Assessing Risk of AR and Organizational Changes Factors in Socio-technical Robotic Manufacturing

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

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, organizational and technical entities (such as AR)). 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 design and execute our study to apply our framework and 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 effectiveness of the risk assessment with respect to AR and organizational changes. Finally, we discuss about validity of our work and we provide our findings and intended future work.

Index Terms

risk assessment, socio-technical systems, augmented reality, organizational factors, human robot collaboration

1 Introduction

Nowadays, socio-technical systems (i.e., systems integrating human, organizational and technical entities) contain 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. From a recent systematic literature review [2], it emerges 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 with a well defined scope on 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]. Specifically, UML is used for partitioning and describing the system. In addition,

Corresponding author: S. Sheikh Bahaei (soheila.sheikh.bahaei@mdu.se)

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

to enable the analyst to imagine possible deviations for each element of the system, guide words and guidelines are provided. 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 (FMEA) 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, the 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). To reduce risk by minimizing its likelihood, mitigation actions are implemented. In our previous work, we proposed a risk assessment framework called FRAAR (Framework for Risk Assessment in AR-equipped socio-technical systems) [10]. 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. To conduct our study in a structured manner, we use guidelines proposed by Runeson and Höst [13]. More specifically, in this paper, we aim at analyzing the applicability and the effectiveness of our previously proposed framework for assessing risk of AR-equipped socio-technical systems with respect to considering effects of AR, organizational changes and support for standards in robotic domain. For this purpose, we use a case of human robot collaboration and we use the percentage of the supported risk assessment steps defined by related safety standards and the percentage of the covered typical human robot interaction failures to demonstrate the applicability and effectiveness of this framework in the robotic domain. In addition, we use the percentage of extension on the identified risk sources with respect to effects of AR and organizational changes in order to illustrate the effectiveness of the framework with respect to effects of AR and organizational changes. We consider relevant safety standards and specifications in the robotic domain. Finally, we discuss about validity of our work and potential future research directions.

The paper is organized as follows. In Section 2, we recall the background and we discuss the related work. In Section 3, we present the research method used in this paper. In Section 4, we report about how we planned and designed our study. In Section 5, we discuss the execution of the study. In section 6, we discuss about the results and threats to validity. In Section 7, we draw our conclusion and we present potential future research directions based on our findings.

2 Background and Related Work

2.1 Background

2.1.1 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* are *faults, errors* and *failures* [14]. *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 [14]. Thus, in case of propagation, *faults* can lead to *errors* and *errors* can lead to *failures*. This causality chain is shown in Fig 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*).

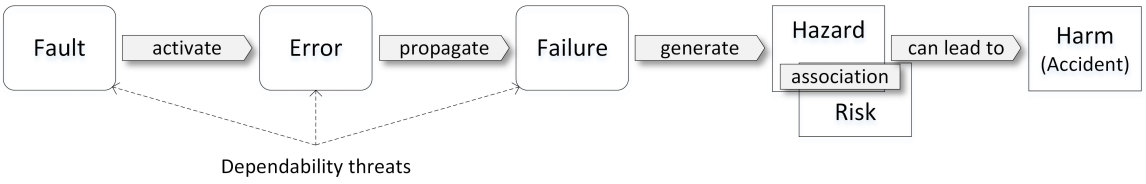

Fig. 1: Relationships between dependability threats, hazard, risk and harm

A failure may manifest itself in different forms which is called failure mode. There are various categorizations for failure modes. Based on [15], failure 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)).

2.1.2 Risk Assessment of AR-equipped Socio-technical Systems: FRAAR [10], which is a framework for risk assessment in AR-equipped socio-technical systems, is proposed based on Concerto-FLA analysis technique [16] 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 Fig. 2 (using a V-model structure), includes four main steps:

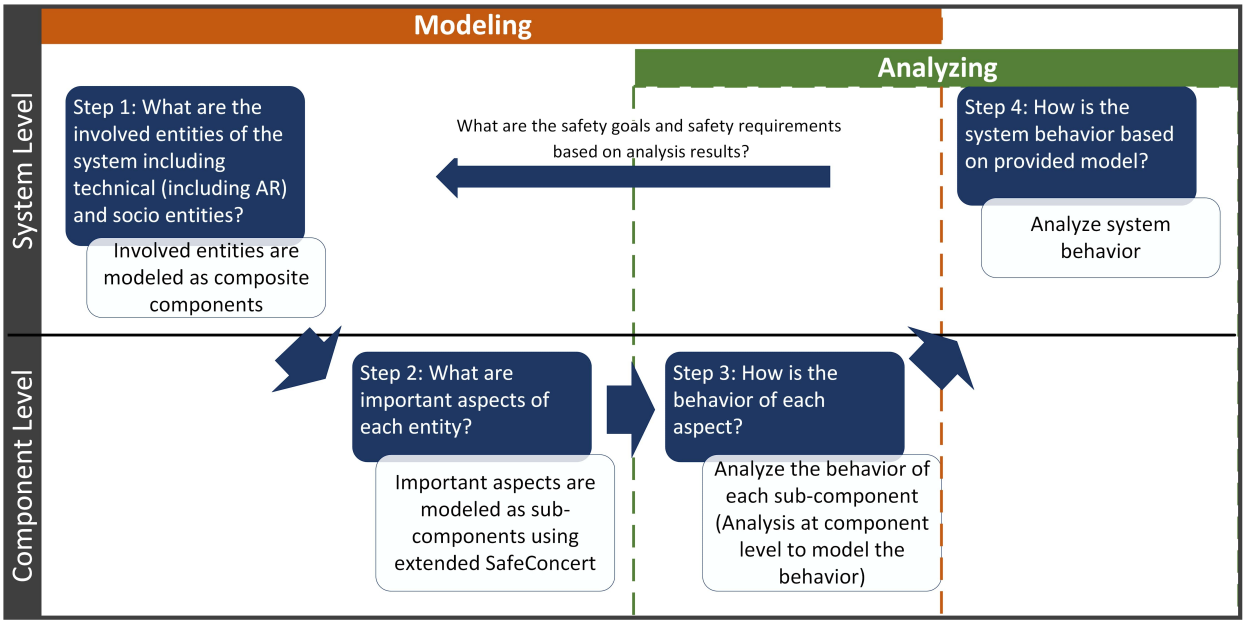

Fig. 2: Methodology of the FRAAR framework [10]

- Step 1: Identifying the involved entities including socio entities and technical (such as AR) entities. 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 [17] 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 Fig. 3 and Fig 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.
- Step 3: Modeling failure behavior of each sub-component by analyzing its behavior at component level. This step is done by using FPTC syntax [18]. 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.
  As an example of a FPTC rule, we can consider "IP1.noFailure → OP1.noFailure", which shows modeling failure behavior of a component with one input (IP1) and one output (OP1). The FPTC rule shows that normal behavior (noFailure) 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 the input, then it is classified as *source* (for example "IP1.noFailure → OP1.late"). If the component provides normal behavior on the output, while there is a failure on the input, then it is classified as *sink* (for example "IP1.late → OP1.noFailure). Finally, if the component transforms failure mode on the input to another failure mode on the output, then it is classified as *transformational* (for example "IP1.late → OP1.omission"). Possible failure modes are explained in Subsection 2.1.1. Using wildcard for an input shows that the behavior of the output will be the same regardless of the failure mode of the input and noFailure is used for modeling normal behavior. For example "IP1.wildcard → OP1.noFailure" shows that there will be normal behavior in the output regardless of the failure mode in the input.
- Step 4: Analyzing system behavior based on the provided model. The calculation is based on Concerto-FLA analysis technique [16], which is an extension of FPTC analysis technique [18] and it is implemented as a plugin within CHESS toolset [19]. 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 Fig. 2, based on the analysis results, safety goals and safety requirements are defined and other iterations of steps can be performed to be able to judge if the risk is reduced to an acceptable level or not.

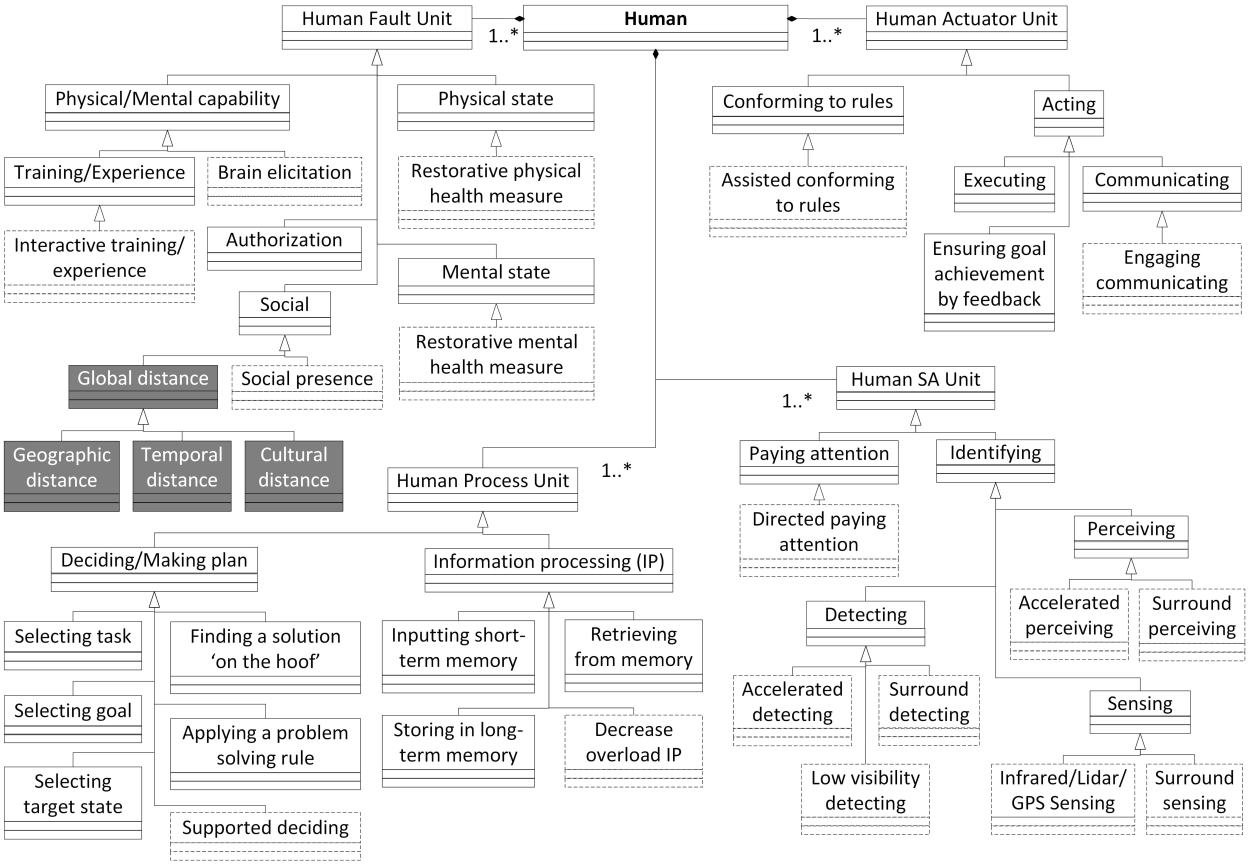

Fig. 3: Extended human modeling elements [11]

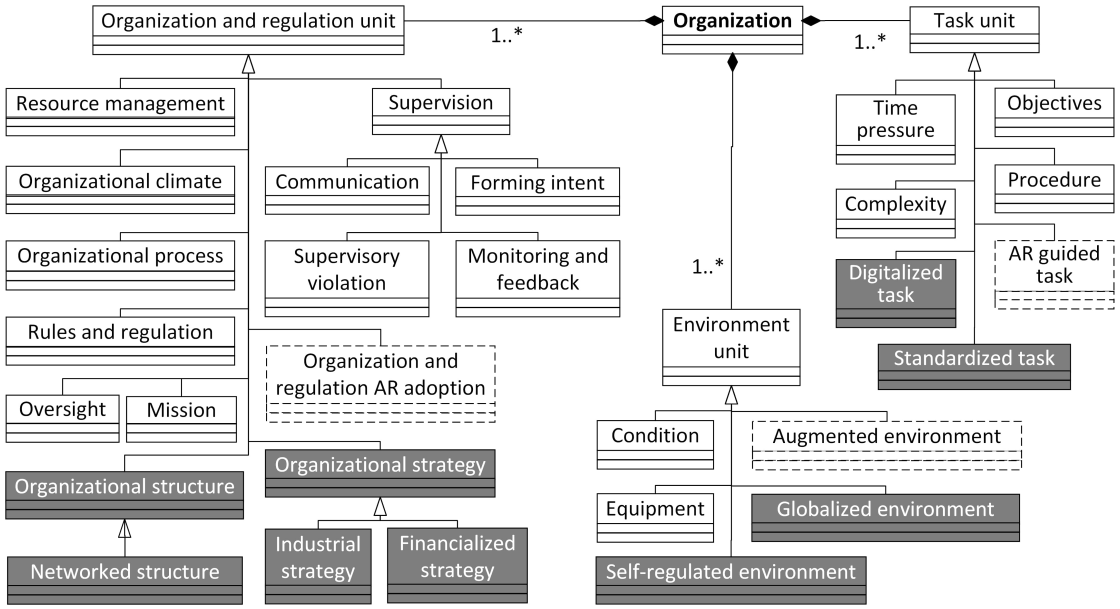

Fig. 4: Extended organization modeling elements [11]

2.1.3 Goal Question Metric method: The Goal Question Metric (GQM) [20] is a method for measuring based on specific purpose. Based on this method, goals shall be defined at the first step. Then, research questions shall be defined based on the goals. Finally, metrics shall 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.

Three levels of this method are shown in Fig 5, the goal which is conceptual level is traced to questions (operational level) which operationally define the goal. Then, metrics, which are defined for interpreting the data with respect to the questions, provide the quantitative level. Different metrics may be used for answering to each question.

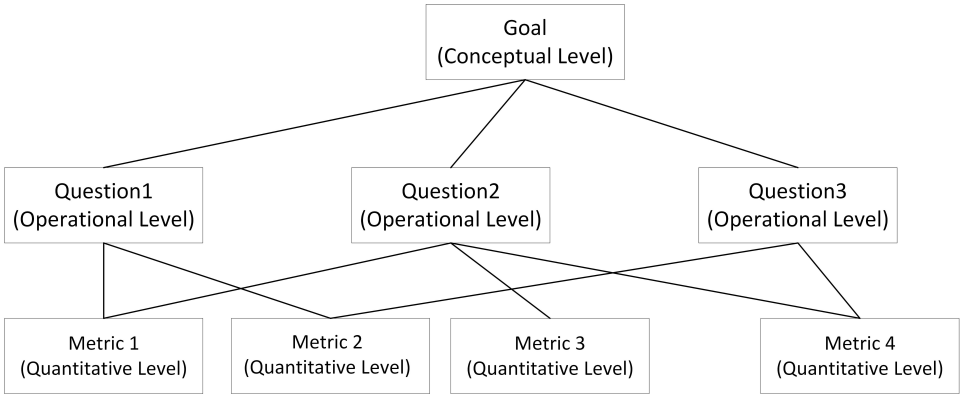

Fig. 5: The Goal Question Metric (GQM) method considering three questions and four metrics [20]

The GQM method has been used in several projects such as NASA Goddard Space Flight Centre environment [21].

2.1.4 Robotic Safety Standards and Generic Standards for Risk Assessment and Control: There are four main relevant standards and one technical specification (TS) for risk assessment in human robot collaboration domain:

- ISO 12100:2010, Safety of machinery - General principles for design - Risk assessment and risk reduction [22]
- ISO 10218-1:2011, Robots and robotic devices – Safety requirements for industrial robots – Part 1: Robots [23]
- 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]

Based on standard ISO 12100:2010 [22], 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 Fig. 6.

Standard ISO 10218-1:2011 [23] 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. It shall be noted that new versions of ISO 10218-2 are under development [27].

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 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 [22], 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

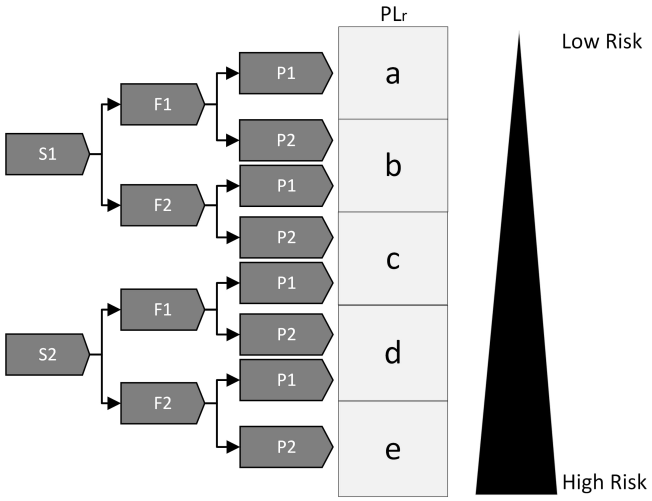

Fig. 6: Determining required performance level based on [26]

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 the risk analysis results

In traditional robot system installations it was not possible for humans 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.

2.2 Related Work

In [28], the authors provide a case study for safety analysis in aircraft ground handling services using STAMP (Systems Theoretic Accident Model and Process) causation model [29]. 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, this model neglects the possible contribution of workers to safety and the complexities that they face. 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 [30], 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 [31], the authors provide a case study for modeling heating, ventilation and air-conditioning (HVAC) systems using FRAM (Functional Resonance Analysis Method) [32]. 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 [33], 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 hazard 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 [34], 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 [35], 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.

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.

In the following sections, we explain the execution of the activities of our study including these steps. Section 4 (planning the study) includes step 1 and step 2. Section 5 (executing the study) includes step 3 and step 4. Finally, Section 6 (discussion on the results and their validity) includes step 5.

4 Planning the Study

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 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 of related safety standards.
2) M2: Percentage of covered typical human robot interaction failures.
3) M3: Percentage of extension 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 Fig. 7.

For the first research question, the metric can be defined as percentage of supported risk assessment steps of related safety standards. The rationale behind selecting this metric is that percentage of supported steps of safety standards shows the extents of support provided for safety standards. 100 percent support means all the steps of safety standards are supported and 0 percent support means that none of the steps can be supported by the framework.

For the second research question, we define the percentage of covered typical human robot interaction failures as the second metric. If all the typical human robot interaction failures are covered by the conceptualizations provided by the framework, it shows 100 percent effectiveness and lack of coverage for any common failure can decrease this percentage. The first defined metric can also support the second research question, because steps defined by safety standards are with the aim of increasing effectiveness of capturing essential information for assessing risk.

For the third research question, we define the percentage of extension on identified risk sources with respect to effects of AR and organizational changes as the third metric. Based on effects of AR and organizational changes, extensions are required to integrate their effects. The percentage of extension is used to illustrate the extent of the extension in comparison to the identified risk sources without considering effects of AR and organizational changes. This metric can also support second research question, because extension for integrating effects of AR and organizational changes increases effectiveness of capturing essential information in socio-technical robotic manufacturing (since it contains AR and organizational changes).

4.2 Selected Case

In this section, we describe an AR-equipped socio-technical system which we selected based on [36] and a taxonomy of typical failures in human robot collaboration proposed in [37]. The reason for selecting the case is that both organizational changes and technological changes are present in this case. Augmented reality which is a technological change is used as a human-machine interface and digitalization which is an organizational change is also present due to the human robot collaboration in diesel manufacturing.

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 [36] provides capabilities to improve safety of collaboration between human and robot in diesel manufacturing. AR-based UI can be implemented using projector-mirror setup. In this setup, as it is

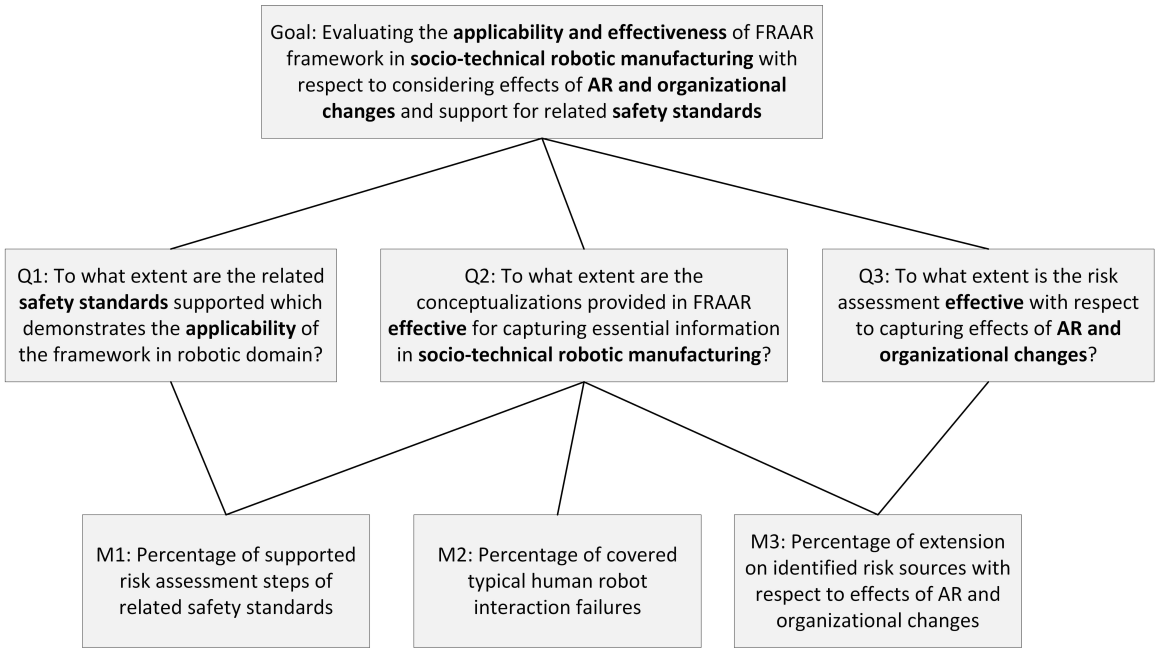

Fig. 7: Defined goal, questions and metrics using GQM method

shown in Fig. 8, the AR indications are shown on the table around the robot and human can collaborate with the robot using the AR indications.

Fig. 8: Robot and AR-based UI using projector-mirror [36]

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 [36] which is part of a real engine assembly task taken from a local company. It contains five sub-tasks which 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). One of these sub-tasks (sub-task 4) is collaborative and we have the focus on that. These sub-tasks. 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 [37] 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*.

4.3 Study Protocol

Based on [13], there are three types of data collection techniques: first degree, second degree and third degree. To make the paper self-contained, we recall the essence of these techniques. In the first degree technique, the researcher is in direct contact with the subjects collecting data in real time such as interview. In the second degree technique, the researcher collects data without interacting with the subjects such as observation. In the third degree technique, the researcher analyzes work artifacts such as 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. Data acquisition in this study means collecting information about the use case, possible scenarios, technical, organizational and human factors. For selecting the case containing augmented reality in a real context, we use [36] which describes an AR-equipped socio-technical system with its real-life context. We use the description provided in [36] to model the system by categorizing technical and socio entities, identifying their important aspects and their relations. We use the real engine assembly task described in [36] for formulating the scenarios and assumptions to analyze the system. In addition, we use technical details described in the related product websites to model technical entities and we use system and task description provided in [36] to model organization and human factors. Additional required information for analyzing the components' behavior is based on assumptions defined by the first author and reviewed by the second author. In addition, we collect data based on Goal Question Metric method (GQM) [20] which is a goal-oriented measurement technique as we explained in Section 2.1.3. 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. The required information is collected with the first author and reviewed by the second author.

5 Executing the Study

5.1 System Modeling

Based on the first step of the FRAAR framework explained in Subsection 2.1.2, in order to model the system, we need to identify the system entities (as we identified in Subsection 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 need to consider some aspects of simplification because for example, a robot arm consists of many different components, each of which can produce a fault. It is of course not possible to cover all the faults, but we consider the most dominant factors. We identify important aspects of the robot collaborating with human using the description provided in [36] and product technical specifications in [38] and [39]. For identifying human and organization important aspects, we use the extended modeling elements of FRAAR framework proposed in [11] and shown in Fig. 3 and Fig. 4.

- Important aspects of robot:
  - Control box: it contains a hardware and a software for receiving command from computing system and providing control commands for controlling the arm and gripper.
    - 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: it contains a hardware and a software for providing command for robot and for providing the required input for 3LCD projector using the received information from RGB-D sensor.
    - 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 in relation to the computing system hardware for conducting the computations.
  - 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 Fig. 9.

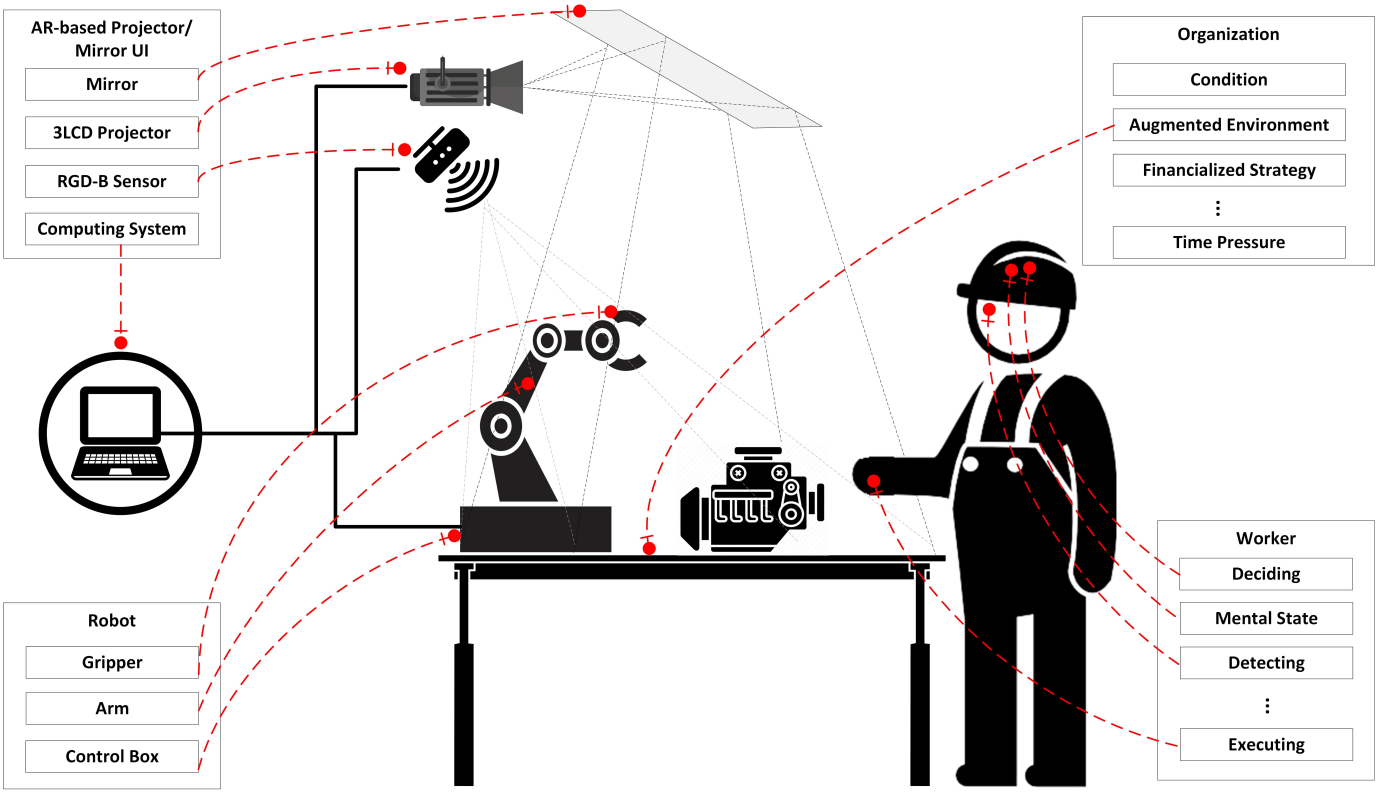

Fig. 9: Integration of human worker, AR-based projector-mirror UI, robot and organizational factors (adapted from [40] and [41])

In Fig. 10, we show how the considered AR-equipped socio-technical system is modeled using the extended modeling language of FRAAR framework. For modeling each entity, which is considered as a composite component, sub-components are determined based on the related important aspects. For example, for human worker, we defined nine important aspects, which are used for defining nine sub-components of the human worker composite component. Human worker contains four inputs. The inputs of each composite component are determined based on the required inputs of the sub-components. 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.

Based on the defined important aspects for robot, the robot composite component has five sub-components and one input coming from a computing system which contains the command 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 sub-components and one input which is connected to the input of the system containing the RGB-D data sensed by sensor, shown by RGB-D.

Based on the ten important aspects of diesel manufacturing organization, ten sub-components and two inputs are determined for organization composite component. One input is coming from mirror and the other input is connected to the input of the system. The input connected to the input of the system integrates influence 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.

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 2.1.2. We assume that the human and the robot collaborate to perform sub-task 4 explained in Subsection 4.2.

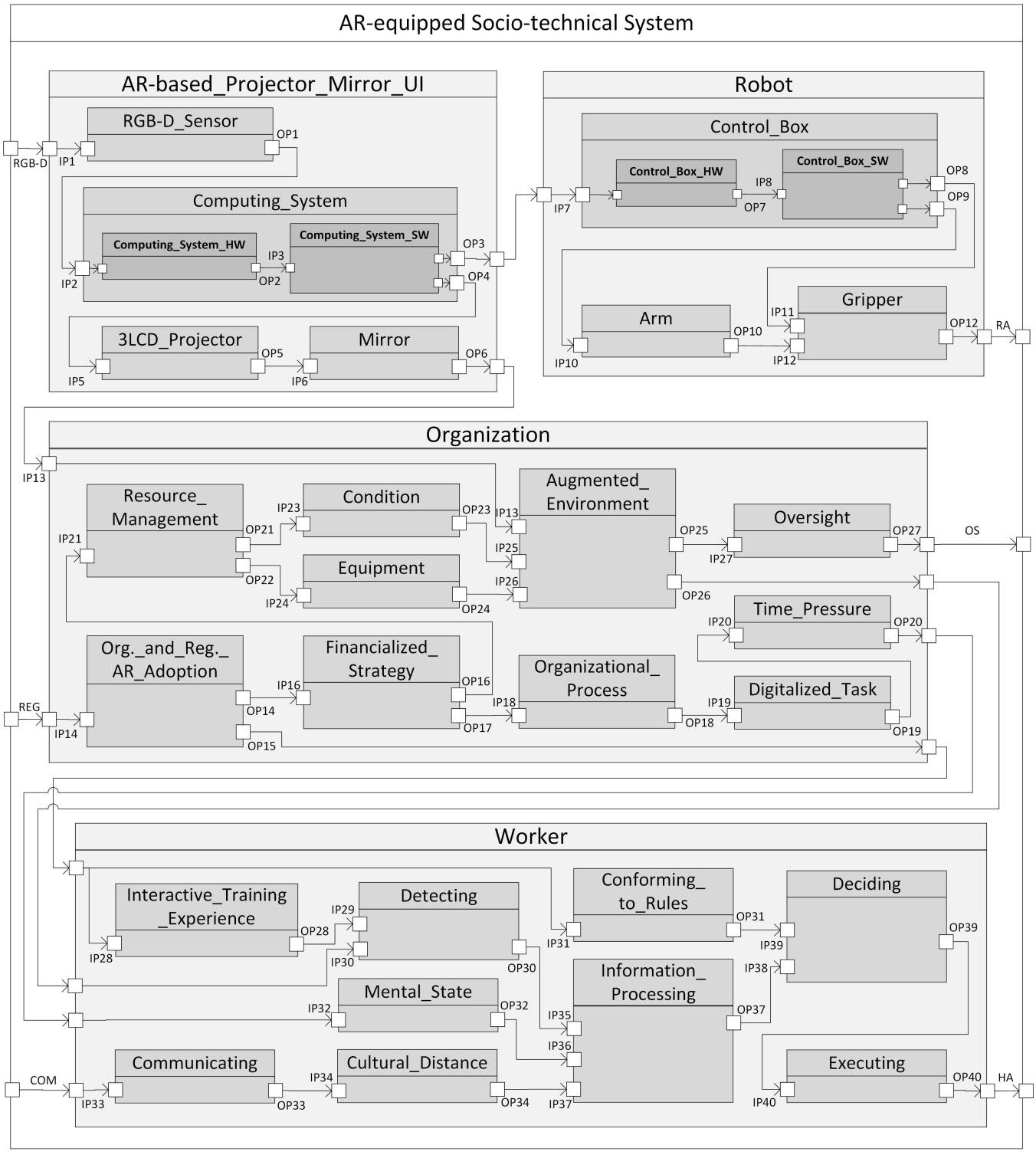

Fig. 10: Modeling of the AR-equipped socio-technical system

Information from accident reports or expert opinion can be used for modeling and analyzing system behavior. We assume 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 and failure in communication. For example, because of increasing power of financial actors, new strategies are assigned to increase production. This can lead to changes in 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. On the other hand, failure in communicating can lead to failure in cultural distance. Failure in mental state and failure in cultural distance both lead to 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 Fig. 11. In this scenario, financialized strategy and communicating behave as source and while there is no failure on their input, they produce valueSubtle failure on their output. Organizational process, digitalized task, time pressure, mental state, cultural distance, 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 there is failure in the provided financialized strategy and valueSubtle failure mode on IP34 means that there is failure in the communicating. ValueSubtle propagates to organizational process, digitalized task, time pressure, mental state, cultural distance, information processing, deciding and executing. The failure propagation is shown by blue color.

Interpretation of 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 OP34. ValueSubtle on OP32 is because of valueSubtle on OP20 and valueSubtle on OP34 is because of valueSubtle on OP33. 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. Failure on OP17 is because of failure in financialized strategy component and failure on OP33 is because of failure in communicating component.

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 and improper communicating. 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 2.1.4, 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 Fig. 6, required performance level is PLr = c, which is quit high.

In this case we define the following safety requirements:

- **Safety requirement 1**: Evaluation for financialized strategies shall be provided.
- **Safety requirement 2**: Evaluation for communicating competence 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 incorrect detecting, incorrect information processing, incorrect deciding and incorrect executing by the human worker (value failure mode).

Modeling failure behavior: The activated FPTC rules are underlined in Fig. 12. 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 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.

Interpretation of 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 in OP37 is because of valueSubtle on OP30 and it is because of valueSubtle on OP26. ValueCoarse on OP27 is because of valueSubtle on OP25. ValueCoarse on OP25 and OP26 is because of failure in augmented environment component.

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 may lead to sever injury since the

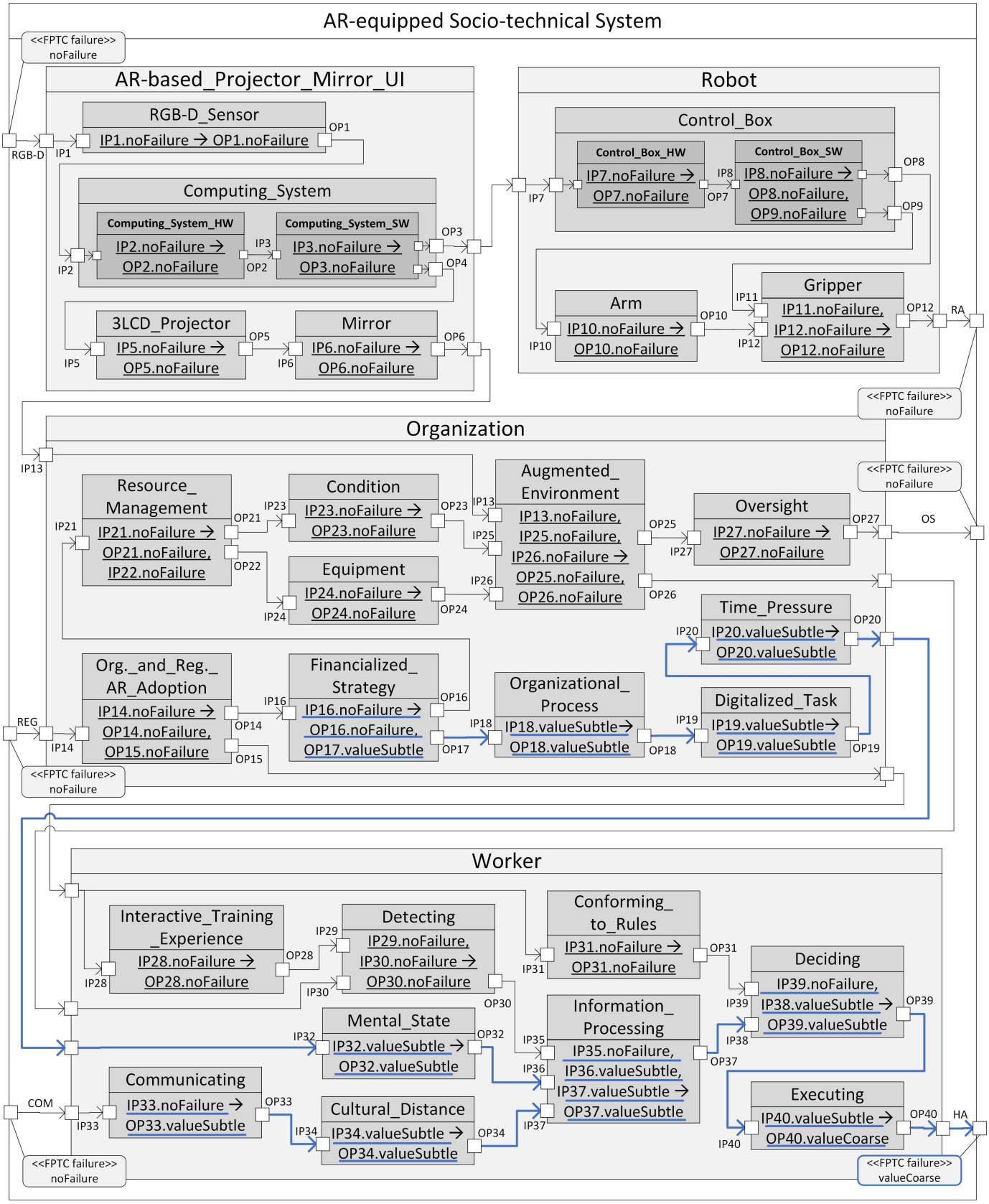

Fig. 11: Analyzing the AR-equipped socio-technical system (Scenario 1)

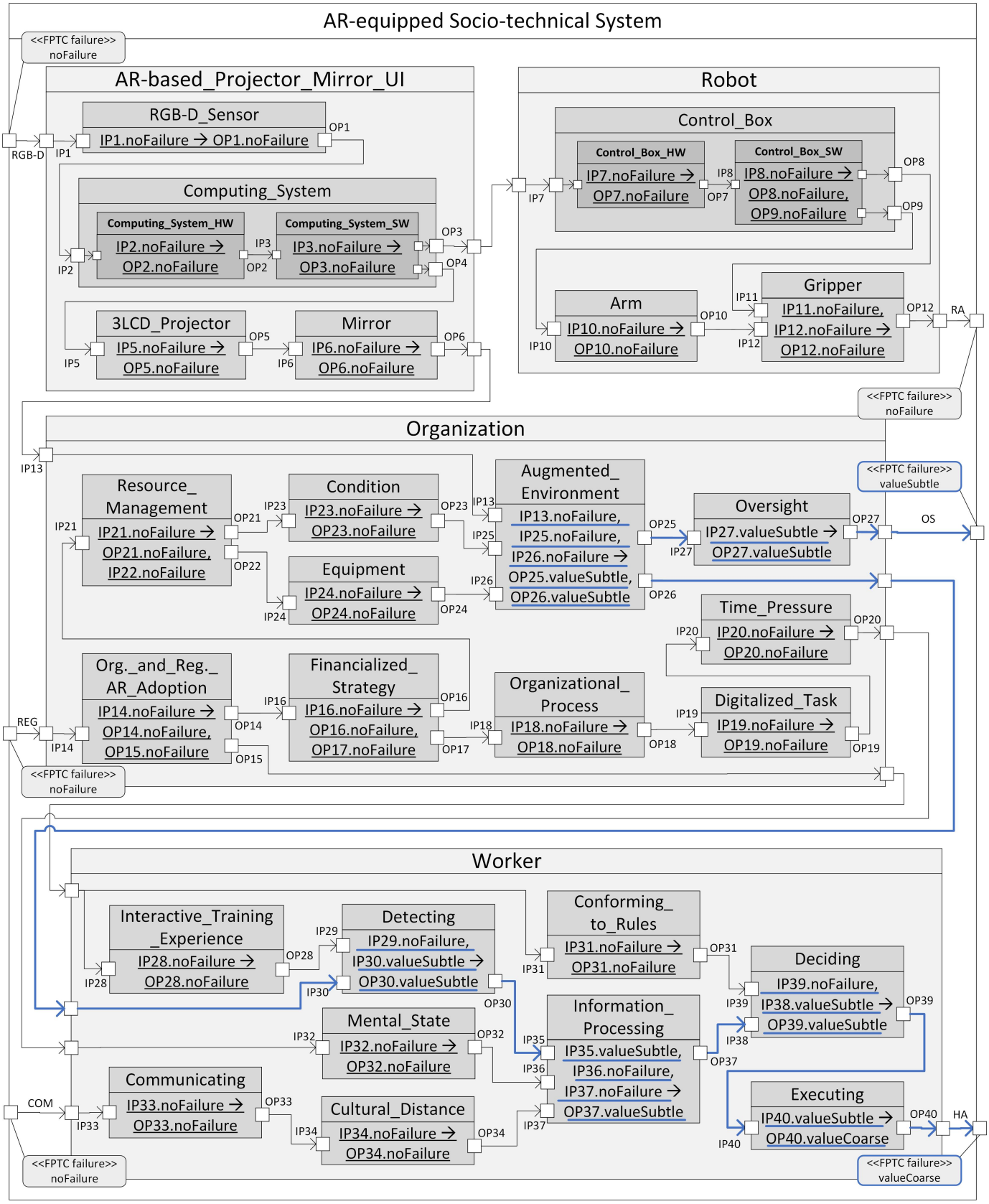

Fig. 12: Analyzing the AR-equipped socio-technical system (Scenario 2)

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 Fig. 6, 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 Fig13. 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.

Interpretation of 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 Fig. 6, 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 I, 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.

6 Discussion on the results and their validity

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.

6.1.1 Results for the First Research Question: In Section 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 activities based on robotic standards explained in Subsection 2.1.4 and we show different steps of FRAAR framework which support them in Table II.

As it is explained in Subsection 2.1.4, based on the 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 Fig. 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 5, can be helpful for *determining the limits containing the intended use and foreseeable misuse*. Thus, we can point out 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 5.2, we can estimate the risk of each hazard

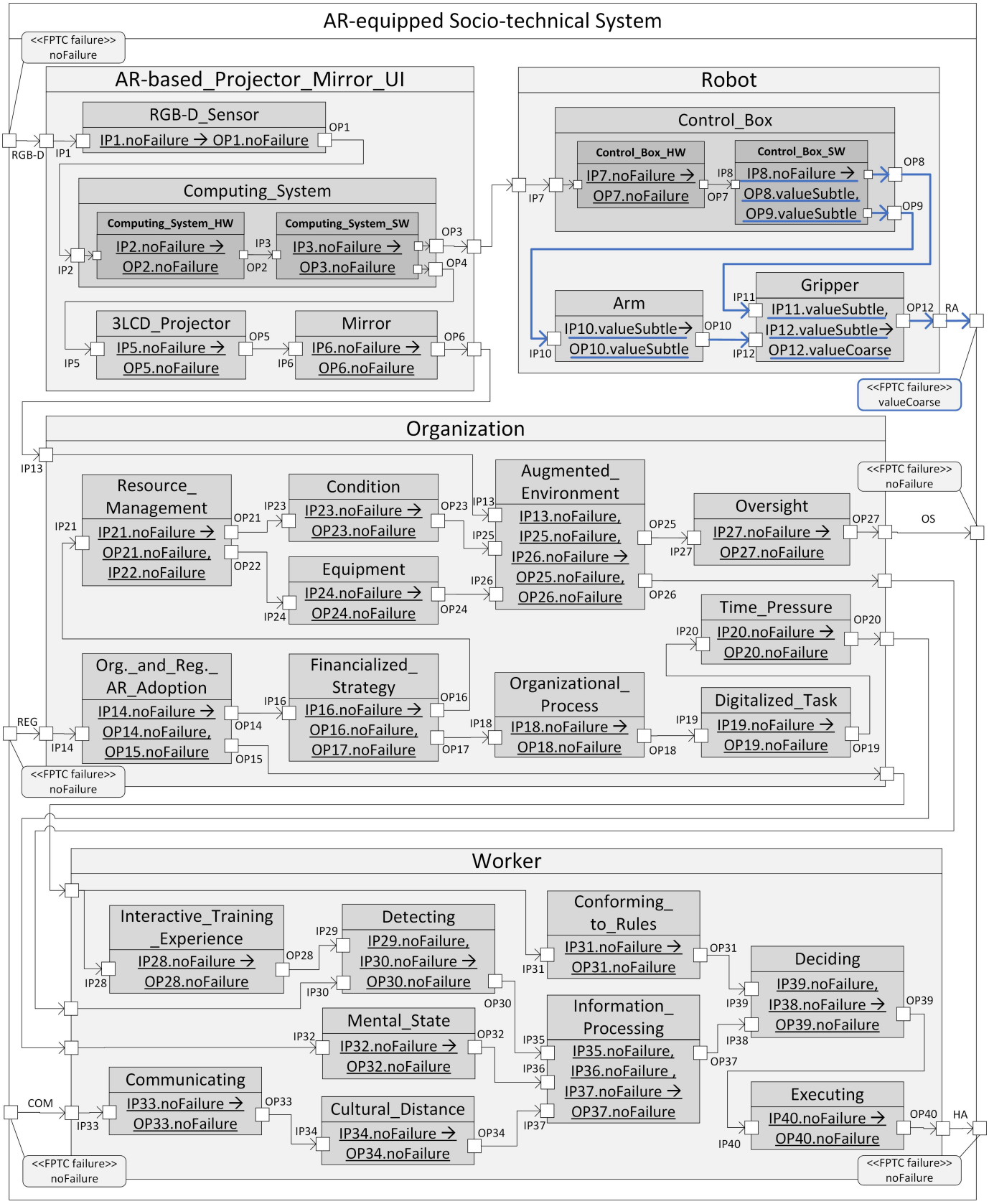

Fig. 13: Analyzing the AR-equipped socio-technical system (Scenario 3)

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
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

TABLE I: Identified list of dependability threats/risk sources

Risk assessment activity 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

TABLE II: Supported risk assessment activities based on robotic standards by FRAAR risk assessment activities

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 5.2.

As it is shown in Table II, 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.

6.1.2 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 [37], explained in Subsection 4.2. In Table III, it is shown how failures are covered by the available modeling elements/failure modes/failure behaviors used 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 2.1.2, *technical failures* can be modeled using *hardware/software components* and then failure behavior can be modeled by defining possible failure modes on 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 *connector*. For example, in modeling and analysis of our selected case in Section 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 [42], where we have previously proposed the extensions in relation to organizational factors. We also illustrated how we can use this component in Section 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 2.1.1.

Processing failures can be modeled by modeling a component failure behavior as *source* as explained in Subsection 2.1.2. 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 [37], *interaction failures* are failures due to uncertainties in interaction between human, environment and other agents. These failures can be modeled by *human/organization components* and their *connectors* 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 [43].

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*

Num	Typical human robot interaction failure	Available modeling element/failure mode/failure behaviors in FRAAR for modeling the failure
1	I. Technical failures	Hardware/software component
2	I.I. Software failures	Software component
3	I.I.I. Design failures	Equipment component
4	I.I.II. Communication failures	Connector
5	I.I.II.I. Incorrect data	Value failure mode
6	I.I.II.II. Bad timing	Early or late failure mode
7	I.I.II.III. Extra data	Commission failure mode
8	I.I.II.IV. Missing data	Omission failure mode
9	I.I.III. Processing failures	Source failure behavior
10	I.I.III.I. Missing events	Omission failure mode
11	I.I.III.II. Timing and ordering	Early or late failure mode
12	I.I.III.III. Abnormal terminations	Commission failure mode
13	I.I.III.IV. Incorrect logic	Value failure mode
14	I.II. Hardware failures	Hardware component
15	I.II.I. Effectors failures	Hardware component
16	I.II.II. Power failures	Hardware component
17	I.II.III. Control failures	Hardware component
18	I.II.IV. Sensors failures	Hardware component
19	II. Interaction failures	Human/organization component
20	II.I. Human errors	Human components
21	II.I.I. Mistakes	Selecting goal component
22	II.I.II. Slips	Acting component
23	II.I.III. Lapses	Information processing component
24	II.I.IV. Deliberate violations	Conforming to rules component
25	II.II. Environmental and Other agents failures	Environment unit component
26	II.II.I. Group-level judgment	Organizational climate component
27	II.II.II. Working environment	Environment unit component
28	II.II.III. Organizational flaws	Organization and regulation unit component
29	II.III. Social norm violations	-

TABLE III: Covered typical human robot interaction failures

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 extension to a great extent.

6.1.3 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 5.2, Table I. 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.

6.2 Discussion on the validity

As it is described in [13], the validity of a study discusses the trustworthiness of the results and the extent the results may be biased by subjective viewpoint of the researcher. In what follows, we discuss three aspects of validity: construct validity, internal and external validity.

6.2.1 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. In addition, we paid careful attention to critical comments provided by external reviewers on the previous versions of this paper (see the technical report published in [44] which is also included in the Ph.D Dissertation [45]).

6.2.2 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 extension with respect to effects of AR and organizational changes as three distinct metrics for measuring support for standards, the extent of applicability of the framework and effectiveness of risk assessment with respect to effects of AR and organizational changes. 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. It is essential to investigate the completeness of the modeling elements and modeling behavior. There are various challenges in this regard since there are unlimited possible scenarios and unlimited ways of modeling and analyzing. We focused on the most dominant factors which significantly influence safety and assumed some relevant scenarios. 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.

6.2.3 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 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.

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 the human-robot collaboration domain. Specifically, we considered the effects of the use of augmented reality as a new technology, the effects of new organizational changes and support for relevant safety standards. 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 effects of AR and organizational changes.

Based on the findings of our study, we demonstrated the applicability and effectiveness of our FRAAR framework in socio-technical robotic manufacturing with respect to considering effects of AR and organizational changes and support for related safety standards. One suggested research direction is to demonstrate the applicability and effectiveness of this framework in other domains considering related safety standards for the intended domain. Based on the limitations of our study (discussed in Subsection 6.2, other research directions can be defined such as defining systematic methods for generating scenarios, evaluating the completeness of the components and their relationships and providing quantitative analysis.

In addition, 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. Furthermore, 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.

References

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

[2] E. H. D. R. da Silva, A. C. Shinohara, E. P. de Lima, J. Angelis, C. G. Machado, Reviewing digital manufacturing concept in the industry 4.0 paradigm, Procedia CIRP 81 (2019) 240–245. doi:https://doi.org/10.1016/j.procir.2019.03.042.

[3] Associated Press in Berlin, Robot kills worker at volkswagen plant in germany (2015). URL https://www.theguardian.com/world/2015/jul/02/robot-kills-worker-at-volkswagen-plant-in-germany

[4] ISO 45001, Occupational health and safety management systems - Requirements with guidance for use (2018). URL 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). URL https://www.iso.org/news/ref2269.html

[6] J. Guiochet, Hazard analysis of human–robot interactions with HAZOP–UML, Safety science 84 (2016) 225–237. doi:https://doi.org/10.1016/j.ssci.2015.12.017.

[7] T. A. Kletz, Hazop—past and future, Reliability Engineering & System Safety 55 (3) (1997) 263–266. doi:https://doi.org/10.1016/S0951-8320(96)00100-7.

[8] G. Booch, J. Rumbaugh, I. Jackobson, UML: Unified Modeling Language (1997).
URL https://people.eecs.berkeley.edu/~brewer/cs169/lecture05-uml.pdf

[9] R. T. Stone, S. Pujari, A. Mumani, C. Fales, 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, SAGE Publications Sage CA: Los Angeles, CA, 2021, pp. 737–741. doi:https://doi.org/10.1177/10711813216510.

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

[11] S. Sheikh Bahaei, 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.

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

[13] P. Runeson, M. Höst, Guidelines for conducting and reporting case study research in software engineering, Empirical software engineering 14 (2) (2009) 131–164. doi:https://doi.org/10.1007/s10664-008-9102-8.

[14] A. Avizienis, J.-C. Laprie, B. Randell, C. Landwehr, Basic concepts and taxonomy of dependable and secure computing, IEEE transactions on dependable and secure computing 1 (1) (2004) 11–33. doi:https://doi.org/10.1109/TDSC.2004.2.

[15] D. J. Pumfrey, The principled design of computer system safety analyses., Ph.D. thesis, University of York (1999).
URL https://etheses.whiterose.ac.uk/9797/1/313770.pdf

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

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

[18] M. Wallace, Modular architectural representation and analysis of fault propagation and transformation, Electronic Notes in Theoretical Computer Science 141 (3) (2005) 53–71. doi:https://doi.org/10.1016/j.entcs.2005.02.051.

[19] A. Cicchetti, F. Ciccozzi, S. Mazzini, S. Puri, M. Panunzio, A. Zovi, 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, ACM, 2012, pp. 362–365. doi:https://doi.org/10.1145/2351676.2351748.

[20] V. R. Basili, Software modeling and measurement: the goal/question/metric paradigm, Tech. rep. (1992).
URL https://dl.acm.org/doi/10.5555/137076

[21] V. R. B. G. Caldiera, H. D. Rombach, The goal question metric approach, Encyclopedia of software engineering (1994) 528–532doi:https://doi.org/10.1002/0471028959.sof142.

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

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

[24] 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

[25] 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/

[26] 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

[27] ISO/FDIS 10218-2, URL: https://www.iso.org/standard/73934.html, accessed: 2024-01-14.

[28] D. Passenier, A. Sharpanskykh, R. J. de Boer, When to STAMP? A case study in aircraft ground handling services, Procedia Engineering 128 (2015) 35–43. doi:https://doi.org/10.1016/j.proeng.2015.11.502.

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

[30] X. Zhang, Y. Sun, Y. Zhang, 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 111 (2021) 102355.

doi:https://doi.org/10.1016/j.simpat.2021.102355.
[31] I. T. de Souza, A. C. Rosa, A. C. J. Evangelista, V. W. Tam, 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 320 (2021) 128895. doi:https://doi.org/10.1016/S0925-7535(03)00047-X.
[32] H. Erik, FRAM: the functional resonance analysis method: modelling complex socio-technical systems, CRC Press, 2017. doi:https://doi.org/10.1201/9781315255071.
[33] A. Hanna, K. Bengtsson, P.-L. Götvall, 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, IEEE, 2020, pp. 424–431. doi:https://doi.org/10.1109/ETFA46521.2020.9212127.
[34] R. Inam, K. Raizer, A. Hata, R. Souza, E. Forsman, E. Cao, 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, IEEE, 2018, pp. 743–751. doi:https://doi.org/10.1109/ETFA.2018.8502466.
[35] S. Sheikh Bahaei, B. Gallina, Technical report on risk assessment of safety-critical socio-technical systems: A systematic literature review, Tech. rep. (2022).
URL http://www.es.mdh.se/publications/6578-
[36] A. Hietanen, R. Pieters, M. Lanz, J. Latokartano, J.-K. Kämäräinen, AR-based interaction for human-robot collaborative manufacturing, Robotics and Computer-Integrated Manufacturing 63 (2020) 101891. doi:https://doi.org/10.1016/j.rcim.2019.101891.
[37] S. Honig, T. Oron-Gilad, Understanding and resolving failures in human-robot interaction: Literature review and model development, Frontiers in Psychology 9 (2018). doi:https://doi.org/10.3389/fpsyg.2018.00861.
[38] Universal robots, URL: https://www.universal-robots.com/cb3/, accessed: 2022-09-05.
[39] RG2-Grippe, URL: https://onrobot.com/en/products/rg2-gripper, accessed: 2022-09-05.
[40] Flaticon database of free icons, URL: https://www.flaticon.com/, accessed: 2022-09-05.
[41] Vecteezy resources of photography, videos and vector illustrations, URL: https://www.vecteezy.com/, accessed: 2022-09-05.
[42] S. Sheikh Bahaei, B. Gallina, K. Laumann, 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.
[43] S. Sheikh Bahaei, 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.
[44] S. Sheikh Bahaei, B. Gallina, Technical report on assessing risk of ar and organizational changes factors in socio-technical robotic manufacturing, Tech. rep. (2022).
URL http://www.es.mdu.se/publications/6579-
[45] S. Sheikh Bahaei, Organizational Changes-aware Safety-centered Risk Assessment in Augmented Reality-equipped Socio-technical Systems, Ph.D. thesis, Mälardalen University (2023).
URL http://www.es.mdu.se/publications/6663-