finally, we use a monitoring system case study to illustrate our contributions.

The rest of the paper is organized as follows. In section 9.2, we provide essential background information. In Section 9.3, we propose our extension on synergy of qualitative and quantitative dependability analysis represented as an extended process. In Section 9.4, we present the extension on a monitoring system case study. In Section 9.5, we provide a discussion about the contribution of our proposed extension. In Section 9.6, we provide related work. Finally, in Section 9.7, we present some concluding remarks and describe the future work.

9.2 Background

This section provides essential background information onto which our work is based. First, the metamodel extensions for modeling AR-equipped socio-technical systems are recalled. Then, toolchain for automated dependability evaluation and a synergy of qualitative and quantitative dependability analysis techniques are recalled. Finally, analyzing socio-technical systems is explained.

9.2.1 Metamodel extensions for AR-equipped Socio-technical Systems

To capture AR-equipped socio-technical systems, constructs for modelling socio and technical (including AR-specific aspects) entities are needed. In [1], new constructs are proposed for modeling human and organization and their related aspects. In [10], AR-related concepts in addition to various socio concepts are considered and modeling elements related to these concepts categorized into two types are proposed. First category is human modeling elements for characterizing human functions (including AR-extended human functions) and human internal states (including AR-related human internal states). Second category is organization modeling elements characterizing external influencing factors on human performance (including AR-related factors).

For example, modeling elements of paying attention, deciding, executing and etc. are used for characterizing human functions. Modeling elements of

human physical state, mental state, experience and etc. are used for characterizing human internal states. Modeling elements of environmental condition, time pressure, supervision and etc. are used for characterizing external influencing factors. Modeling element of surround detecting is an AR-extended modeling element, which characterize AR-extended human function. The reason is that using AR technology would help human to detect surrounding environment, thus augmenting the human to an extended human.

9.2.2 Toolchain for Automated Dependability Evaluation

A toolchain is introduced in [13], to perform the dependability analysis automatically. The toolchain contains five metamodels and four model-transformation algorithms. The relationship between five models (m1...m5) conforming to these five metamodels and four model-transformations (t1...t4) are shown in Figure 9.1.

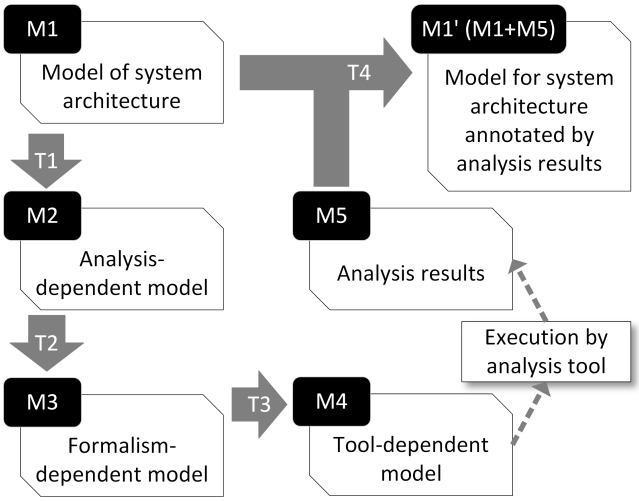


Figure 9.1: Relationship between models and transformations adapted from [13]

- **Metamodel 1:** This metamodel contains constructs to model various concepts of system architecture. The extended metamodel explained in Subsection 9.2.1 provides the constructs for preparing a model of system architecture at this level.

- **Metamodel 2:** This metamodel contains constructs for performing the intended analysis. The model prepared at this level is analysis-dependent. For example, in order to perform performance analysis, only information related to performance are extracted from the system architecture and other details are not considered.

- **Metamodel 3:** This metamodel contains constructs for implementing the analysis model in a specific formalism. The model prepared at this level is formalism-dependent. For example, Stochastic Petri Nets (SPNs) [14] or Fault Tree [15] can be considered as formalisms for the analysis.

- **Metamodel 4:** This metamodel contains constructs for preparing the code of the implementation by a specific tool. The model prepared at this level is tool-dependent. For example, a file including header, variable definitions, etc. that can be provided as input of a tool is a model at this level.

- **Metamodel 5:** This metamodel contains constructs for describing the results provided by the analysis tool. For example, a text file conforming to standard interchange formats such as XML can be considered as a model at this level.

- **Model-transformation 1:** This transformation extracts the information required for the intended analysis from the mass of information representing the system architecture. It is applied to m1 to produce m2.

- **Model-transformation 2:** This transformation implements the analysis algorithm using the intended formalism. It is applied to m2 to produce m3.

- **Model-transformation 3:** This transformation provides the implemented code to be used as the input of the analysis tool. It is applied to m3 to produce m4.

- **Model-transformation 4:** This transformation propagates the analysis results back into the system architecture. It uses m5 and m1 to produce a modified version of m1, which contains analysis results in addition to system architecture.

This toolchain presents how the dependability analysis can be implemented to perform the analysis automatically.

9.2.3 Synergy of Qualitative and Quantitative Dependability Analysis Techniques

A synergy of qualitative and quantitative dependability analysis techniques is proposed in [16]. It contains State-based analysis and Failure Logic Analysis (FLA). State-based analysis technique [17] is a quantitative technique, which is implemented based on the toolchain explained in Subsection 9.2.2. FLA is a qualitative analysis based on qualitative behavior of components and their causes.

It is required to have information or assumptions about the system architecture to be used for modeling system architecture. Formalism used in state-based analysis is Stochastic Petri Nets (SPNs) [14] with general probability distributions. There are three types of behavior modeling used in these two analysis techniques, which are simple stochastic behavior, error model and Failure Logic Analysis (FLA) [16]. These three types of behavior modeling are described in the following paragraphs.

Simple stochastic behavior uses probability distribution for specifying the time to the occurrence of a failure and the time required to fix the component after failure occurrence, if available. Possible failure modes and their probabilities also can be provided. As it is shown in Figure 9.2, exponential distribution with rate of 1.0e-6 per hour of operation is used for illustrating time to failure of this hardware component. Possible failure modes in case of failure in the output and their probabilities are shown in this example, which are omission (means output is not provided when expected) with probability of 80% and valueSubtle (means output is not in the expected range and it is not detected by user) with probability of 20%.

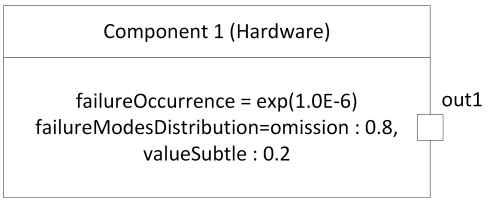


Figure 9.2: Modeling a hardware component with stochastic behavior [16]

Error model is defined by using a set of finite state machines modeling internal faults, external faults and their probabilities. It also models transitions between states. Error models are used when there are detail information about the component's failure behavior [16]. For example in Figure 9.3, a software

is modeled by two error models modeling internal fault occurrence and effect of external faults. In the top part of the picture, probability of occurrence of internal fault is defined as exp(6.0E-6) and it would propagate to an undetected error state leading to output failure mode omission with weight 0.8 or it would propagate to an error state incorrect value with weight 0.2. In the bottom part of the picture, omission external fault is considered propagated to undetected error state leading to omission failure mode in the output.

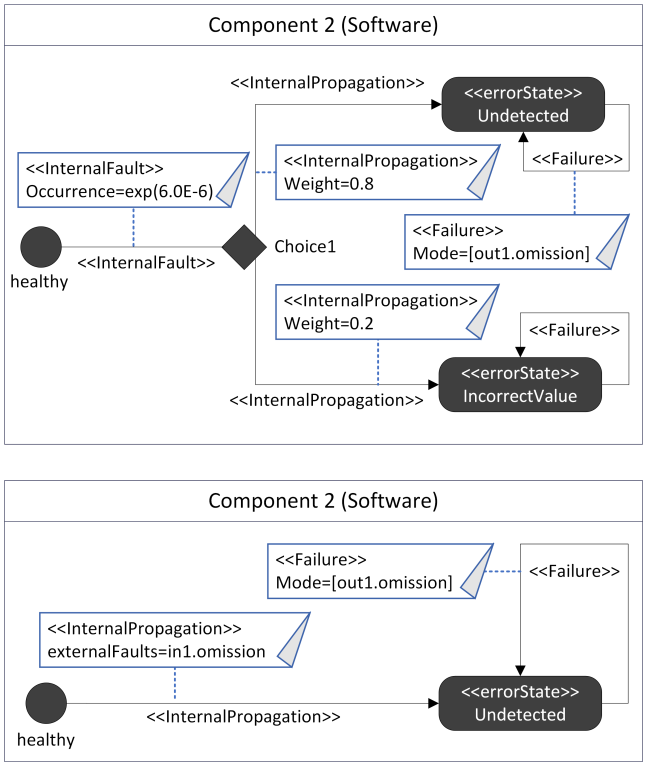


Figure 9.3: Modeling a software component with error models [16]

FLA behavior is defined by assigning possible failure modes in the input to possible failure modes in the output. In this type of behavior modeling probabilities are not considered. For example, in Figure 9.4, a software is modeled by defining FLA behavior. In this example, there are two inputs (In1, In2) and two outputs (out1, out2) for the software component. NoFailure (normal be-

havior) at input In1 and valueSubtle at input In2 will lead to valueSubtle at out1 and noFailure at out2. Relationship of other possible failure modes at inputs and outputs are defined similarly.

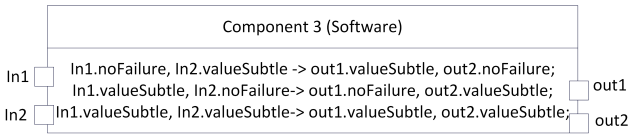


Figure 9.4: Modeling a software component with FLABehavior [16]

The following metrics can be measured by the quantitative analysis:

- Reliability: the probability that the system continuously remains in proper state from the time 0 up to time t.
- Availability:
  - Immediate: the probability that the system is in proper state at time t.
  - In a time interval: the fraction of time that system is in proper state in a given time interval.
- Probability of Failure on Demand (PFD): the probability that the system fails to provide a requested service. It can be obtained by calculating 1 minus immediate reliability.

9.2.4 Analyzing Socio-technical systems

Concerto-FLA analysis technique [1] is an extension of FLA qualitative technique. The extension includes capabilities for analyzing socio aspects. It is implemented as plug-in within CHESS toolset [18] for developing high integrity socio-technical systems. Model of the system architecture is used for running the analysis and results are back propagated in order to support an iterative and incremental system development [13]. Formalism used in Concerto-FLA is based on fixed-point calculation used in FPTC technique [19]. In Concerto-FLA analysis technique [1], FPTC rules are used.

FPTC rules are expressions for illustrating components' behavior by relating input failure modes to output failure modes. Failure modes include early (provided function early), late (provided function late), commission (provided function at a time which is not expected), omission (not provided function at a

time which is expected), valueSubtle (provided wrong value after computation that user can not detect it) and valueCoarse (provided wrong value after computation that user can detect it). Components' behavior can be classified as the following categories:

- Sink: when component detects failure in the input and corrects it in the output.
- Propagational: when component propagates the same failure mode or normal behavior in the input to the output.
- Transformational: when component transforms the failure mode in the input to another failure mode in the output.

FPTC syntax for modeling failure behavior at component and connector level is as follows:

behavior = expression+
expression = LHS '→' RHS
LHS = portname'.' bL | portname
'.' bL (',' portname '.' bL) +
RHS = portname'.' bR | portname
'.' bR (',' portname '.' bR) +
failure = 'early' | 'late' | 'commission' | 'omission' |
'valueSubtle' | 'valueCoarse'
bL = 'wildcard' | bR
bR = 'noFailure' | failure

NoFailure shows normal behavior. Wildcard on a specific input shows that the output is provided regardless of the failure mode or normal behavior of this specific input. For example, IP1.wildcard → OP1.noFailure is an example of a FPTC rule which shows that regardless of the failure mode or normal behavior on the input port with the name IP1 the output on the port OP1 will be provided with normal behavior. This shows the behavior of a component with sink behavior.

9.3 Proposed Analysis Process

In this section, we propose an extension based on AR-related modeling extensions and Concerto-FLA analysis technique [1]. We build on top of the

synergy of qualitative and quantitative analysis in [16]. We aim at extending this synergy by incorporating socio-related and AR-related aspects explained in Subsection 9.2.1. Our proposed analysis process is illustrated in Figure 9.5.

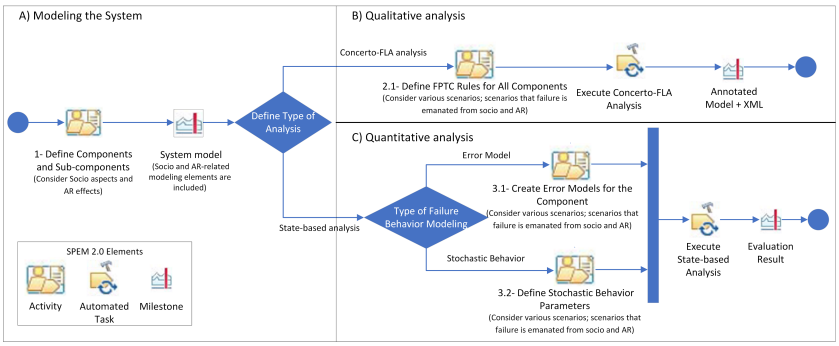


Figure 9.5: The proposed extended analysis process

The added value with respect to the synergy of quantitative and qualitative analysis is the possibility of analyzing various socio and AR-related aspects and their effects on system behavior. AR-related metamodel extensions are used in the system modeling by including AR-related modeling elements in the system model. In case of using qualitative analysis, Concerto-FLA analysis can be used for defining FPTC rules for AR-related components and automated analysis is used for obtaining the annotated model by analysis results. In case of quantitative analysis, error model or stochastic behavior are used for analyzing system behavior including AR-related effects.

Part A of Figure 9.5 contains the activity that should be done for preparing the system model. This activity is defining components and sub-components. Then, we need to decide about analysis type. If we need to do qualitative analysis, we perform the next activities based on Concerto-FLA analysis technique (Part B), otherwise we perform based on State-based analysis technique (Part C).

Based on Part B of the figure, FPTC rules should be defined for all components. Then, Concerto-FLA analysis will be executed and model annotated by analysis results will be provided.

Based on Part C of the figure, failure behavior modeling type should be defined. If we want to use error model, then we need to create the error model of the desired component. If we want to use stochastic behavior, then we should define the related parameters. Next step is to execute the state-based analysis

and to measure the evaluation result.

Result of the analysis can be used for hazard identification, defining safety goals and safety requirements.

We explain the activities of all the steps in the following subsections and in Table 9.1, we compare these steps of our proposed extended process with the previous process in [16].

9.3.1 Define Components and Sub-components

Main entities incorporating in a system are considered as the main components. It is important to consider socio entities, which are human and organization. Defining sub-components are based on important aspects of each entity. In technical components, important aspects are defined based on technical description of the system. Human important aspects are defined based on human functions and human internal states. Organization important aspects are defined based on organizational important aspects. Human and organization modeling elements introduced in the extended metamodel explained in Subsection 9.2.1 are the modeling constructs that can be used for defining human and organization sub-components. For example, condition, environment and any other influencing factor on human performance can be considered as organizational important aspects. The extensions include AR-related aspects, which should be considered in defining sub-components.

9.3.2 Define FPTC rules for All Components

This activity should be done based on the syntax explained in Subsection 9.2.4. In order to define FPTC rules, each component/sub-component should be analyzed individually. We should define the possible failure modes at each of their inputs and outputs for various scenarios. Then, FPTC rules can be used for relating the failure modes at inputs to the failure modes at outputs. For example, a camera would not receive the input (raw image) because of the obstacle in front of it. Input failure mode in this example is omission as explained in Subsection 9.2.4. Based on technical analysis of the camera, we would model it as propagational (explained in Subsection 9.2.4). It means that the failure mode in input propagates to the output port and it does not provide the output.

9.3.3 Create Error Models for the Component

This activity should be done based on the syntax explained in Subsection 9.2.3. In order to define error models, the intended component/sub-component should be analyzed individually. State machine for each component including internal and external faults and their probabilities should be defined for various scenarios.

9.3.4 Define Stochastic Behavior Parameters

This activity should be done based on the syntax explained in Subsection 9.2.3. In order to define stochastic behavior parameters, the intended component/sub-component should be analyzed individually. Possible failure modes and their probabilities should be defined for various scenarios.

Table 9.1: Comparison of our Proposed Extended Process with the Previous Process in [16]

Steps	**In the previous process in** [16]	**In our proposed extended process**
Define components and sub-components	Technical components/sub-components are defined.	Technical + socio + AR-related components/sub-components are defined.
Define FPTC rules for all components	Scenarios including failures emanated from technical components/sub-components are considered.	Scenarios including failures emanated from technical + socio + AR-related components/sub-components are considered.
Create error models for the component	Scenarios including failures emanated from technical components/sub-components are considered.	Scenarios including failures emanated from technical + socio + AR-related components/sub-components are considered.
Define stochastic behavior parameters	Scenarios including failures emanated from technical components/sub-components are considered.	Scenarios including failures emanated from technical + socio + AR-related components/sub-components are considered.

9.4 Case study

In this section, we design a case study with the objective of presenting the analysis capabilities provided by the proposed process. First step is to model the system, as shown in part A of the process. Then, Concerto-FLA analysis can be used for qualitative analysis (Part B) and state-based analysis can be

used for quantitative analysis (Part C). We consider an industrial monitoring system introduced in [20]. We use this system as a case study for analyzing AR-equipped socio-technical system.

The industrial monitoring system uses a sensor for receiving raw data. Raw data is processed in server and it is organized to be represented to the user for making decisions. AR can be used for providing graphical or textual instructions for solving a problem, configuring an equipment or maintenance activities. In this example, we consider using AR for providing visual alarm in case of problem in a special equipment under control.

9.4.1 Modeling the System

This system includes technical and socio entities. Technical entity is the monitoring system and socio entities are the user and organization. We model each of these entities based on their description and based on their important aspects.

The technical components of this system are defined based on description of monitoring system as follows:

- **Sensor**: it is a hardware component. It can be various sensors, for example a camera receiving raw data of a specific equipment, which is considered for monitoring.
- **Server**: it is a hardware component. It is a computer that contains processing unit for processing the data.
- **Processing unit**: it is a software component. It processes the received data from sensor and organizes it in a format to be used by the user.
- **AR application interface**: it is a hardware component. It is the interface between the user and the server. It is an screen containing AR technology notations.

The user can be characterized based on its important aspects, which are human functions and human internal states. We use four following modeling elements of the extended metamodel explained in Subsection 9.2.1.

- **Directed paying attention**: it refers to an AR-extended human function. It models the function paying attention when it is directed to a specific position by using AR technology. For example, in this case study, if there is something strange related to the equipment which is under monitoring, then AR technology can be used for displaying a red circle around the strange area.

Thus, the user attention will be directed to the position to make a decision to prevent any probable risk.

- **Training**: it refers to training received by the human.
- **Deciding**: it refers to human deciding function.
- **Executing**: it refers to human executing function.

The organization can be modeled based on important organizational aspects. We use the following modeling elements of the extended metamodel explained in Subsection 9.2.1.

- **Condition**: it refers to the condition of the organization where the monitoring task is performed.
- **Organization and regulation AR adoption**: it refers to an AR-extended aspect. It models the adoption process needed in the organization to be able to use AR.
- **AR guided task**: it refers to the task that AR is used for guiding the human to do that. For example, a task should be defined in an organization that in case of special AR alarm the user should react.

Based on the described entities and their important aspects, we provided the model shown in Figure 9.6. Sensor receives raw data (shown by RD in Figure 9.6) and provides the output for processing unit. Data is processed in processing unit and its output is shown in AR application interface to the user. Organization and regulation AR adoption is influenced by regulation authorities (REG) and it affects on AR guided task defined by organization. AR guided task is also influenced by condition of the organization, which is influenced by condition out of the organization (shown by CON). Output of monitoring system which is a visual description on a screen influences on human directed paying attention and output of the organization influences on training. Finally, human deciding function is influenced by directed paying attention and training. Human executing function is influenced by human deciding function. Output of the system, which is output of the human component is human function (HF).

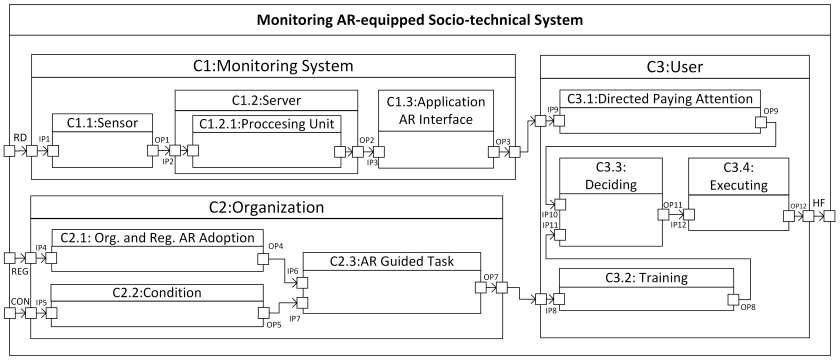


Figure 9.6: Modeling the system

9.4.2 Qualitative Analysis

As it is shown on part B of the Figure 9.5, in order to provide the qualitative analysis, we need to define FPTC rules for all components. These rules should be defined based on individual analysis of components and based on the assumptions of various scenarios. For example, we provide the FPTC rules for a specific scenario and we provide the system behavior based on failure propagation.

- **Definition of scenario**: We assume that the equipment under monitoring is in a situation that it can harm a person. The information is received by the sensor and it is processed by the processing unit and a visual alarm is displayed on the AR display. However, we assume that there is a failure in organization and regulation AR adoption. For example, organization should update regulations in order to include AR related considerations and trainings. Since there is no rule defined in the organization, the required training is not provided for the user. The user's attention is directed to the alarm, but the user does not take the correct decision and does not provide the required execution function to prevent the harm.

- **Modeling of the failure behavior**: In this scenario the organization and regulation AR adoption is behaving as a source (source behavior is explained in Subsection 9.2.4). The input of this component receives noFailure, but in the output it provides valueSubtle. The reason is that organization has not updated rules and regulations to adopt AR (valueSubtle) and the user does not receive the required AR-related training (omission). Since the user

does not receive the required AR-related training, the deciding component provides valueSubtle failure mode in its output. Thus, the user does not provide the required execution (omission). Monitoring system components are behaving as propagational and propagate noFailure from input to output.

- **Analyzing the system behavior**: Analysis annotations are shown in Figure 9.7. ValueSubtle in OP4 means that the AR adoption in organization and regulation is not performed correctly. ValueSubtle failure mode transforms to omission in AR guided task and it propagates in training. Then, in deciding it transforms to valueSubtle and in executing transforms to omission. The failure propagation is shown by blue color.

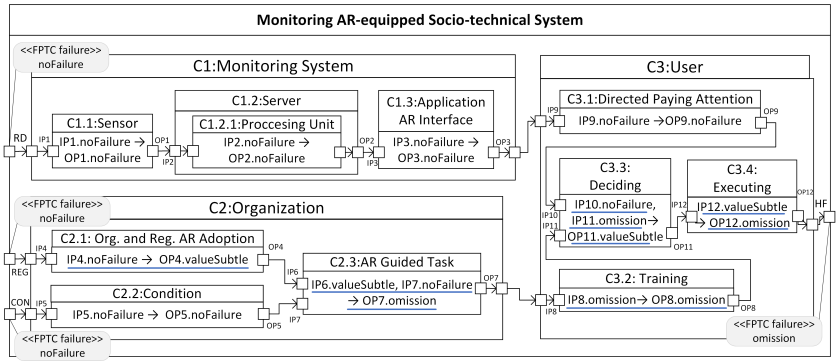


Figure 9.7: Qualitative analysis of the system

- **Interpreting the results**: Based on the back propagation of the results, we can explain how the rules have been triggered. Omission in HF is because of valueSubtle in OP11. ValueSubtle in OP11 is because of omission in IP11 and we continue until IP4, which is input port of organization and regulation AR adoption. Thus, this component caused the failure in the system output. The identified hazard is as follows:

  – **Hazard**: Lack of required AR training

  The reason for this hazard is failure in organization and regulation AR adoption. System failure in this scenario may lead to fatal injuries for people around the intended equipment. Thus, safety goal should be defined to overcome this risk. For example, for this scenario, safety goal can be defined as follows:

– **Safety goal**: The organization should update rules and regulations based on AR and should provide the required AR training.

By using the qualitative analysis and by considering various possible scenarios, various safety goals can be defined. Based on safety goals, system design can be updated and analysis of system behavior can be performed for more iterations to reach the accepted level of safety.

9.4.3 Quantitative Analysis

Based on part C of the Figure 9.5, in order to provide the quantitative analysis, we should model the failure behavior using error models or stochastic parameters. Similar to qualitative analysis these models should be defined based on individual analysis of components and based on the assumptions of various scenarios. For example, we provide stochastic behavior modeling for a specific scenario and we provide the analysis result.

- **Definition of scenario**: Similarly, we assume that the equipment under monitoring is in a situation that it can harm a person. The information should be received by the sensor and it should be processed by the processing unit. Then, a visual alarm should be illustrated through AR display and the user should decide based on illustrated alarm and based on received training from organization to execute a needed task preventing the risk.
- **Modeling of the failure behavior**: In this scenario, for each component we consider possible failure modes and their probabilities as it is shown in Figure 9.8. Probabilities can be defined based on previous accident reports or based on expert opinion. For example, in this scenario, organization has not updated rules and regulations based on AR technology. Thus, failure probability in the Org. and Reg. AR adoption component is high (0.9).
- **Analyzing the system behavior**: In order to perform the analysis, we can consider the hazard related to this scenario and calculate the intended measure or failure mode probability in system output. We consider the same hazard as the one we considered in qualitative analysis, which is lack of required AR training. In this case, we want to calculate the probability of omission failure mode in system output. The result for this assumed scenario is shown in Figure 9.8. Calculation is an automatic task, which can be performed by running the analysis in the toolset. For example, failure in output of executing function would be of type omission or valueSubtle. The probability of omission failure mode is calculated based on the probability

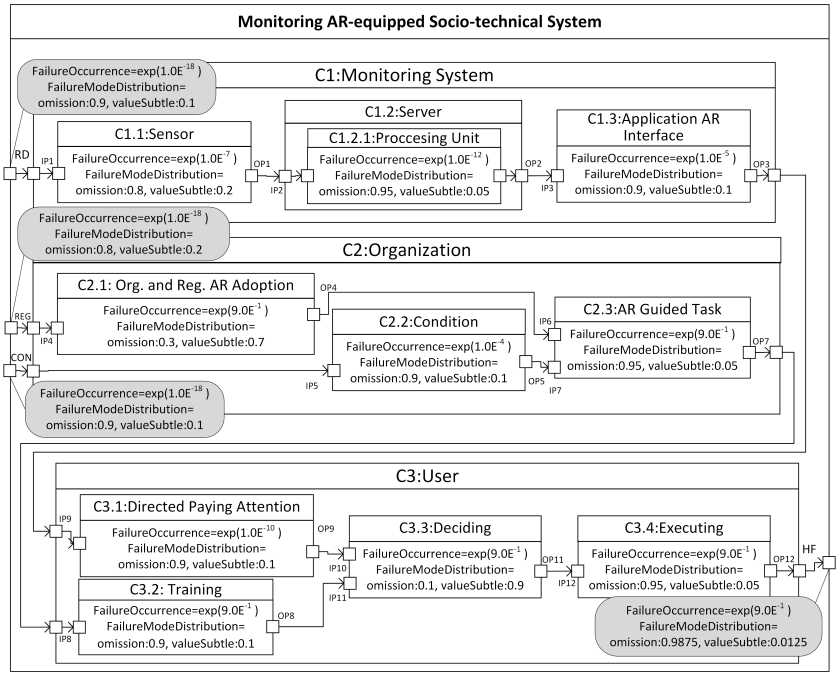


Figure 9.8: Quantitative analysis of the system

of executing function providing an omission failure mode while its input can be different failure modes with different probabilities and all the possible conditions should be considered in the calculation. In this example, probability of failure occurrence in system output (human function) is 0.9, which shows that the reliability of the system from time 0 up to time 1000 hours is around $1 - 0.900 = 0.100$. The probability for omission failure mode will be $0.9 * 0.9875 = 0.88875$.

- **Interpreting the results**: Based on the back propagation of the results, we can explain how the hazard would happen and how much is the probability. For example, in this scenario the probability of omission failure mode in output is 0.88875 and the reason is high probability of failure in organization and regulation AR adoption.

  Similar to the previous scenario, safety goal can be defined in order to decrease the probability and prevent the risk. The probability can be helpful to

decide if a special failure mode in the system output should be overcome or it can be ignored due to low probability of its occurrence.

By using the quantitative analysis and by considering various possible scenarios, various safety goals can be defined. Based on safety goals, safety requirements can be defined and system design can be updated. Then, analysis of system behavior can be performed for more iterations to reach the accepted level of safety.

9.5 Discussion

As it is shown in the case study, there are human functions extended by augmented reality (directed paying attention) and there are AR-related organizational factors (organization and regulation AR adoption and AR guided task). Using modeling elements characterizing AR-extended human functions and AR-related organizational factors in modeling and analyzing the system provides the possibility to include their effects on system behavior while performing the analysis. In the extended metamodel used in our process, there are various modeling elements, which can be helpful in order to incorporate new features of organizations and their effects on human. Globalization, digitalization and networking structure of organizations are also considered. There are various factors leading to post normal accidents discussed in [21] and these factors are included in the extensions as it is explained in [11]. For example, industrial strategy is an organizational modeling element, which can be used to incorporate effect of industrial strategy on system behavior. A failure in industrial strategy can influence on human performance and can lead to system failure. Thus, it is important to model this factor in system modeling and it is crucial to consider its effects on analysis while we analyze system behavior.

Similar to the modeling and analysis capabilities for components, modeling and analysis constructs can be used for modeling and analysis of connections between components. It is an important feature, because based on accident reports there are a lot of situations that failure in the system is not caused by failure in components, but it is caused by failure in connections between components. It is important that we consider various scenarios including ones which system failure is emanated from failures in connections between components.

Analysis results can be used for preparing safety case and arguments to show that a system is acceptably safe. It is required to have several analysis iterations and brainstorm the possible scenarios and possible failures for all

components, subcomponents and connections.

9.6 Related Work

In [22], a framework is proposed for integrated socio-technical enterprise modelling. In this framework, social aspects in addition to technical aspects, internal and external aspects are considered. Eight constructs such as goal, structure, task, etc. are mapped to enterprise models such as goal model, organizational model and process model respectively. The framework is illustrated on a case study from healthcare industry.

In [6], a framework is proposed for conducting the analysis of cyber-socio-technical systems. Concepts and a language are developed to characterize varieties of entities from a knowledge management perspective. Effects of modern challenges of digitalization on organizations and systems in various domains are included in the proposed framework.

Similarity of these studies with our work is consideration of social aspects and their interactions between various entities. The difference is that we also consider augmented reality effects on different socio aspects and its effects on system behavior in general. We incorporate augmented reality effects in the modeling and analysis process.

In [23], a literature review on various studies of risk management on socio-technical systems with the existence of digital transformation is proposed. Various studies are identified and they are categorized based on the steps they have considered for risk management and if human, organization and technology are considered in these steps. The results show that the researches are increasing on human and organization in addition to technology. However, in the risk controlling step, approaches considering all dimensions of socio-technical systems, are required. In our study, we considered all dimensions of socio-technical systems in risk identification, calculation of failure propagation and system behavior analysis. In addition, we considered effect of augmented reality on various parts of socio-technical systems.

9.7 Conclusion

In this paper, we proposed an extension on the synergy of qualitative and quantitative dependability analysis techniques by incorporating AR and socio aspects. We presented this extension by an extended process. In the proposed

extended process, we used extended metamodels for capturing AR-related aspects and considered their effects on system behavior. By implementing the proposed process in the CHESS toolset, it is possible to automatically calculate the failure propagation and failure mode probabilities for AR-equipped socio-technical systems. We illustrated the proposed process for analysis on an industrial monitoring system.

Further research is required to show the potential of the proposed process in more complex case studies within different domains. In addition, we plan to evaluate our proposed process by preparing a questionnaire and collecting expert opinions.

Bibliography

[1] B. Gallina, E. Sefer, and A. Refsdal, "Towards safety risk assessment of socio-technical systems via failure logic analysis," in *International Symposium on Software Reliability Engineering Workshops (ISSRE)*, pp. 287–292, IEEE, 2014.

[2] D. Van Krevelen and R. Poelman, "A Survey of Augmented Reality Technologies , Applications and Limitations," *The International Journal of Virtual Reality (IJVR)*, vol. 9, no. 2, pp. 1–20, 2010.

[3] D. Shcmidt, "Guest editor's introduction: Model-driven engineering," *IEEE Computer*, vol. 2, no. 39, pp. 25–31, 2006.

[4] D. Stewart, J. J. Liu, D. Cofer, M. Heimdahl, M. W. Whalen, and M. Peterson, "AADL-Based safety analysis using formal methods applied to aircraft digital systems," *Reliability Engineering & System Safety*, vol. 213, p. 107649, 2021.

[5] P. H. Feiler and D. P. Gluch, *Model-based engineering with AADL: an introduction to the SAE architecture analysis & design language*. Addison-Wesley, 2012.

[6] R. Patriarca, A. Falegnami, F. Costantino, G. Di Gravio, A. De Nicola, and M. L. Villani, "WAx: An integrated conceptual framework for the analysis of cyber-socio-technical systems," *Safety science*, vol. 136, p. 105142, 2021.

[7] S. Sheikh Bahaei and B. Gallina, "Augmented reality-extended humans: towards a taxonomy of failures – focus on visual technologies," in *European Safety and Reliability Conference (ESREL)*, Research Publishing, Singapore, 2019.

[8] S. Sheikh Bahaei, B. Gallina, K. Laumann, and M. Rasmussen Skogstad, "Effect of augmented reality on faults leading to human failures in socio-technical systems," in *International Conference on System Reliability and Safety (ICSRS)*, IEEE, 2019.

[9] L. Montecchi and B. Gallina, "SafeConcert: A metamodel for a concerted safety modeling of socio-technical systems," in *International Symposium on Model-Based Safety and Assessment (IMBSA)*, pp. 129–144, Springer, 2017.

[10] S. Sheikh Bahaei and B. Gallina, "Extending SafeConcert for Modelling Augmented Reality-equipped Socio-technical Systems," in *International Conference on System Reliability and Safety (ICSRS)*, IEEE, 2019.

[11] S. Sheikh Bahaei and B. Gallina, "A Metamodel Extension to Capture Post Normal Accidents in AR-equipped Socio-technical Systems," in *European Safety and Reliability Conference (ESREL)*, Research Publishing, Singapore, 2021.

[12] CHESS, "ARTEMIS-JU-100022 – Composition with guarantees for high-integrity embedded software components assembly." http://www.chess-project.org.

[13] L. Montecchi, P. Lollini, and A. Bondavalli, "A reusable modular toolchain for automated dependability evaluation," in *Proceedings of the 7th International Conference on Performance Evaluation Methodologies and Tools*, pp. 298–303, Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 2013.

[14] G. Balbo, "Introduction to Stochastic Petri Nets," in *School organized by the European Educational Forum*, pp. 84–155, Springer, 2000.

[15] D. F. Haasl, N. H. Roberts, W. E. Vesely, and F. F. Goldberg, "Fault tree handbook," tech. rep., Nuclear Regulatory Commission, 1981.

[16] L. P. Bressan, A. L. de Oliveira, L. Montecchi, and B. Gallina, "A Systematic Process for Applying the CHESS Methodology in the Creation of Certifiable Evidence," in *2018 14th European Dependable Computing Conference (EDCC)*, pp. 49–56, IEEE, 2018.

[17] CHESS-SBA, "CHESS State-Based Analysis," 2021. https://ic.unicamp.br/ leonardo/tools.html.

[18] S. Mazzini, J. M. Favaro, S. Puri, and L. Baracchi, "CHESS: an Open Source Methodology and Toolset for the Development of Critical Systems.," in *Join Proceedings of EduSymp and OSS4MDE*, pp. 59–66, 2016.

[19] M. Wallace, "Modular architectural representation and analysis of fault propagation and transformation," *Electronic Notes in Theoretical Computer Science*, vol. 141, no. 3, pp. 53–71, 2005. Proceedings of the Second International Workshop on Formal Foundations of Embedded Software and Component-based Software Architectures (FESCA).

[20] D. Pavlov, I. Sosnovsky, V. Dimitrov, V. Melentyev, and D. Korzun, "Case study of using virtual and augmented reality in industrial system monitoring," in *2020 26th Conference of Open Innovations Association (FRUCT)*, pp. 367–375, IEEE, 2020.

[21] J.-C. Le Coze, *Post Normal Accident: Revisiting Perrow's Classic*. CRC Press, 2020.

[22] A. Fayoumi and R. Williams, "An integrated socio-technical enterprise modelling: A scenario of healthcare system analysis and design," *Journal of Industrial Information Integration*, vol. 23, p. 100221, 2021.

[23] J. S. Menzefricke, I. Wiederkehr, C. Koldewey, and R. Dumitrescu, "Socio-technical risk management in the age of digital transformation-identification and analysis of existing approaches," *Procedia CIRP*, vol. 100, pp. 708–713, 2021.

Chapter 10

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

Soheila Sheikh Bahaei and Barbara Gallina.
Technical Report, ISRN MDH-MRTC-345/2022-1-SE, Mälardalen Real-Time Research Center, Mälardalen University, December 2022.

Abstract

One of the most important activities in the safety lifecycle of (socio-technical) safety-critical 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 safety-critical socio-technical 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/globalization, 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, considering their safety perspective, modeling formality, type of analysis, tool support, application domain and supported standards. Finally, we also provide our findings and possible future works based on the analysis of the primary studies, their potential applications and their challenges.

10.1 Introduction

Risk assessment is an essential part of the activities required during the safety lifecycle of (socio-technical) safety-critical 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 human machine teaming should also 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 capture the adequate 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 and increasing automation, 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 meth-

ods 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, they only focus on the human entities of the socio-technical systems. In so doing 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 SLR features of these methods, their methodological steps and their enrichments 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/globalization 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 on risk assessment of safety-critical socio-technical systems, 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 impacts and challenges and provide research directions for future works based on the findings.

The paper is organized as follows. In Section 10.2, we recall the background and discuss related work. In Section 10.3, we present the research method. In Section 10.4, we report about the results of the SLR, which we conducted. In Section 10.5, we discuss the results and threats to validity. In Section 10.6, we draw our conclusion and we present potential future research directions based on our findings.

10.2 Background and Related Work

10.2.1 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 10.1, socio-technical systems are systems including technical and socio entities such as human and organization [2]. Thus, the socio related risks and the risks related to socio and technical teaming are as important as technical related risks to be considered in risk assessment process. Safety-critical systems are "systems whose failure could result in loss of life, significant property damage, or damage to the environment" [3]. Thus,

it is highly important to perform the risk assessment in these systems according to accepted safety standards. Regarding approaches and other practices for performing risk assessment, 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 10.1. *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. 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] another safety perspective (*safety engineering today*) is also introduced 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.

Table 10.1: Comparison between safety perspectives

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

10.2.2 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 hazard identification and accident investiga-

tion methods 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 integrating 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 pub-

lished 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 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.

10.3 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.

10.3.1 Planning the SLR

This subsection describes the execution of the recalled phases.

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 has 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 really important to consider AR-related aspects of the system during the risk assessment. In addition, as mentioned in Section 10.1, 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/globalization, 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.

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 the 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 effects of AR 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 provided explanation in Section 10.2))
- 1.10. Which safety perspective is supported? (safety I, safety II, safety III or safety engineering today (based on Table 10.1))

- **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 Figure 10.1.

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:

- **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

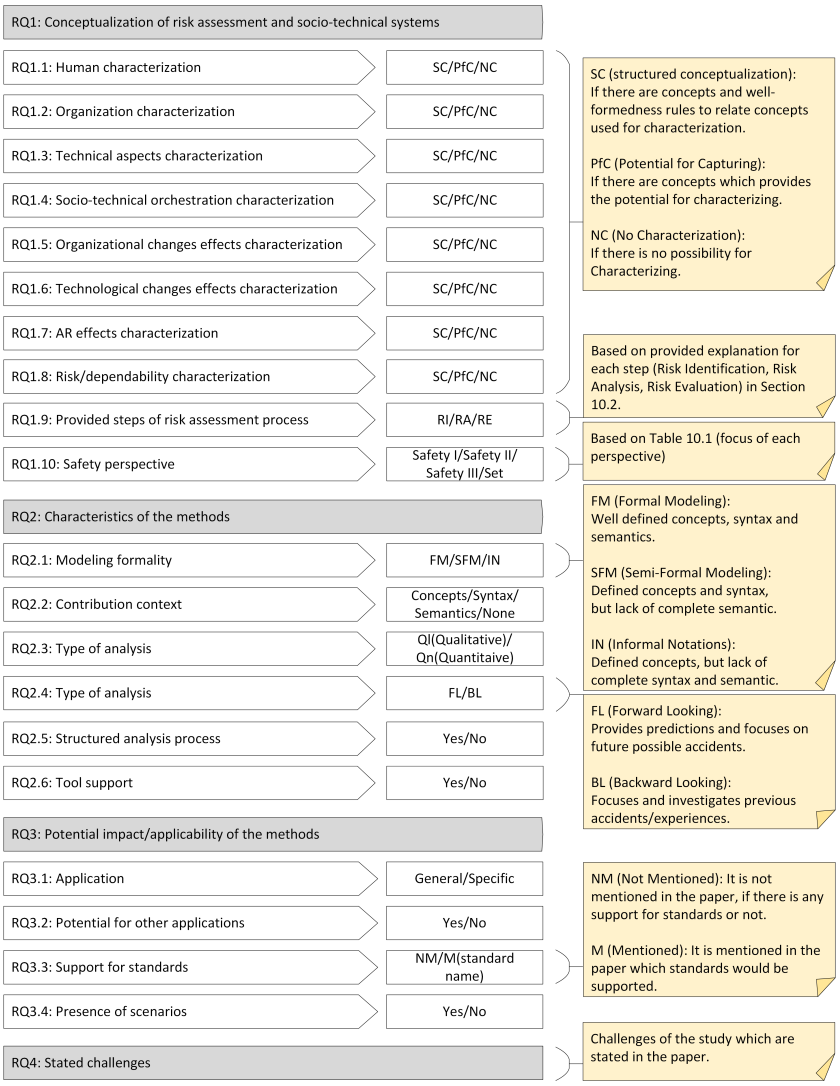


Figure 10.1: Defined abbreviations for possible options of extracted information in relation to each research question

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) systematic literature reviews best practices [26] regarding the usage of the evaluation results. 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 10.2.

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

Table 10.2: Inclusion and exclusion criteria

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.

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 10.3.

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 10.4. 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

Table 10.3: Data extraction criteria

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

be completely analyzed and final primary studies are selected from these papers by considering inclusion and exclusion criteria after reading the papers completely.

10.3.2 Conducting the SLR

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

Data Collection

We apply the SLR protocol described in subsection 10.3.1. 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

Table 10.4: Quality assessment criteria

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 unclear.
	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.

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 Figure 10.2. Then, we performed the title screening 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 10.3.1. 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

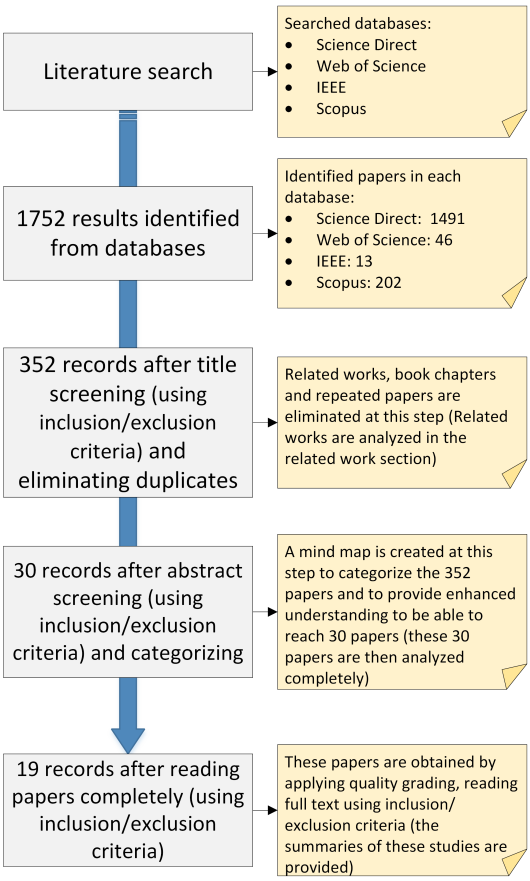


Figure 10.2: Process for papers selection

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 Figure 10.3. 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 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.

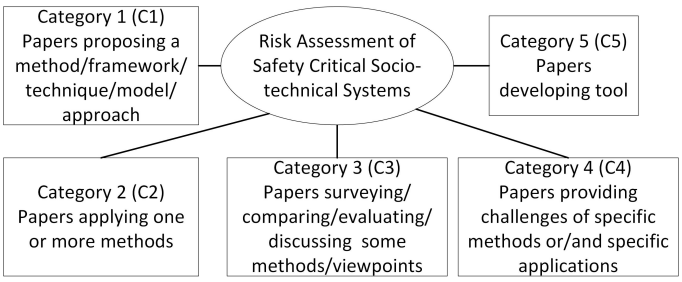


Figure 10.3: Proposed high-level categorization for the identified papers

Most of the studies are assigned to C1 (as we expect, because we did a title screening before this step). We divide this category into four subcategories shown in Figure 10.4. 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 men-

tioned techniques.

Category C1-4 is also divided to five subcategories which are shown in Figure 10.4. In the following paragraphs, we explain about STAMP, STPA, FRAM, Safety II, PRA and BNs briefly and we provide an example.

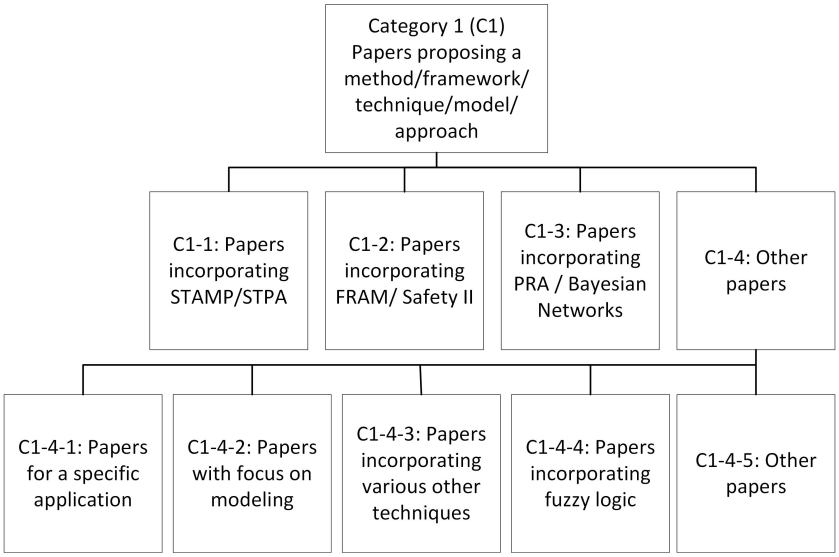


Figure 10.4: Subcategories of the Category 1

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.

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 10.5 using the quality assessment criteria defined in Table 10.4.

Data Extraction

The study overview of the identified 19 primary studies are presented in Table 10.6. 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.

10.4 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

Table 10.5: Quality grading of the primary studies using Table 10.4

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

related to research questions shown in Tables 10.7 - 10.11.

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)

Table 10.6: Selected primary studies

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

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 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 informa-

tion 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 **sce-**

nario 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 taxon-

omy 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 so-**

cio (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 er-

ror, 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 called 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 [2] 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) [54]. 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.

10.5 Discussion

10.5.1 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 10.7, 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

Table 10.7: Summary of the reviewed primary studies in relation to the first research question

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.

consider conceptualizing the related aspects to enable capturing their effects on system safety and risk assessment.

As it is shown in Table 10.8, 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 10.9, in most papers the modeling formality is in the level of informal notation and we can conclude that more research is re-

Table 10.8: Summary of the reviewed primary studies in relation to the first research question (Con.)

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.

quired 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 consider both of them since we learn from the past to prevent the accidents in

Table 10.9: Summary of the reviewed primary studies in relation to the second research question

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

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 10.10, 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

Table 10.10: Summary of the reviewed primary studies in relation to the third 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

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 10.11, 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 scenar-

Table 10.11: Summary of the reviewed primary studies in relation to the fourth 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

ios 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.

10.5.2 Threats to Validity

In this subsection, we discuss about validity of the results based on the guideline provided in [55] and [12]. Specifically, we discuss about possible threats regarding publication bias, identification of primary studies and data extraction consistency.

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 Subsection 10.3.1. 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.

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.

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 10.3 and we defined the abbreviations used for extracted data in Figure 10.1. 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 Tables10.7 - 10.11 with information in relation to each research question.

10.6 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 stan-

dards 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.

Bibliography

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

[2] B. Gallina, E. Sefer, and A. Refsdal, "Towards safety risk assessment of socio-technical systems via failure logic analysis," in *International Symposium on Software Reliability Engineering Workshops*, pp. 287–292, IEEE, 2014.

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

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

[5] J.-C. Le Coze, "Globalization and high-risk systems," *Policy and practice in health and safety*, vol. 15, no. 1, pp. 57–81, 2017.

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

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

[8] C. Dallat, P. M. Salmon, and 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*, vol. 119, pp. 266–279, 2019.

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

[10] N. Leveson, "A new accident model for engineering safer systems," *Safety Science*, vol. 42, no. 4, pp. 237–270, 2004.

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

[12] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," tech. rep., Keele University and Durham University Joint Report, 2007.

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

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

[15] E. Hollnagel, R. L. Wears, and 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.

[16] N. Leveson, "Safety III: A systems approach to safety and resilience," 2020.

[17] T. Aven, "A risk science perspective on the discussion concerning Safety I, Safety II and Safety III," *Reliability Engineering & System Safety*, vol. 217, p. 108077, 2022.

[18] T. Aven, "Risk assessment and risk management: Review of recent advances on their foundation," *European Journal of Operational Research*, vol. 253, no. 1, pp. 1–13, 2016.

[19] H. J. Pasman, W. J. Rogers, and 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*, vol. 55, pp. 80–106, 2018.

[20] A. Adriaensen, W. Decré, and 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*, vol. 5, no. 4, 2019.

[21] R. Sadeghi and F. Goerlandt, "The State of the Practice in Validation of Model-Based Safety Analysis in Socio-Technical Systems: An Empirical Study," *Safety*, vol. 7, no. 4, p. 72, 2021.

[22] N. Berx, W. Decré, I. Morag, P. Chemweno, and L. Pintelon, "Identification and classification of risk factors for human-robot collaboration from a system-wide perspective," *Computers & Industrial Engineering*, vol. 163, p. 107827, 2022.

[23] H. M. Cooper, "Organizing knowledge syntheses: A taxonomy of literature reviews," *Knowledge in society*, vol. 1, no. 1, pp. 104–126, 1988.

[24] B. Kitchenham, E. Mendes, and 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*, pp. 1–10, 2006.

[25] M. Gusenbauer and 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*, vol. 11, no. 2, pp. 181–217, 2020.

[26] M. Campoverde-Molina, S. Luján-Mora, L. Valverde, *et al.*, "Systematic literature review on software architecture of educational websites," *IET Software*, vol. 15, no. 4, pp. 239–259, 2021.

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

[28] J. P. Castellanos Ardila, B. Gallina, and F. Ul Muram, "Compliance checking of software processes: A systematic literature review," *Journal of Software: Evolution and Process*, p. e2440.

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

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

[31] M. M. Chatzimichailidou and I. M. Dokas, "RiskSOAP: Introducing and applying a methodology of risk self-awareness in road tunnel safety," *Accident Analysis & Prevention*, vol. 90, pp. 118–127, 2016.

[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*, vol. 83, no. 2, pp. 229–240, 2004.

[33] M. Schönbeck, M. Rausand, and J. Rouvroye, "Human and organisational factors in the operational phase of safety instrumented systems: A new approach," *Safety science*, vol. 48, no. 3, pp. 310–318, 2010.

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

[35] M. Farcasiu and I. Prisecaru, "MMOSA–a new approach of the human and organizational factor analysis in PSA," *Reliability Engineering & System Safety*, vol. 123, pp. 91–98, 2014.

[36] I. Bider and 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*, pp. 1–12, IEEE, 2015.

[37] A. Mazaheri, J. Montewka, J. Nisula, and P. Kujala, "Usability of accident and incident reports for evidence-based risk modeling–A case study on ship grounding report," *Safety science*, vol. 76, pp. 202–214, 2015.

[38] K. Klockner and Y. Toft, "Accident modelling of railway safety occurrences: the safety and failure event network (SAFE-Net) method," *Procedia Manufacturing*, vol. 3, pp. 1734–1741, 2015.

[39] A. Dehghan Nejad, R. Gholamnia, and A. Alibabaee, "A new framework to model and analyze organizational aspect of safety control structure," *International Journal of System Assurance Engineering and Management*, vol. 8, no. 2, pp. 1008–1025, 2017.

[40] C. Leong, T. Kelly, and R. Alexander, "Incorporating epistemic uncertainty into the safety assurance of socio-technical systems," *Electronic Proceedings in Theoretical Computer Science*, vol. 259, pp. 56–71, 2017.

[41] W. Li, L. Zhang, and W. Liang, "An Accident Causation Analysis and Taxonomy (ACAT) model of complex industrial system from both system safety and control theory perspectives," *Safety science*, vol. 92, pp. 94–103, 2017.

[42] P.-c. Li, L. Zhang, L.-c. Dai, X.-f. Li, and Y. Jiang, "A new organization-oriented technique of human error analysis in digital NPPs: Model and classification framework," *Annals of Nuclear Energy*, vol. 120, pp. 48–61, 2018.

[43] E. Zarei, M. Yazdi, R. Abbassi, and F. Khan, "A hybrid model for human factor analysis in process accidents: FBN-HFACS," *Journal of loss prevention in the process industries*, vol. 57, pp. 142–155, 2019.

[44] A. Lališ, R. Patriarca, J. Ahmad, G. Di Gravio, and B. Kostov, "Functional modeling in safety by means of foundational ontologies," *Transportation research procedia*, vol. 43, pp. 290–299, 2019.

[45] B. Accou and G. Reniers, "Developing a method to improve safety management systems based on accident investigations: The SAfety FRactal ANalysis," *Safety science*, vol. 115, pp. 285–293, 2019.

[46] G. Fu, X. Xie, Q. Jia, Z. Li, P. Chen, and 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*, vol. 134, pp. 47–82, 2020.

[47] J. I. Single, J. Schmidt, and J. Denecke, "Ontology-based computer aid for the automation of HAZOP studies," *Journal of Loss Prevention in the Process Industries*, vol. 68, p. 104321, 2020.

[48] B. Ryan, D. Golightly, L. Pickup, S. Reinartz, S. Atkinson, and N. Dadashi, "Human functions in safety-developing a framework of goals, human functions and safety relevant activities for railway socio-technical systems," *Safety science*, vol. 140, p. 105279, 2021.

[49] S. Sheikh Bahaei, B. Gallina, and M. Vidović, "A case study for risk assessment in AR-equipped socio-technical systems," *Journal of Systems Architecture*, vol. 119, p. 102250, 2021.

[50] N. Chouchani, S. Debbech, and M. Perin, "Model-based safety engineering for autonomous train map," *Journal of Systems and Software*, vol. 183, p. 111082, 2022.

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

[52] B. Accou and F. Carpinelli, "Systematically investigating human and organisational factors in complex socio-technical systems by using the "SAfety FRactal ANalysis" method," *Applied ergonomics*, vol. 100, p. 103662, 2022.

[53] L. Montecchi and B. Gallina, "SafeConcert: A metamodel for a concerted safety modeling of socio-technical systems," in *International Symposium on Model-Based Safety and Assessment*, pp. 129–144, Springer, 2017.

[54] S. Debbech, S. C. Dutilleul, and P. Bon, "An Ontological Approach to Support Dysfunctional Analysis for Railway Systems Design," *J. Univers. Comput. Sci.*, vol. 26, no. 5, pp. 549–582, 2020.

[55] 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.

Chapter 11

Paper E: Technical Report on Assessing Risk of AR and Organizational Changes Factors in Socio-technical Robotic Manufacturing

Soheila Sheikh Bahaei and Barbara Gallina.
Technical Report, ISRN MDH-MRTC-346/2022-1-SE, Mälardalen Real-Time Research Center, Mälardalen University, December 2022.

Abstract

Technological changes such as the use of Augmented Reality (AR) along with the advent of new organizational changes such as digitalization are on the one hand positively changing the way of working but on the other hand they are introducing new risks, potentially leading to not only normal but also post-normal accidents. In our previous work, we have incrementally proposed a novel framework, called FRAAR, for risk assessment of AR-equipped socio-technical systems (i.e., systems integrating human, organisational and technical entities). We have also partly evaluated our framework via an industrial automotive study and by providing comparison and positioning with respect to other related works in a systematic literature review. In this paper, we conduct a new study to evaluate the applicability and effectiveness of our framework in a different domain. To do that, we choose a digitalized socio-technical factory system, focusing on the human-robot collaboration for a realistic diesel engine assembly task using AR-based user interface in an organization affected by organizational changes. Then, we discuss about the extent the conceptualizations provided by the framework are effective to capture the essential information for risk assessment in socio-technical robotic manufacturing, the extent the robotic safety standards are supported (to demonstrate the applicability of the framework in the robotic domain) and the extent of development in risk assessment with respect to AR and organizational changes. Finally, we discuss about validity of our work and we provide our findings and possible future works.

11.1 Introduction

In the contemporary socio-technical systems (i.e., systems integrating human, organisational and technical entities), there is a growth in technological changes such as the use of augmented reality in addition to organizational changes such as digitalization. On one hand, these changes have the potential to improve the system performance but, on the other hand, they may introduce new dependability threats to the system leading to hazards and ultimately to normal as well as post-normal accidents. Post normal accidents [1] are new kinds of accidents due to the new organizational changes such as digitalization and globalization. Since these new organizational changes may introduce new kinds of dependability threats, it is important to consider these new threats while assessing the risk of the recent systems.

Within, the industrial automation sector and more specifically within the manufacturing sector, for instance, automation is being digitalized and the robotic fabrication is being transformed into collaborative fabrication. Hence, the complexity of the digital manufacturing is increasing and potentially leading to post-normal accidents. A recent systematic literature review [2], for instance, shows that technological changes influence digital manufacturing and new challenges are brought in. Within robotic manufacturing, accidents can happen. As it was reported in [3], a robot killed a worker at Volkswagen plant in Germany. This accident happened when the worker was setting up the stationary robot and the robot grabbed and crushed him against a metal plate. To prevent these accidents, it is necessary to investigate the changes and their effects on safety.

Based on ISO 45001:2018 [4], which is a standard that asks businesses to look at hazards posed by "the design of work areas, processes, installations, machinery/equipment, operating procedures and work organization, including their adaptation to the needs and capabilities of the workers involved" [5], *risk* is defined as "a combination of the likelihood of occurrence of a work-related hazardous event or exposure and the severity of injury or ill-health that can be caused by the event or exposure", while *hazard* is defined as source with a potential to cause injury and ill-health. In the specific context of collaborative fabrication, robotics standards also apply and define specific practices for assessing risk.

To assess risk of socio-technical systems (including robotic systems), various techniques exist. In [6], for instance, the author proposes a technique for hazard analysis of human-robot interactions based on the HAZOP (Hazard Operability) technique [7] and UML (Unified Modeling Language) [8]. Specif-

ically, UML is used for partitioning and describing the system. In addition, guide words and guidelines to enable the analyst to imagine possible deviations for each element of the system. The deviations are then transferred to HAZOP tables and their causes, consequences and recommendations are provided. An ergonomic risk assessment is conducted in [9] using process-failure mode effect analysis for different automation levels in human-robot interaction. The proposed risk assessment can be used by manufacturers to assess risk before installing robots in the intended environment. In qualitative assessment, level of severity of potential harm is determined, which can be catastrophic, critical or minimal. In quantitative assessment, metrics are determined and they are compared with risk criteria or critical number (multiplication of severity of the accident and occurrence of the event). Implementing actions for minimizing the likelihood of the risk is considered for risk reduction. In our previous work, we proposed a risk assessment framework called FRAAR (Framework for Risk Assessment in AR-equipped socio-technical systems). This framework includes modelling capabilities for capturing effects of augmented reality and organizational changes on socio-technical system's behavior. To demonstrate the effectiveness of the FRAAR's modelling capabilities for capturing risks caused by human, technical, organizational and AR-related aspects, we conducted a case study based on an automotive case [10]. As documented in [11], our framework also includes modelling capabilities to capture the global distance [12] and factors related to organizational changes leading to post normal accidents [1] such as digitalization. So far, however, our framework has not been applied to the robotic systems, which are systems incorporating organizational changes besides augmented reality.

Hence, in this paper, we fill this gap and choose a digitalized socio-technical factory system with the focus on human robot collaboration for a realistic diesel engine assembly task using AR-based user interface. We use guidelines proposed by Runeson and Höst [13] to conduct our study in a structured manner.

More specifically, in this paper, we aim at analyzing applicability and effectiveness of our previously proposed framework for assessing risk of AR-equipped socio-technical systems with respect to consideration of AR effects, organizational changes and support for standards in robotic domain. For this purpose, we use a case of human robot collaboration and we use percentage of supported risk assessment steps defined by related safety standards and percentage of covered typical human robot interaction failures to demonstrate the applicability and effectiveness of this framework in robotic domain. In addition, we use percentage of identified risk sources with respect to AR and organizational changes in order to illustrate the extension provided by the frame-

work with respect to AR and organizational changes. We consider related safety standards such as ISO 10218-1:2011 [14], which is a standard for robots and robotic devices. We undertake this study based on the guidelines proposed by Runeson and Höst [13]. Finally, we discuss about validity of our work and potential future research directions.

The paper is organized as follows. In Section 11.2, we recall the background and we discuss the related work. In Section 11.3, we present the research method used in this paper. In Section 11.4, we report about how we planned and designed our study. In Section 11.5, we discuss the execution of the study. In section 11.6, we discuss about the results and threats to validity. In Section 11.7, we draw our conclusion and we present potential future research directions based on our findings.

11.2 Background and Related Work

11.2.1 Background

Basic Concepts

For sake of clarity and self-containment, in what follows, we recall the definitions of some key terms (*dependability threats*, *hazard*, *risk*, *harm* and *accident*). *Dependability threats* [15] are *faults, errors* and *failures*. *Fault* is cause of *error*, *error* is cause of *failure* and *failure* (service failure) is deviation of the provided service with respect to the correct service. Thus, in case of propagation, *faults* can lead to *errors* and *errors* can lead to *failures*. This causality chain is shown in Figure 11.1. As it is shown in this figure, *dependability threats* can lead to *hazard*, which is associated with a specific *risk* and *hazard* can lead to *harm* (sometimes referred to as *accident*).

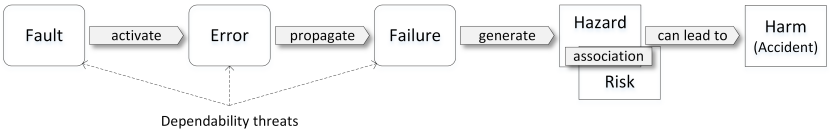


Figure 11.1: Relationships between dependability threats, hazard, risk and harm

A failure may manifest itself in different forms which is called failure mode. There are various categorizations for failure modes. Based on [16], fail-

ure modes are categorized to three categories: 1) provisioning (omission (no output is provided), commission (output is provided when not expected)), 2) timing (early (output is provided too early), late (output is provided too late)), 3) value (course (output not in expected range of value and user can detect), subtle (output not in expected range of value and user can not detect)).

Risk Assessment of AR-equipped Socio-technical Systems

In [10], FRAAR (Framework for Risk Assessment in AR-equipped socio-technical systems) is proposed based on ConcertoFLA analysis technique [17] and by integrating modeling extensions for modeling various socio factors, AR-related factors and factors related to organizational changes. The methodology of the provided framework, shown in Figure 11.2 (using a V-model structure), includes four main steps:

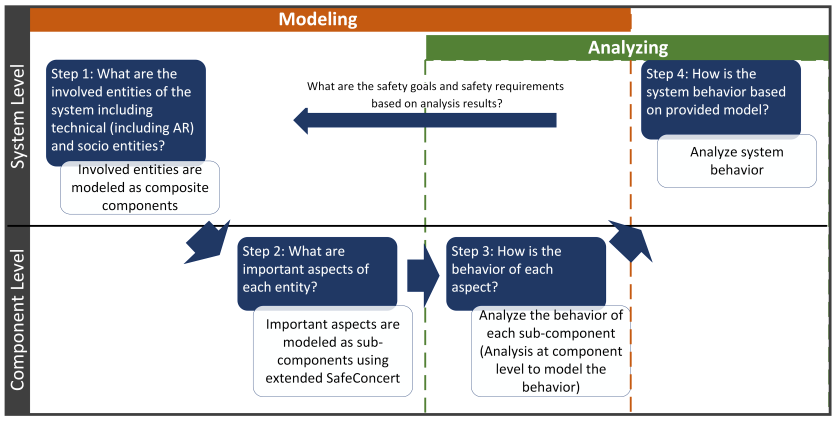


Figure 11.2: Methodology of the FRAAR framework [10]

- Step 1: Identifying the involved entities including socio entities and technical entities (such as AR). The entities are modeled as composite components at system level.
- Step 2: Identifying the important aspects of each entity. This step is done based on SafeConcert modeling language [18] and the extended modeling elements proposed in [11]. The important aspects are modeled as sub-components of the composite component modeling the related entity. Based on extended

SafeConcert modeling language, system element can be a *component* or a *connector* (for modeling connections) and a component can be *socio, software* or *hardware* component. Extended modeling elements include constructs for modeling socio entities which are human and organization shown in Figure 11.3 and Figure 11.4. Modeling elements with gray color show the elements related to organizational changes and modeling elements with dotted line border are AR-related modeling elements.

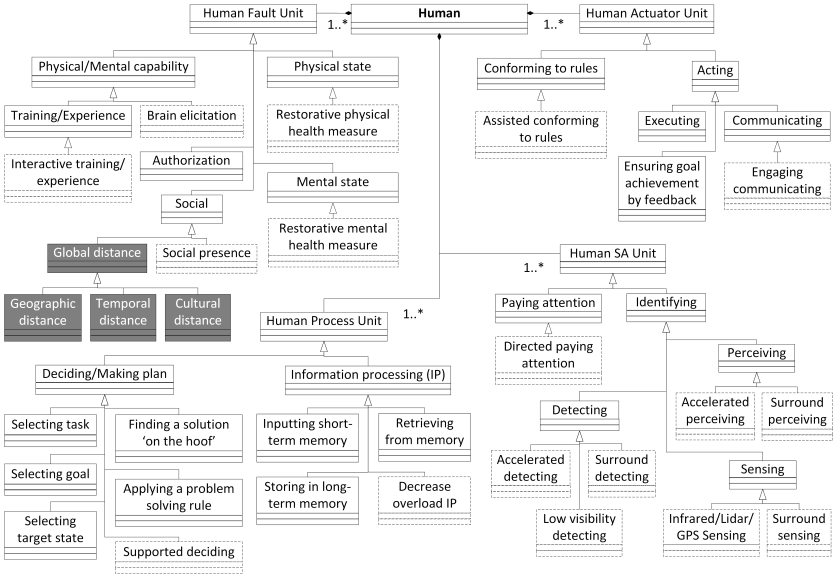


Figure 11.3: Extended human modeling elements [11]

- Step 3: Modeling failure behavior of each sub-component by analyzing its behavior at component level. This step is done by using FPTC syntax [19]. Based on this syntax, FPTC rules are used as logical expressions for relating combinations of input failure modes to output failure modes in each sub-component.

  FPTC syntac for modeling failure behavior is expressed as follows:

  behavior = expression+

  expression = LHS '→' RHS

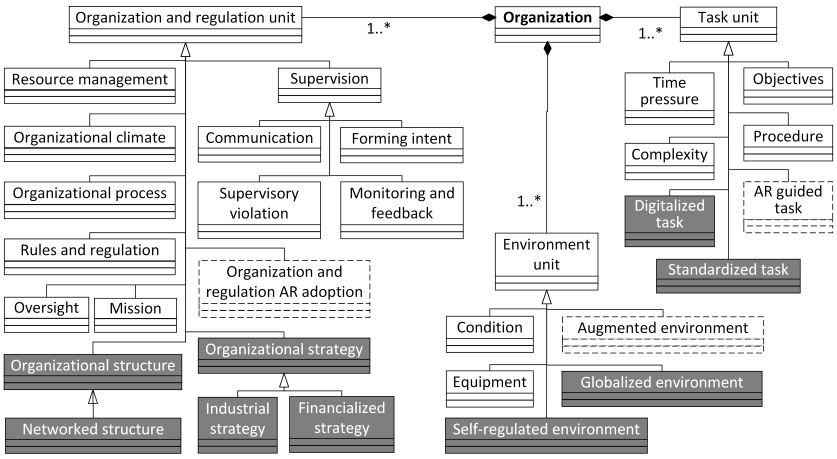


Figure 11.4: Extended organization modeling elements [11]

LHS = portname‘.’ bL | portname

‘.’ bL (‘,’ portname ‘.’ bL) +

RHS = portname‘.’ bR | portname

‘.’ bR (‘,’ portname ‘.’ bR) +

failure = ‘early’ | ‘late’ | ‘commission’ | ‘omission’ |

‘valueSubtle’ | ‘valueCoarse’

bL = ‘wildcard’ | bR

bR = ‘noFailure’ | failure

Using wildcard for an input shows that the behavior of output will be the same regardless of the failure mode of this input and noFailure is used for modeling normal behavior. As an example of a FPTC rule, we can consider ”IP1.noFailure → OP1.noFailure”, which shows modeling failure behavior of a component with input IP1 and output OP1. The FPTC rule shows that normal behavior on IP1 is propagated to OP1. In this case the component’ behavior is classified as *propagational*. If the component produces a failure on the output, while there is normal behavior on its input, then it is classified as *source*. If the component provides normal behavior on its output, while

there is a failure on the input, then it is classified as *sink*. Finally, if the component transforms failure mode on its input to another failure mode on the output, then it is classified as *transformational*.

- Step 4: Analyzing system behavior based on the provided model. The calculation is based on ConcertoFLA analysis technique [17], which is an extension of FPTC analysis technique [19] and it is implemented as a plugin within CHESS toolset [20]. This technique performs qualitative analysis by automatic calculation of failure propagation. Similar to FPTC technique, the system architecture is considered as token-passing network and tokenset is set of possible failures that may be propagated along a connection. Maximal tokenset is calculated for each connection using a fixed-point calculation to obtain system behavior.

The added value of FRAAR framework in comparison to Concerto-FLA is integration of more socio factors, AR-related factors and factors related to organizational changes in the modeling and analyzing processes. Provided failure calculation can be used for identifying and analyzing the possible hazards and their associated risk (by using related safety standards).

As it is shown in Figure 11.2, based on the analysis results, safety goals and safety requirements are defined and another iterations of steps can be performed to judge if the risk is reduced to an acceptable level or not.

Goal Question Metric method

The Goal Question Metric approach (GQM) [21] is a method for measuring based on specific purpose. Based on this method, goals should be defined at the first step. Then, research questions should be defined based on the goals. Finally, metrics should be defined based on the research questions and in a way to reach the defined goals. In this way the metrics provide the possibility to analyze goal achievement. It has been used in several projects such as NASA Goddard Space Flight Centre environment [22].

Robotic Safety Standards

There are six main relevant standards and technical specification for risk assessment in human robot collaboration domain:

- ISO 12100:2010, safety of machinery - General principles for design - Risk assessment and risk reduction [23]

- ISO 10218-1:2011, Robots and robotic devices – Safety requirements for industrial robots – Part 1: Robots [14]
- ISO 10218-2:2011, Robots and robotic devices – Safety requirements for industrial robots – Part 2: Robot systems and integration [24]
- ISO/TS 15066:2016, Robots and robotic devices - Collaborative robots [25]
- ISO 13849-1:2015, Safety of machinery – Safety-related parts of control systems - Part 1: General principles for design [26]
- ISO/DIS 10218-1 Robotics - Safety requirements for robot systems in an industrial environment - Part 1: Robots (under development)

Based on standard ISO 12100:2010 [23] risk is "combination of the probability of occurrence of harm and the severity of that harm". Severity of the harm (S) is classified as S1 (for occasions with slight injuries which are reversible) and S2 (for occasions with serious injuries or death which are irreversible). Probability of occurrence of harm (P) is classified as P1 for occasions where there is chance of avoidance or significant decrement in effects, otherwise it is classified as P2. Based on standard ISO 13849-1:2015 [26], safety-related PLr (required performance level) is determined based on severity of injury (S), possibility of avoiding or limiting harm and probability of occurrence (P) and frequency and/or exposure to hazard (F). Frequency and/or exposure to hazard is classified as F1 for occasions with exposure time less than or equal to 1/20 of overall operating time or frequency of less than or equal to once per 15 min, otherwise it is classified as F2. Determining the required performance level is shown in Figure 11.5.

Standard ISO 10218-1:2011 [14], provides guidelines and requirements for design, measures and use of industrial robots. Basic hazards are recognized for industrial robots and industrial robot systems. However, it is discussed that the numbers and types of hazards are different for various kinds of robots with different automation process and installation complexity. In addition, the sources of the hazards are specific for each particular robot. Standard ISO 10218-2:2011 [24], which is complementary part of ISO 10218-1:2011 specifies the requirements for robot systems, integration and their installation. It also contains significant hazards for robot and robot systems. However, other hazards for specific applications must be addressed based on individual basis.

Based on technical specification ISO/TS 15066:2016 [25] collaborative operation means "state in which a purposely designed robot system and an operator work within a collaborative workspace". The aim of using collaborative

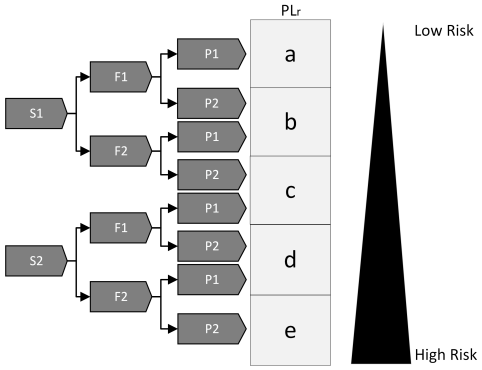


Figure 11.5: Determining required performance level based on [26]

robots is to integrate the competencies of robots such as repetitive performance, precision, power and endurance with the skills and abilities of human. Traditional applications prevented human intervention during the robot activity and it caused lower speed and not being able to automate some operations. In order to have collaboration between human and robot operations, it is essential to consider safety related issues and assess the risk during the collaboration.

Based on standard ISO 12100:2010 [23], risk assessment is the process containing risk analysis and risk evaluation. Risk analysis is the process containing defining the limits of the machine, identifying hazards and estimating the risk. Risk evaluation is "judgment, on the basis of risk analysis, of whether the risk reduction objectives have been achieved". This process is more extended in ISO 10218-2: 2011 by considering robot system which contains industrial robot, end-effector(s) and any supporting machinery, equipment or sensors. In addition, task identification is considered during the risk assessment process to determine the potential occurrence of hazardous situations. Finally, in ISO/TS 15066:2016 the risk assessment is defined containing the following actions:

- Risk analysis
  - Determining the limits of the robot system (intended use and foreseeable misuse)
  - Identifying the hazards and associated hazardous situations
    - considering robot related hazards

    * considering hazards related to the robot system
    * considering application related hazards
    * identifying tasks
  - Estimating the risk of each hazard and hazardous situation
- Risk evaluation
  - Evaluating the risk and taking decision about necessity of reducing the risk based on risk analysis results

In traditional robot system installations it was not possible for human to work in close proximity to robots unless the power of the robot was disconnected. Since in human robot collaboration they can operate in the same workspace while the power of the robot is connected, it is of high importance to take into account potential hazards and their related risk. Technical measures for risk reduction are based on main principles defined in ISO/TS 15066:2016: 1) hazard elimination by design or hazard reduction by substitution. 2) preventing the human to face the hazards or providing a safe state before human come to the hazardous situation, 3) risk reduction during the interventions.

11.2.2 Related Work

In [27], the authors provide a case study for safety analysis in aircraft ground handling services using STAMP (Systems Theoretic Accident Model and Process) causation model [28]. Based on the case study, the limitations of using this model as an organizational management theory are discussed. For example, it is discussed that behavior of people is not represented and by placing a control on behavior without knowing its driving forces, the possible contribution of workers to safety and the complexities that they face are neglected. In addition, it is recommended in this study to use complementary approaches to STAMP in order to consider social dynamics and understanding emergent behavior of systems before introducing control. In [29], the authors provide a case study for modeling and situational awareness analysis of human-computer interaction in the aircraft cockpit. It considers the model with three modules: pilot agent, technical system and environment modules. Two scenarios with human-computer interaction are used and the results are compared with past studies to illustrate the advantages. In [30], the authors provide a case study for modeling heating, ventilation and air-conditioning (HVAC) systems using FRAM (Functional Resonance Analysis Method) [31]. In order to decrease the

complexity of the FRAM model representation, a layered FRAM is presented in this study. Scenarios containing dynamic nature of complex socio-technical systems are considered and the results show better view of the functions and facilitation in analyzing the model.

In [32], the authors discuss the challenges of providing safety in an intelligent human robot collaborative station using the current safety standards and the need for updating and improving them. As it is explained in this paper, according to robotic safety standards, it is mandatory to have risk assessment process for all robotic applications. However, the standards do not support the collaboration in an efficient manner. Manual assembly station from a truck engine final assembly line is used as a use-case and five hazards are identified and described. For each hazards some recommendations are provided to reduce the risk. Finally, a new collaboration mode called "Deliberation in planning and acting" is suggested to include advanced control strategies and improve the current standards. For implementing the suggested mode, control system component should be added to support the deliberation and to provide an agreed plan for safe collaboration. Good understanding of the system and well received education and training is also required by the operator.

In [33], the authors propose a systematic risk assessment approach and apply it to an automated warehouse use case. Based on the proposed approach, different humans with different levels of interaction are identified and their safety requirements are provided. In addition, a list of hazards and their related scenarios are identified using HAZOP method. Finally, the hazards are analyzed, and safety requirements and recommendations are generated to be used in the next risk mitigation phase. Furthermore, a simulation setup is implemented for risk management process using a Virtual Robot Experimentation Platform (V-REP).

In [34], the authors conduct a comprehensive and systematic literature review characterizing works on risk assessment of safety-critical socio-technical systems based on development of conceptualization of socio-technical systems including technological and organizational changes, evolution of safety standards and safety perspectives. In this paper, we aim at investigating applicability and effectiveness of our previously proposed risk assessment framework for AR-equipped socio-technical systems in human robot collaboration domain by considering related safety standards.

11.3 Research Methodology

This section describes the research method that we used for conducting and reporting our study. The research method is based on the guidelines for conducting and reporting case studies by Runeson and Höst [13]. Based on the guidelines, a case study is "an empirical method aimed at investigating contemporary phenomena in their context". There are five main steps for conducting and reporting a case study:

1. **Case study design:** In this step, objectives should be defined and the case study should be planned. In order to define objectives, a set of research questions can be defined. In order to plan the case study, the case (object of study) and case study protocol should be defined.

2. **Preparation for data collection:** In this step, procedures and protocols for data collection should be defined. The principal decisions on methods for collecting data are taken in the design step (defining the case study protocol) and the details of procedures are defined in this step.

3. **Collecting evidence:** In this step, the case study should be executed and data should be collected according to case study protocol. It is important to have several data sources to limit the effects of one data source interpretation. The collected data should provide the ability to address research questions.

4. **Analysis of collected data:** In this step, the collected data should be analyzed by defining an analysis methodology. There would be conclusions from the analysis such as recommendations for future studies.

5. **Reporting the results:** In this step, the results should be reported. The results include answers to the research questions, conclusions, suggestions for future research direction. Threats to validity can be analyzed with proposing countermeasures to reduce them.

we regroup these steps into 3 main activities as follows. Activity one, called planning the study, includes: step 1 and step 2; activity two, called executing the study, includes: step 3, step 4 and activity three, called discussion on the results and their validity which refers to step 5. We explain execution of these activities in the following sections.

11.4 Planning the Study

11.4.1 Objectives

We aim at evaluating the applicability and effectiveness of the FRAAR framework for the purpose of assessing risk of an AR-equipped socio-technical system in human robot collaboration domain with respect to considering effects of AR, organizational changes and support for standards. Based on this objective, we define the following research questions (Qs):

1. Q1: To what extent are the related safety standards in the robotic domain supported (which demonstrates the applicability of the framework in robotic domain)?
2. Q2: To what extent are the conceptualizations provided by the framework effective to capture the essential information for assessing risk in the socio-technical robotic factory?
3. Q3: To what extent is the risk assessment effective with respect to capturing factors related to effects of AR and organizational changes?

Based on these research questions, we define metrics for characterizing and answering the research questions.

Metrics based on Qs:

1. M1: Percentage of supported risk assessment steps provided by standards.
2. M2: Percentage of covered typical human robot interaction failures.
3. M3: Percentage of extensions on identified risk sources with respect to effects of AR and organizational changes.

We show the defined goal, research questions and metrics based on GQM model in Figure 11.6.

11.4.2 Selected Case

In this subsection, we describe an AR-equipped socio-technical system which we selected based on [35] and a taxonomy of typical failures in human robot collaboration proposed in [36].

The system contains the following entities:

- **Technical entities:**

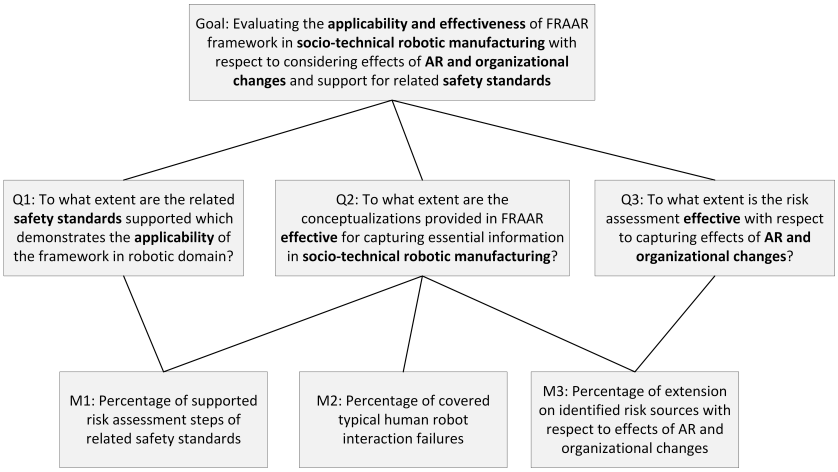


Figure 11.6: Defined goal, questions and metrics using GQM method

- A robot collaborating with the human worker for the engine assembly task.
- An AR user interface for illustrating information such as instructions and robot status to the human worker.

- **Socio entities:**

- A human worker who is working in local diesel engine manufacturing company.
- Diesel manufacturing organization which is responsible for providing rules and regulations, proper work conditions and etc.

Interactive AR-based user interface (UI) proposed in [35] provides capabilities to improve safety of collaboration between human and robot in diesel manufacturing. There are two types of implementations for the AR-based UI: using projector-mirror setup (Figure 11.7) or wearable AR gear (HoloLens) (Figure 11.8). In projector-mirror setup the AR indications are shown on the table around the robot, while in wearable AR HoloLens the indications are shown on the display of the headset used by the human worker. We focus on projector-mirror setup.

The AR-based UI provides six main indications: 1) danger zone which is the region the worker should avoid, 2) changes of human zone, 3) GO and

Figure 11.7: Robot and AR-based UI using projector-mirror [35]

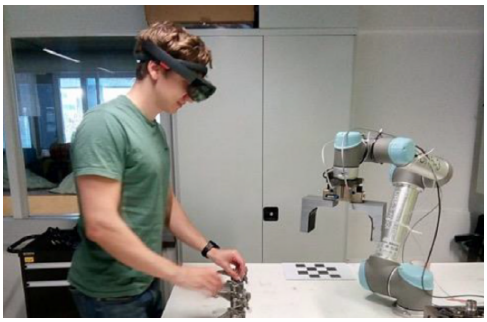


Figure 11.8: Robot and Human using AR-based wearable HoloLens UI[35]

STOP button for starting and stopping the robot, 4) CONFIRM button for verifying and changing of regions, 5) ENABLE button for enabling GO and CONFIRM buttons and 6) a graphical display box containing the instructions and status of the robot.

The considered task is based on [35] which is a part of a real engine assembly task taken from a local company. It contains five sub-tasks which one of them (sub-task 4) is collaborative and we have the focus on that. These subtasks are: 1) installing 8 rocker arms (by human), 2) installing the engine frame (by robot), 3) Inserting 4 frame screws (by robot), 4) installing the rocker shaft (bringing and providing required force by robot and accurate positioning by human), 5) inserting the nuts on the shaft (by robot). The rocker shaft weights 4.3 kg and it is helpful to use a robot for bringing it. However, it is also crucial to consider safety issues while the human is in close distance and dropping the

shaft on human worker's hands would lead to serious injuries.

In [36] a taxonomy of typical failures in human-robot collaboration is provided based on a literature review conducted in the paper. Based on this taxonomy there are two main types of failures in human robot collaboration: *technical failures* and *interaction failures*. *Technical failures* are categorized to *hardware* and *software failures*. *Interaction failures* are categorized to *human errors, environment and other agents*, and *social norm violations*. *Software failures* are categorized to *design failures, communication failures* (categorized to *incorrect data, bad timing, extra data* and *missing data*), and *processing failures* (categorized to *missing events, timing and ordering, abnormal terminations* and *incorrect logic*). *Hardware failures* are categorized to *effectors, power, control* and *sensors failures*. *Human errors* are categorized to *mistakes, slips, lapses* and *deliberate violations*. *Environment and other agents failures* are categorized to *group-level judgment, working environment* and *organizational flaws*.

11.4.3 Study Protocol

Based on [13], there are three types of data collection techniques: first degree (researcher in direct contact with the subjects collecting data in real time such as interview), second degree (researcher collects data without interacting with the subjects such as observation) and third degree (analysis of work artifacts such as using archival data). In this study, we use the third degree data collection technique. However, we use multiple sources of evidence in order to increase trustworthiness of the work. For selecting the case containing augmented reality in a real context, we use [35] which describes an AR-equipped socio-technical system with its real-life context. In order to model technical entities, we use technical details described in the related product websites. In addition, we collect data based on Goal Question Metric method (GQM) [21] which is a goal-oriented measurement technique as we explained in Subsection 11.2.1. Based on this technique, the goal of the study is defined and then research questions are defined based on the goal to trace goal to data intended to define the goal operationally. Finally, metrics are defined based on the research questions for characterizing and answering them to achieve the goal.

11.5 Executing the Study

11.5.1 System Modeling

Based on the first step of the FRAAR framework explained in Subsection 11.2.1, in order to model the system, we need to identify the system entities (as we identified in Subsection 11.4.2). Then, based on the second step, we need to identify the important aspects of each entity. Important aspects are required for modeling sub-components of each composite component representing the related entity. We identify important aspects of the robot collaborating with human using the description provided in [35] and product technical specifications in [37] and [38]. For identifying human and organization important aspects, we use the extended modeling elements of FRAAR framework extracted from [11] and shown in Figure 11.3 and Figure 11.4.

- Important aspects of robot:
  - Control box hardware: it is a hardware for receiving command from computing system and providing control commands for controlling the arm and gripper using its related software.
  - Control box software: it is a software in relation to control box hardware for providing the commands.
  - Arm: it is a hardware for receiving command from control box and providing the required movement.
  - Gripper: it is a hardware for receiving command from control box and providing the required movement.
- Important aspects of projector-mirror UI:
  - RGB-D sensor: it is a hardware for capturing color image (RGB) and depth information from the scene and providing the required information to be sent to the computing system.
  - Computing system hardware: it is a hardware in relation to the computing system software for conducting the computations.
  - Computing system software: it is a software for providing command for robot and for providing the required input for 3LCD projector using the received information from RGB-D sensor.
  - A 3LCD video projector: it is a hardware for receiving information from computing system and providing a 1920*1080 color image with 50 Hz frame rate.

  - Mirror: it is a hardware for increasing the projection area.

- Important aspects of human worker:
  - Mental state: it refers to mental state of human that may influence on human behavior. For example, there may be problem in mental state because of time pressure and it would influence on worker behavior and leads to wrong decision and execution.
  - Detecting: it refers to human detecting function.
  - Deciding: it refers to human deciding function.
  - Executing: it refers to human executing function.
  - Information processing: it refers to human information processing function.
  - Communicating: it refers to human communicating function (for example with other people).
  - Cultural distance: it refers to a factor related to organizational changes. For example, if there is any misunderstanding between the worker and the manager due to distance between their cultures.
  - Interactive training/experience: it refers to a factor related to AR. When AR is used in the system, it is required for the worker to have training/experience to be able to work with AR interface.
  - Conforming to rules: it refers to a human function for conforming to rules.

- Important aspects of diesel manufacturing organization:
  - Financialized strategy: it refers to a factor related to the effects of new organizational changes that causes increasing power of financial actors leading to new strategies.
  - Time pressure: it refers to a factor that may influence on human behavior, because time pressure may cause wrong decision and execution by human.
  - Condition: it refers to the condition provided by the organization.
  - Augmented environment: it refers to the environment provided by using augmented reality. For example, when a projector is used for illustrating AR information, the augmented environment is the virtual displayed information along with the physical environment of the user.
  - Resource management: it refers to managing the resource in organization.

- Organization and regulation AR adoption: it refers to updating rules and regulations based on changes due to AR.
- Equipment: it refers to equipment used for performing the task.
- Organizational process: it refers to daily corporate decisions.
- Oversight: it refers to providing feedback for managers.
- digitalized task: it refers to a factor integrating effects of organizational changes. It refers to task definition provided by organization while the task is digitalized as an organizational change.

An overview of the integration of human worker, AR-based projector-mirror UI, robot and organizational factors is provided in Figure 11.9.

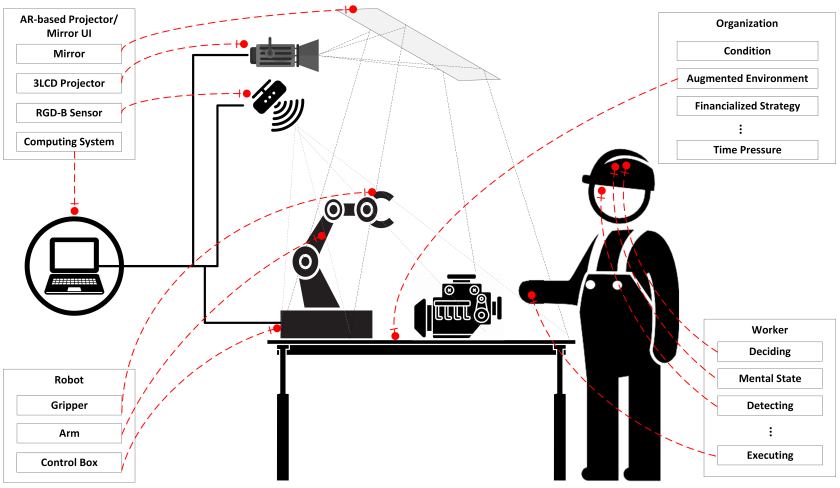


Figure 11.9: Integration of human worker, AR-based projector-mirror UI, robot and organizational factors (adapted from [39] and [40])

In Figure 11.10, we show how the considered AR-equipped socio-technical system is modeled using the extended modeling language of FRAAR framework. Human worker contains nine sub-components with four inputs. Three of human inputs are from organization and one is from system input as communicating input. Interactions between different sub-components are shown in the figure. The output of human worker is Human Action shown by HA.

Robot has five sub-components and one input coming from a computing system which contains the commands which should be executed by the robot.

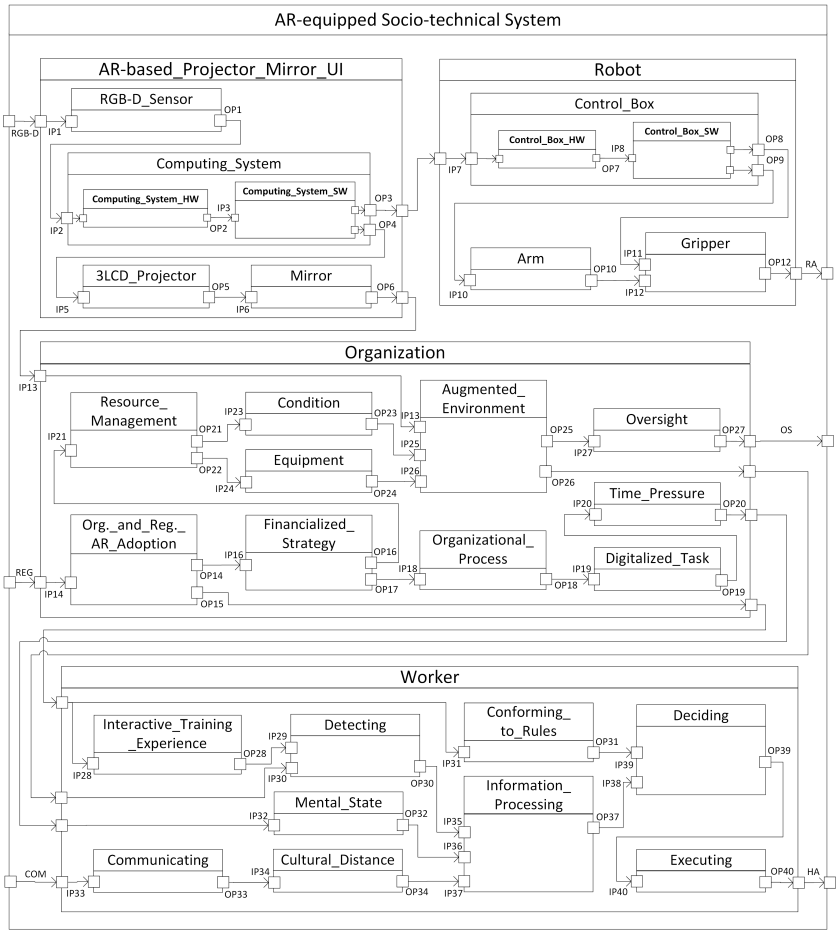


Figure 11.10: Modeling of the AR-equipped socio-technical system

Output of the robot is robot action which is shown by RA. AR-based Projector-mirror UI has six sub-components and one input which is input of the system containing the RGB-D data sensed by sensor, shown by RGB-D.

Organization has ten sub-components and two inputs, one coming from mirror and the other input is connected to the input of the system. The input coming from the system input is influences from regulation authorities shown by REG. The organization has four outputs. One of them is connected to system output shown by OS, which is output of oversight sub-component and provides the feedback for managers about the organization. The other three outputs are from augmented environment, time pressure and organization and regulation AR adoption, which are connected to worker inputs.

11.5.2 System Analysis

This subsection reports on the analysis of the system based on step 3 and step 4 of the FRAAR framework explained in Subsection 11.2.1. We assume that human worker and robot are collaborating to perform sub-task 4 explained in Subsection 11.4.2 and we consider three scenarios as examples and we show the analysis results.

Scenario 1:

Description of the scenario: In this scenario, we assume that failure in the system is emanated from the financialized strategy. For example, because of increasing power of financial actors, new strategies are assigned to increase production. This can lead to changes on definitions of organization process and it causes changes in definition of the digitalized task (for example the collaboration between human and robot should be performed with higher speed). It can cause time pressure for worker. Time pressure can cause improper mental state, incorrect information processing, incorrect deciding and incorrect executing by the human worker and the human worker may move his/her hands under the rocker shaft when the robot is bringing it to install it (value failure mode). The result is a post normal accident, because it is due to new organizational changes.

Modeling failure behavior: The activated FPTC rules are underlined in Figure 11.11. In this scenario, financialized strategy behaves as source and while there is no failure on its input, it produces valueSubtle failure on its output. Organizational process, digitalized task, time pressure, mental state, information processing and deciding sub-components behave as propagational and propagate valueSubtle from their inputs to their outputs and executing sub-component transforms valueSubtle to valueCoarse. The reason is that value

failure in executing function can be detected by user.

Analysis of system behavior: ValueSubtle failure mode on IP18 means that the there is failure in the provided financialized strategy. ValueSubtle propagates to organizational process, digitalized task, time pressure, mental state, information processing, deciding and executing. The failure propagation is shown by blue color.

Interpreting the results: Based on the back propagation of the results, we can explain how the rules are triggered. ValueCoarse on OP40 is because of valueSubtle on OP39 and it is because of valueSubtle on OP37. ValueSubtle on OP37 is because of valueSubtle on OP32 and it is because of valueSubtle on OP20. ValueSubtle on OP20 is because of valueSubtle on OP19 and it is because of valueSubtle on OP18. Finally, valueSubtle on OP18 is because of valueSubtle on OP17.

The results can be helpful to support hazard identification and analysis required by safety standards used in robotic and human robot collaboration.

In this case, unexpected movement by human is the identified hazard and the reason is improper financialized strategy leading to time pressure. System failure in this scenario would lead to sever injury since the human worker would move his/her hands under the rocker shaft when the robot is bringing the shaft to install it. Based on the standard ISO 13849-1:2015 [26] explained in Subsection 11.2.1, severity is s2 and frequency and duration of exposure to the risk is f1 and the possibility of avoiding the risk is p1. Thus, based on Figure 11.5, required performance level is PLr = c, which is quit high.

In this case we define the following safety requirement:

- **Safety requirement**: Evaluation for financialized strategies shall be provided.

Scenario 2:

Description of the scenario: In this scenario, we assume there is failure in the augmented environment, while there is no failure in the augmented reality information provided by the projector and there is also no failure in the condition and equipment provided by the organization. However, the table used for projection of AR information has some patterns on it and it causes that the worker misread (value failure mode) the AR information shown by projector. This leads to wrong detecting, wrong information processing, wrong deciding and wrong executing by the human worker (value failure mode).

Modeling failure behavior: The activated FPTC rules are underlined in Figure 11.12. In this scenario, augmented environment behaves as source and while there is no failure on its inputs, it produces valueSubtle failure on

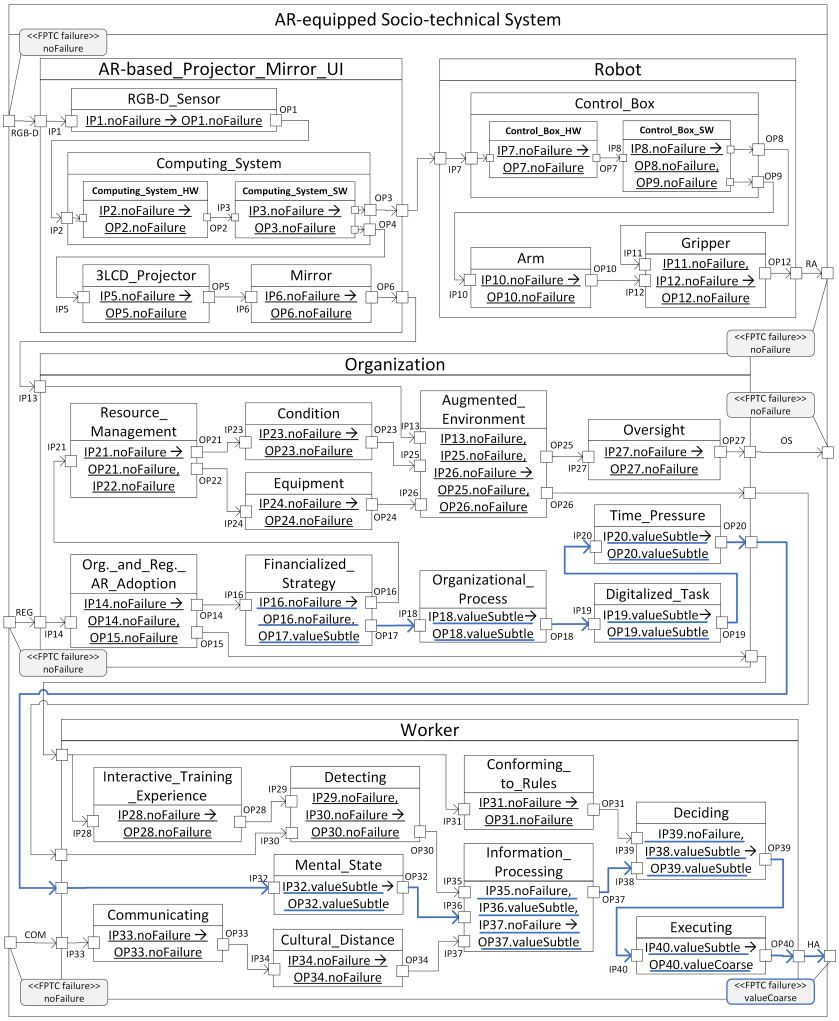


Figure 11.11: Analyzing the AR-equipped socio-technical system (Scenario 1)

its output. Oversight, detecting, information processing and deciding sub-components behave as propagational and propagate valueSubtle from their inputs to their outputs and executing sub-component transforms valueSubtle to valueCoarse. The reason is that value failure in executing function can be detected by user.

Analysis of system behavior: ValueSubtle failure mode on IP30 means that the detected AR information by the user is incorrect. ValueSubtle propagates to information processing, deciding, and executing. The failure propagation is shown by blue color. ValueSubtle failure mode on IP27 means that the oversight received from the organization is not correct. However, since it is not detected by managers it is propagated as valueSubtle and it is not transformed to valueCoarse.

Interpreting the results: Based on the back propagation of the results, we can explain how the rules are triggered. ValueCoarse on OP40 is because of valueSubtle on OP39 and it is because of valueSubtle on OP37. ValueSubtle on OP37 is because of valueSubtle in OP30 and it is because of valueSubtle on OP26.

In this case also, unexpected movement by human (failure in human action) is the identified hazard and the reason is failure in augmented environment. Similar to the previous scenario, system failure in this scenario would lead to sever injury since the human worker may move his/her hands under the rocker shaft when the robot is bringing the shaft to install it. In this case also severity is s2 and frequency and duration of exposure to the risk is f1 and the possibility of avoiding the risk is p1. Thus, based on Figure 11.5, required performance level is PLr = c, which is quit high.

To reduce this risk, it is possible to limit the speed of the robot using mechanical safety design of the gripper. However, it may affect on system performance and efficiency. Another possibility is to provide necessary display requirements as part of safety requirements in order to prevent intervention in the augmented environment. Thus, in this case we define the following safety requirement:

- **Safety requirement**: The environment shall conform to the requirements of AR integration.

Scenario 3:

Description of the scenario: In this scenario, we assume there is failure in control box software. This can lead to failure in arm and gripper movements leading to drop of shaft (value failure mode).

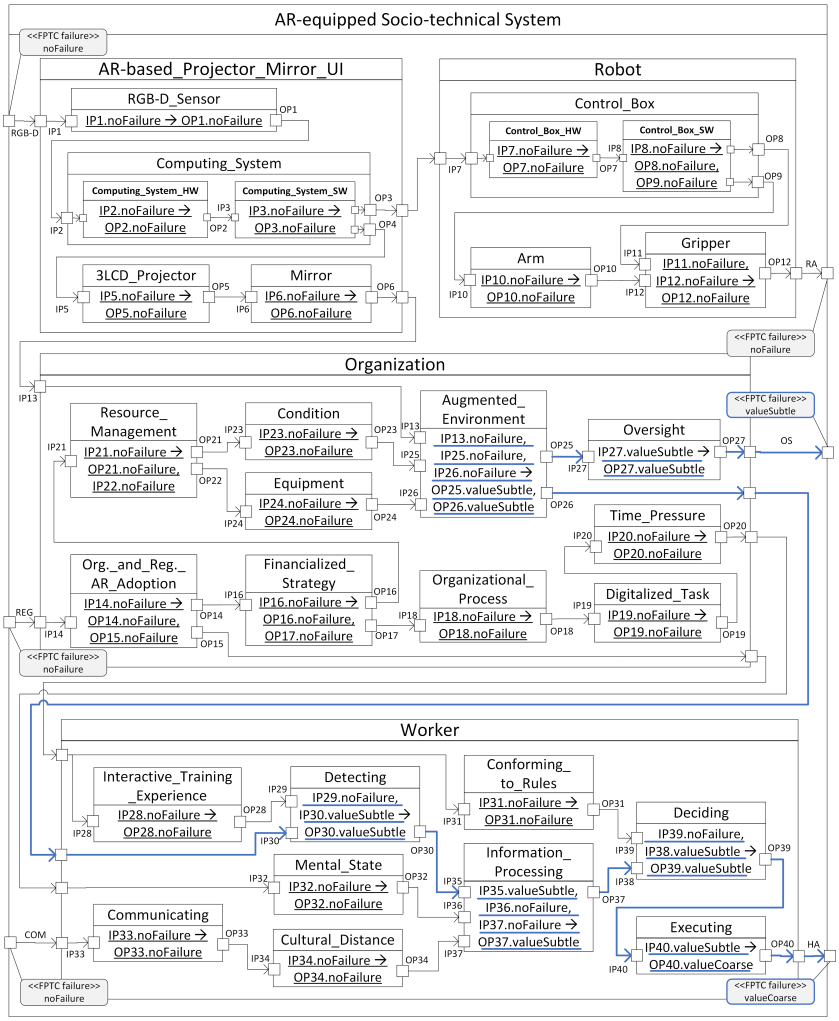


Figure 11.12: Analyzing the AR-equipped socio-technical system (Scenario 2)

Modeling failure behavior: The activated FPTC rules are underlined in Fig11.13. In this scenario, control box software behaves as source and while there is no failure on its input, it produces valueSubtle failure on its output. Arm sub-component behaves as propagational and propagates valueSubtle from its input to its output and gripper sub-component transforms valueSubtle to valueCoarse. The reason is that value failure in robot movement can be detected by user.

Analysis of system behavior: ValueSubtle failure mode in IP10 means that the there is failure in the provided command from control box. ValueSubtle propagates to gripper. The failure propagation is shown by blue color.

Interpreting the results: Based on the back propagation of the results, we can explain how the rules are triggered. ValueCoarse on OP12 is because of valueSubtle on OP8 and OP10 and valueSubtle on OP10 is because of valueSubtle on OP9. ValueSubtle on OP8 and OP9 is because of failure in control box software.

In this case, drop of shaft is the identified hazard and the reason is improper provided command by control box. System failure in this scenario would lead to sever injury since the human worker's hands may be under the rocker shaft when the robot drops it. In this case severity is s2 and frequency and duration of exposure to the risk is f1 and the possibility of avoiding the risk is p2. Thus based on Figure 11.5, required performance level is PLr = d, which is high.

In this case we define the following safety requirement:

- **Safety requirement**: The computing system shall actively monitor the status of the control box.

Similarly, we can consider various other scenarios and update the system analysis based on them to investigate further risk sources, their effects and related safety requirements.

In this section, we applied the FRAAR framework for three example scenarios using some important aspects of socio and technical entities to illustrate how the modeling and analysis is conducted and how we can identify risk sources and related safety requirements. There is the possibility to consider more important aspects and extend the modeling and analysis. For example, in Table 11.1 and 11.2 , we provide further possible risk sources in relation to socio aspects using the extended modeling elements which are integrated in the FRAAR framework. We show the risk sources in connection with effects of organizational changes or AR with gray color to be able to illustrate the extent of risk assessment extension with respect to effects of AR and organizational changes.

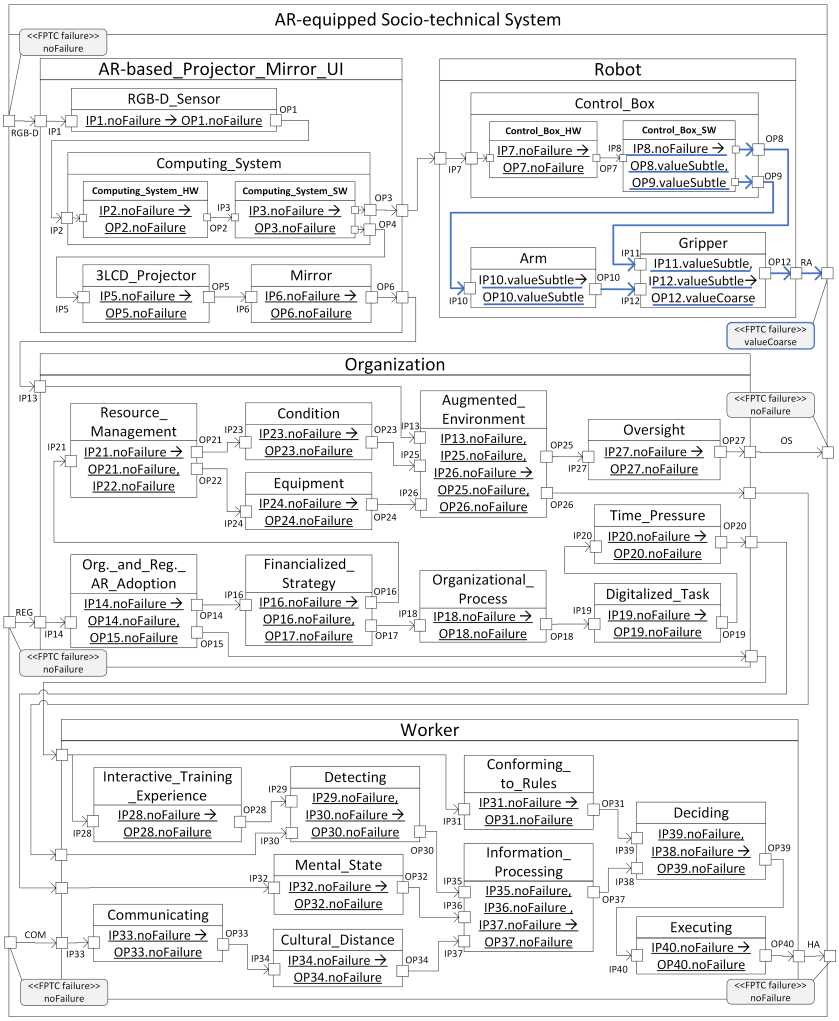


Figure 11.13: Analyzing the AR-equipped socio-technical system (Scenario 3)

Table 11.1: Identified list of dependability threats/risk sources

Identified risk sources	**Description**	**Safety requirement**
Training/experience problem	The required training is not (properly) provided for the user to perform the assembly task	Training shall be provided based on best practices
Interactive training/experience problem	The required training is not (properly) provided for the user to work with AR interface	AR-related training shall be provided based on best practices
Social presence problem	The user is fully taken by AR technology and miss the connectivity with other people and environment	The user shall receive notification through the system in case of receiving crucial communication requirement
Cultural distance problem	Communication between user and manager is affected by culture causing misinterpretation	Guidelines shall be provided for defining critical communication keywords
Physical state problem	There is injury or physical problem in the user body	Minimum level of required physical state for starting the work shall be defined
Mental state problem	There is problem in psychological state of the user	Minimum level of required psychological state for starting the work shall be defined
Deciding/ making plan problem	There is problem in deciding and making plan	Evaluation for deciding competence shall be provided
Supported deciding problem	Problem in deciding which is based on guidance provided by AR technology	Evaluation of AR notifications for supporting deciding shall be provided
Information processing problem	the user has problem in processing information	Evaluation for information processing competence shall be provided
Paying attention problem	The user has problem in paying attention during the task performance	Evaluation of AR notifications for paying attention competence shall be provided
Directed paying attention problem	There is problem in directing attention of user by AR-based UI	Evaluation of AR notifications for directed paying attention shall be provided
Identifying problem	The user has identification problem	Evaluation for identifying competence shall be provided
Perceiving problem	The user has perceiving problem	Evaluation for perceiving competence shall be provided
Surround perceiving problem	The user can not perceive surrounding environment as it is intended by AR	Evaluation of AR notifications for surround perceiving shall be provided
Sensing problem	The user has problem in sensing	Evaluation for sensing competence shall be defined
Accelerated perceiving problem	The user can not accelerate perceiving as it is intended by AR	Evaluation of AR notifications for accelerated perceiving shall be provided
Conforming to rules problem	The user has problem in conforming to rules	Evaluation for conforming to rules competence shall be provided
Executing problem	The user has problem in executing	Evaluation for executing competence shall be provided
Communicating problem	The user has problem in communicating	Evaluation for communicating competence shall be provided
Ensuring goal achievement by feedback problem	The user has problem in ensuring goal achievement by feedback	Evaluation for ensuring goal achievement by feedback competence shall be defined

Table 11.2: Identified list of dependability threats/risk sources (Cont.)

Identified risk sources	**Description**	**Safety requirement**
Resource management problem	There is problem in managing resources in the organization	Guidelines shall be provided for resource management
Organizational process problem	There is problem in daily corporate decisions	Guidelines shall be provided for organizational process
Organizational climate problem	There is problem in organization culture and policy	Guidelines shall be provided for organizational climate
Rules and regulations problem	There is problem in rules and regulations	Guidelines shall be provided for organizational rules and regulations
Oversight problem	There is problem in providing feedback for managers	Guidelines shall be provided for organizational oversight
Networked structure of organization problem	There is problem because of the networked structure of organization	Guidelines shall be provided for organizing networked structure
Supervision communication problem	There is problem in communication between the supervisors	Guidelines shall be provided for communication at supervision level
Monitoring and feedback problem	There is problem in monitoring and feedback	Guidelines shall be provided for monitoring and feedback
Organization and regulation AR adoption problem	Rules and regulations are not updated based on changes due to AR	Updates shall be provided for rules and regulations based on AR changes
Organizational industrial strategy problem	There is problem in industrial strategy defined by organization	Evaluation of organizational industrial strategy shall be provided based on best practices
Organizational financialized strategy problem	There is problem in financialized strategy defined by organization	Evaluation of organizational financialized strategy shall be provided based on best practices
Condition problem	There is problem in condition	Conditional evaluation shall be provided
Equipment problem	There is problem in equipment required for performing the task	Equipment evaluation shall be provided
Self-regulated environment problem	There is problem in self-regulated environment of the organization	Evaluation of self-regulated environment of the organization shall be provided based on best practices
Augmented environment problem	There is problem in the integration of AR and the environment	The environment shall conform to the requirements for AR integration
Time pressure problem	Time pressure is imposed by organization	Evaluation for time adequacy shall be provided
Task objectives problem	Task objectives are not (properly) defined	Guidelines shall be provided for defining task objectives
Task complexity problem	The task is too complex	Defined tasks shall be evaluated in terms of complexity
Digitalized task problem	There is a problem due to the digitalization of the task	Evaluation of digitalization shall be provided
AR guided task problem	There is a problem in the definition of the task which is guided by AR	Evaluation of definition of AR guided task shall be provided
Standardized task problem	There is a problem due to the standardization	Evaluation of standardization shall be provided

As it is shown in this table, there are various risk sources in relation to effects of AR and organizational changes which are identified and analyzed using the extended modeling elements.

11.6 Discussion on the results and their validity

11.6.1 Discussion on the results

In this subsection, we discuss on the results and how metrics are calculated to answer the research questions to reach the goal.

Results for the First Research Question

In Section 11.5, we illustrated how the framework can be applied in robotic domain and how the standards can be used for evaluating the risk. In order to calculate the percentage of supported risk assessment steps provided by related safety standards (first metric), we show the risk assessment steps based on robotic standards explained in Subsection 11.2.1 and we show different activities of FRAAR framework which support them in Table 11.3.

As it is explained in Subsection 11.2.1, based on extended risk assessment definition provided in ISO/TS 15066:2016 [25], risk assessment contains two main activities: *risk analysis* and *risk evaluation*. The first step in risk analysis is *determining the limits of the robot system (intended use and foreseeable misuse)*. In step 1 of the FRAAR framework shown in Figure 11.2, involved entities should be defined. Then, in step 2, important aspects of each entity should be modeled and in step 3, the behavior of each aspect is analyzed. Defining the entities, modeling their important aspects and their behavior as we illustrated in Section 11.5, can be helpful for *determining the limits containing the intended use and foreseeable misuse*. Thus, we can conclude that these activities required for risk assessment are supported by the first three steps of the FRAAR framework. The second step of risk analysis is *identifying the hazards and associated hazardous situations (considering hazards related to robot, robot system and application and identifying tasks)*. This step is also supported by the analysis results from step 4 of the FRAAR framework. Furthermore, *estimating the risk of each hazard and hazardous situation* is supported by the analysis results from step 4. In addition, as we explained in the three example scenarios in Subsection 11.5.2, we can estimate the risk of each hazard and hazardous situation. Finally, *risk evaluation and deciding about necessity of reducing the*

Table 11.3: Supported risk assessment steps based on robotic standards by FRAAR risk assessment activities

Risk assessment step based on standard	**FRAAR risk assessment activity**
1. Risk analysis	Defining the involved entities and their important aspects, modeling their behavior and analyzing system behavior (step 1, 2, 3 and 4)
1.1. Determining the limits of the robot system	Defining the involved entities, their important aspects and their behavior (step 1, 2 and 3)
1.1.1. Defining intended use	Defining the involved entities, their important aspects and their behavior (step 1, 2 and 3)
1.1.1. Defining foreseeable misuse	Defining the involved entities, their important aspects and their behavior (step 1, 2 and 3)
1.2. Identifying the hazards and associated hazardous situations	Analyzing system behavior (step 4)
1.2.1. Considering robot related hazards	Analyzing system behavior (step 4) by considering technical hazards
1.2.2. Considering hazard related to robot system	Analyzing system behavior (step 4) by considering technical and socio hazards
1.2.3. Considering application related hazards	Analyzing system behavior (step 4) by considering technical and socio hazards
1.2.4. Identifying tasks	Defining the involved entities and their important aspects (step 1 and 2)
1.3. Estimating the risk of each hazard and hazardous situation	Analysis results from step 4
2. Risk evaluation	Analysis results from step 4
2.1. Evaluating the risk and taking decision about necessity of reducing the risk based on risk analysis results	Analysis results from step 4

risk is also supported by analysis results from step 4 of the FRAAR framework as it was explained for three example scenarios in Subsection 11.5.2.

As it is shown in Table 11.3, all tasks/sub-tasks defined based on standards in robotic domain are supported by FRAAR framework and it shows that 100 percent of risk assessment steps of robotic safety standards are supported using the FRAAR framework.

Results for the Second Research Question

For this research question we calculate the second metric (percentage of covered typical human robot interaction failures). However, first and third metric are also in alignment with demonstrating the effectiveness of the framework

in socio-technical robotic manufacturing with respect to considering effects of AR and organizational changes and support for related safety standards. In order to calculate the percentage of covered typical human robot interaction failures, we use the taxonomy proposed in [36], explained in Subsection 11.4.2. In Table 11.4, it is shown how failures are covered by the available modeling elements/failure modes/failure behaviors in FRAAR risk assessment framework.

As it is shown in this table, 28 failures of the total 29 failures are covered by the available modeling elements, failure modes and failure behaviors in the FRAAR framework. Based on these results about 96 percent of the typical human robot interaction failures are supported by FRAAR framework, which is a generic risk assessment framework. In the following paragraphs, we explain more about details of the assignments shown in the table.

As we explained in Subsection 11.2.1, *technical failures* can be modeled using *hardware/software components* and then failure behavior can be modeled by defining possible failure modes in the inputs and by defining FPTC rules for each component. Similarly, *software and hardware failures* can be modeled using *software and hardware components* and *communication failures* can be modeled using *connectors*. For example, in modeling and analysis of our selected case in Section 11.5, we show how the software and hardware components are used for modeling technical failures. Equipment component can be used for modeling *design failures*. More details about equipment component are in [41], where we have previously proposed the extensions in relation to organizational factors. We also illustrated how we can use this component in Section 11.5. *Incorrect data, bad timing, extra data* and *missing data* can be modeled by using *value failure mode, early/late, commission* and *omission failure modes* as explained in Subsection 11.2.1.

Processing failures can be modeled by modeling a component failure behavior as *source* as explained in Subsection 11.2.1. It shows that a technical component is producing failure and there is problem in the processing. *Missing events, timing and ordering, abnormal terminations* and *incorrect logic* can be modeled by using different failure modes in the source behavior.

Effectors failures, power failures, control failures and *sensor failures* can be modeled using hardware component and defining their behavior and possible failure modes.

Based on the definition provided in [36], *interaction failures* are failures due to uncertainties in interaction between human, environment and other agents. These failures can be modeled by *socio components* and human errors can be modeled by using human components.

For *mistakes, slips, lapses* and *deliberate violations* there are specific com-

Table 11.4: Covered typical human robot interaction failures

Typical human robot interaction failure	**Available modeling element/failure mode/-failure behaviors in FRAAR for modeling the failure**
1. Technical failures	Technical components
1.1. Software failures	Software component
1.1.1. Design failures	Equipment component
1.1.2. Communication failures	Connector
1.1.2.1. Incorrect data	Value failure mode
1.1.2.2. Bad timing	Early or late failure mode
1.1.2.3. Extra data	Commission failure mode
1.1.2.4. Missing data	Omission failure mode
1.1.3. Processing failures	Source failure behavior
1.1.3.1. Missing events	Omission failure mode
1.1.3.2. Timing and ordering	Early or late failure mode
1.1.3.3. Abnormal terminations	Commission failure mode
1.1.3.4. Incorrect logic	Value failure mode
1.2. Hardware failures	Hardware component
1.2.1. Effectors failures	Hardware component
1.2.2. Power failures	Hardware component
1.2.3. Control failures	Hardware component
1.2.4. Sensors failures	Hardware component
2. Interaction failures	Socio components
2.1. Human errors	Human components
2.1.1. Mistakes	Selecting goal component
2.1.2. Slips	Acting component
2.1.3. Lapses	Information processing component
2.1.4. Deliberate violations	Conforming to rules component
2.2. Environmental and other agents failures	Environment unit component
2.2.1. Group-level judgment	Organizational climate component
2.2.2. Working environment	Environment unit component
2.2.3. Organizational flaws	Organization and regulation unit component
2.3. Social norm violations	-

ponents named *selecting goal, acting, information processing* and *conforming to rules components*, respectively. These components can be used for modeling the assigned failures as it is completely explained in [42].

Finally, *environment and other agents failures* and *working environment failures* can be modeled using *environment unit component*, organizational flaws can be modeled using *organization and regulation unit component* and *group-level judgement* (for example failure due to effects of group-level judgements on human actions) can be modeled using *organization climate component*. There are no associated modeling element for modeling *social norm violations* (for example failure in robot behavior due to not being in compliance with social norm).

Most of the failures in the considered taxonomy are technical failures and failures related to socio aspects are not intensely investigated, while these socio failures, in addition to effects of AR and organizational changes are considered in our extensions to a great extent.

Results for the Third Research Question

In order to calculate the percentage of extension in risk assessment with respect to effects of AR and organizational changes (third metric), we use the number of identified risk sources which are in connection with AR and organization changes divided by the total number of identified possible risk sources discussed in Subsection 11.5.2, Table 11.1. There are 16 identified risk sources in connection with AR and organizational changes in total of 41 identified possible risk sources, which shows 39 percent extension in the risk assessment with respect of effects of AR and organizational changes. From the 16 identified risk sources in connection with AR and organizational changes, 7 of them are in connection with organizational changes with the potential to result in post normal accidents. Therefore, 17 percent extension in risk assessment is provided in order to prevent post-normal accidents.

11.6.2 Discussion on the validity

As it is described in [13], validity of a study discusses the trustworthiness of the results and to what extent the results may be biased by subjective viewpoint of the researcher. We use three aspects of validity, which are introduced in the study containing construct validity, internal and external validity.

Construct validity

This aspect refers to the extent of representation of operational measures based on research questions. We defined operational measures based on the research questions using GQM method. We considered defining operational measures in a way to be able to use data which is possible for us to collect and use it to answer the research questions. For example, we defined typical human robot interaction failure coverage as operational measure in order to measure effectiveness of capturing the essential information for assessing risk in socio-technical robotic factory. This selection was affected by considering that it was possible for us to measure coverage using a typical failure taxonomy in human robot collaboration domain. Thus, some extent of subjectivity is not avoidable, meanwhile we tried to perform it with subjectivity as low as possible.

Internal validity

This aspect refers to considering different causal relations affecting an investigated factor and not missing some of them. In our case, we considered percentage of supported risk assessment steps based on standards, percentage of human robot interaction failure coverage and percentage of extensions with respect to effects of AR and organizational changes as three distinct metrics for measuring support for standards, the extent of effectiveness of the framework and development of risk assessment with respect to effects of AR and organizational changes, respectively. We defined our goal, research questions and metrics based on GQM method in order to consider causal relations affecting our goal, which can be helpful to increase internal validity. However, we are aware of some limitations in relation to internal validity. For example, in the system modeling and designing various scenarios, we considered different assumptions, which can lead to missing some causal relations affecting on system behavior. In modeling and analyzing system behavior, we have considered simplifications and in reality, much more effort is required to investigate various causal relations and to investigate fulfillment of the assumptions.

External validity

This aspect refers to possibility of generalization of the findings. We have discussed about generalization of the FRAAR risk assessment in [10] and one of the main purposes of the empirical study conducted in this paper is demonstrating the applicability of the framework in a new domain, which is in line with demonstrating that the framework can be used as a general framework in

different domains for risk assessment of AR-equipped socio-technical systems with respect to effects of AR and organizational changes.

11.7 Conclusion and Future Work

In this paper, we provided a complementary evaluation of FRAAR framework for risk assessment of a socio-technical system in human robot collaboration domain with respect to effects of use of augmented reality as a new technology and with respect to new organizational changes. We used a digitalized socio-technical factory system containing human robot collaboration using AR-based user interface. We evaluated effectiveness of the framework by calculating the percentage of the covered typical failure modes in the human robot collaboration domain, the percentage of supported risk assessment steps based on safety standards in robotic domain and the percentage of development of the identified risk sources with respect to AR effects and organizational changes.

In future, we aim at conducting a comparative study to compare the results of applying FRAAR risk assessment framework with other risk assessment frameworks in the context of AR-equipped socio-technical systems. In addition, we plan to implement the conceptual extensions proposed in the FRAAR framework by proposing extensions in syntax and semantics of the extended modeling language to enable automating the analysis process and providing tool support. Another important issue for further research is also investigating on risk reduction and defining measures for mitigating the identified risks.

Bibliography

[1] J.-C. Le Coze, *Post Normal Accident: Revisiting Perrow's Classic*. CRC Press, 2020.

[2] E. H. D. R. da Silva, A. C. Shinohara, E. P. de Lima, J. Angelis, and C. G. Machado, "Reviewing digital manufacturing concept in the industry 4.0 paradigm," *Procedia CIRP*, vol. 81, pp. 240–245, 2019.

[3] Associated Press in Berlin, "Robot kills worker at volkswagen plant in germany," 2015. https://www.theguardian.com/world/2015/jul/02/robot-kills-worker-at-volkswagen-plant-in-germany.

[4] ISO 45001:2018, "Occupational health and safety management systems - Requirements with guidance for use," 2016. https://www.sis.se/en/produkter/management-system/occupational-health-and-safety-management-systems/ss-iso-450012018/.

[5] E. Gasiorowski-Denis, "Toward a healthier manufacturing industry," 2018. https://www.iso.org/news/ref2269.html.

[6] J. Guiochet, "Hazard analysis of human–robot interactions with HAZOP–UML," *Safety science*, vol. 84, pp. 225–237, 2016.

[7] T. A. Kletz, "Hazop—past and future," *Reliability Engineering & System Safety*, vol. 55, no. 3, pp. 263–266, 1997.

[8] G. Booch, J. Rumbaugh, and I. Jackobson, "UML: Unified Modeling Language," 1997.

[9] R. T. Stone, S. Pujari, A. Mumani, C. Fales, and M. Ameen, "Cobot and robot risk assessment (carra) method: an automation level-based safety

assessment tool to improve fluency in safe human cobot/robot interaction," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 65, pp. 737–741, SAGE Publications Sage CA: Los Angeles, CA, 2021.

[10] S. Sheikh Bahaei, B. Gallina, and M. Vidović, "A case study for risk assessment in AR-equipped socio-technical systems," *Journal of Systems Architecture*, vol. 119, p. 102250, 2021.

[11] S. Sheikh Bahaei and B. Gallina, "A Metamodel Extension to Capture Post Normal Accidents in AR-equipped Socio-technical Systems," in *European Safety and Reliability Conference (ESREL)*, Research Publishing, Singapore, 2021.

[12] J. Noll and S. Beecham, "Measuring global distance: A survey of distance factors and interventions," in *International Conference on Software Process Improvement and Capability Determination*, pp. 227–240, Springer, 2016.

[13] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," *Empirical software engineering*, vol. 14, no. 2, pp. 131–164, 2009.

[14] ISO(10218-1), "Robots and robotic devices – Safety requirements for industrial robots – Part 1: Robots," 2011. https://www.iso.org/standard/51330.html.

[15] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," *IEEE transactions on dependable and secure computing*, vol. 1, no. 1, pp. 11–33, 2004.

[16] D. J. Pumfrey, *The principled design of computer system safety analyses.* PhD thesis, University of York, 1999.

[17] B. Gallina, E. Sefer, and A. Refsdal, "Towards safety risk assessment of socio-technical systems via failure logic analysis," in *2014 IEEE International Symposium on Software Reliability Engineering Workshops*, pp. 287–292, IEEE, 2014.

[18] L. Montecchi and B. Gallina, "SafeConcert: A metamodel for a concerted safety modeling of socio-technical systems," in *International Symposium on Model-Based Safety and Assessment*, pp. 129–144, Springer, 2017.

[19] M. Wallace, "Modular architectural representation and analysis of fault propagation and transformation," *Electronic Notes in Theoretical Computer Science*, vol. 141, no. 3, pp. 53–71, 2005.

[20] A. Cicchetti, F. Ciccozzi, S. Mazzini, S. Puri, M. Panunzio, A. Zovi, and T. Vardanega, "CHESS: a model-driven engineering tool environment for aiding the development of complex industrial systems," in *Proceedings of the 27th IEEE/ACM International Conference on Automated Software Engineering*, pp. 362–365, ACM, 2012.

[21] V. R. Basili, "Software modeling and measurement: the goal/question/-metric paradigm," tech. rep., 1992.

[22] V. R. B. G. Caldiera and H. D. Rombach, "The goal question metric approach," *Encyclopedia of software engineering*, pp. 528–532, 1994.

[23] ISO(12100), "safety of machinery - General principles for design - Risk assessment and risk reduction," 2010. https://www.iso.org/standard/51528.html.

[24] ISO(10218-2), "Robots and robotic devices – Safety requirements for industrial robots – Part 2: Robot systems and integration," 2011. https://www.iso.org/standard/41571.html.

[25] ISO/TS15066, "Robots and robotic devices - Collaborative robots," 2016. https://www.sis.se/en/produkter/manufacturing-engineering/industrial-automation-systems/industrial-robots-manipulators/isots150662016/.

[26] ISO(13849-1), "Safety of machinery – Safety-related parts of control systems – Part 1: General principles for design," 2015. https://www.iso.org/standard/69883.html.

[27] D. Passenier, A. Sharpanskykh, and R. J. de Boer, "When to STAMP? A case study in aircraft ground handling services," *Procedia Engineering*, vol. 128, pp. 35–43, 2015.

[28] N. Leveson, "A new accident model for engineering safer systems," *Safety science*, vol. 42, no. 4, pp. 237–270, 2004.

[29] X. Zhang, Y. Sun, Y. Zhang, and S. Su, "Multi-agent modelling and situational awareness analysis of human-computer interaction in the aircraft cockpit: A case study," *Simulation Modelling Practice and Theory*, vol. 111, p. 102355, 2021.

[30] I. T. de Souza, A. C. Rosa, A. C. J. Evangelista, V. W. Tam, and A. Haddad, "Modelling the work-as-done in the building maintenance using a layered FRAM: A case study on HVAC maintenance," *Journal of Cleaner Production*, vol. 320, p. 128895, 2021.

[31] H. Erik, *FRAM: the functional resonance analysis method: modelling complex socio-technical systems*. CRC Press, 2017.

[32] A. Hanna, K. Bengtsson, P.-L. Götvall, and M. Ekström, "Towards safe human robot collaboration-risk assessment of intelligent automation," in *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, vol. 1, pp. 424–431, IEEE, 2020.

[33] R. Inam, K. Raizer, A. Hata, R. Souza, E. Forsman, E. Cao, and S. Wang, "Risk assessment for human-robot collaboration in an automated warehouse scenario," in *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, vol. 1, pp. 743–751, IEEE, 2018.

[34] S. Sheikh Bahaei and B. Gallina, "Risk Assessment of Safety Critical Socio-technical Systems: a Systematic Literature Review," *Submitted to Safety Science Journal*, 2022.

[35] A. Hietanen, R. Pieters, M. Lanz, J. Latokartano, and J.-K. Kämäräinen, "AR-based interaction for human-robot collaborative manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 63, p. 101891, 2020.

[36] S. Honig and T. Oron-Gilad, "Understanding and resolving failures in human-robot interaction: Literature review and model development," *Frontiers in Psychology*, vol. 9, 2018.

[37] "Universal robots." https://www.universal-robots.com/cb3/. Accessed: 2022-09-05.

[38] "RG2-Grippe." https://onrobot.com/en/products/rg2-gripper. Accessed: 2022-09-05.

[39] "Flaticon database of free icons." https://www.flaticon.com/. Accessed: 2022-09-05.

[40] "Vecteezy resources of photography, videos and vector illustrations." https://www.vecteezy.com/. Accessed: 2022-09-05.

[41] S. Sheikh Bahaei, B. Gallina, K. Laumann, and M. Rasmussen Skogstad, "Effect of augmented reality on faults leading to human failures in socio-technical systems," in *International Conference on System Reliability and Safety (ICSRS)*, IEEE, 2019.

[42] S. Sheikh Bahaei and B. Gallina, "Augmented reality-extended humans: towards a taxonomy of failures – focus on visual technologies," in *European Safety and Reliability Conference (ESREL)*, Research Publishing, Singapore, 2019.