Elicitation and Documentation of Explainability Requirements in a Medical Information Systems Context

Christian Kücherer[a], Linda Gerasch[b], Denise Junger[c] and Oliver Burgert[d]
Faculty of Informatics, Reutlingen University, Reutlingen, Germany

Keywords: Requirements Engineering, Requirements Documentation, Explainability, Design Science, Medical Information System, Requirements for Explainable AI, XAI.

Abstract: **[Context and motivation]** Ongoing research indicates the importance of explainabilty as it provides the rationale for results and decision of information systems to users. Explainability must be considered and implemented in software at the early stage in requirements engineering (RE). For the completeness of software requirements specifications, the elicitation and documentation of explainability requirements is essential. **[Problem]** Although there are existing studies on explainability in RE, it is not clear yet, how to elicit and document such requirements in detail. Current software development projects miss a clear guidance, how explainability requirements should be specified. **[Solution Idea]** Through a review of literature, existing works for elicitation and documentation of explainability requirements are analyzed. Based on these findings, a template and additional guiding for capturing explainability requirements is developed. Following a design science approach, the template is applied and improved in a research project of the medical information domain. **[Contribution]** The overview of related work presents the current state of research for the documentation of explainability requirements. The template and additional guiding can be used in other information system context for RE elicitation and documentation. The application of the template and the elicitation guidance in a real world case show the refinement and an improved completeness of existing requirements.

1 INTRODUCTION

The specification of requirements for information systems (IS) is essential to address user needs. Current IS becoming more and more semi-intelligent by the utilization of Artificial Intelligence (AI). They support users in decision-making or take over even making decisions for them. However, often the reasons why a system made a certain decision stays unclear for users. This might lead to a loss of user's trust in the IS. This effect has been investigated by Chazette et al. (Chazette et al., 2019) who argue that explainability requirements for IS capture user's expectation to understand the systems' decisions. Such explanations should be based on transparency, traceability and trust and belong to the category of non-functional requirements (NFR). In their study Chazette et al. provide the example of a commuter who uses a navigation system during transportation. Due to road works or an accident, the most efficient route becomes blocked. The navigation systems will show an alternative route, unknown to the user. S/he might be unsure if this route is a good choice. The user expects the navigation app to provide an understandable reason for the chosen alternative. This expectation should be captured in an *explainability requirement*. Thus, it is essential for analysts and engineers to gain an overview of practice and research of explainability requirements.

IS must provide the rationale behind decisions or complex results to assure users' confidence into the system. This characteristic must be considered in the software requirement specification as explainability requirements. However, the principles of how to capture explainability requirements are still vague yet. Analysts need practical support to decide relevant aspects of explainability for users. In particular a guidance is necessary to elicit and document such explainability requirements. However, there is not yet much experience reported, to specify explainability requirements in practice. Within this paper, we develop a template and guidance to supports analysts in capturing and documenting explainability requirements and

[a] https://orcid.org/0000-0001-5608-482X
[b] https://orcid.org/ 0009-0002-9757-4782
[c] https://orcid.org/0000-0002-7895-3210
[d] https://orcid.org/0000-0001-7118-4730

Kücherer, C., Gerasch, L., Junger, D. and Burgert, O.
Elicitation and Documentation of Explainability Requirements in a Medical Information Systems Context.
DOI: 10.5220/0013470600003929

In *Proceedings of the 27th International Conference on Enterprise Information Systems (ICEIS 2025) - Volume 2*, pages 83-94
ISBN: 978-989-758-749-8; ISSN: 2184-4992

the rationales behind. By the application of the template and guidance in a research project of the medical domain, we show that it improves existing explianbility requirements and shapes understanding of the analyst to find new ones.

2 RELATED WORK

Method. With a lightweight literature search we want to explore the state of the art *to elicit and document explainability requirements*. This question is described as Research Question 1 (RQ) in Tab. 2. We scope the search to the sources IEEE[1] and ACM [2] as the major computer science literature databases with a technical and practical research focus. Springer link[3] was selected due to their large collection of RE related literature. Science Direct[4] completes this view as it covers many studies of applied requirements. We included peer reviewed papers after publication date of 2015 to April 2023, and such that contain a description of *explainability* as requirement, The search string is deviated from RQ1 and is `(Explainability AND "Requirements Engineering" AND (Elicitation OR Documentation OR visualization))`

Tab. 1 shows the classification of the studies to requirements *elicitation* and *documentation* or both as one aspect of the answer to RQ1 with a color scheme. Elicitation techniques are much more in focus (13 studies) than documentation (2 studies). Only four studies cover both RE activities.

Duque Anton et al. document a cross domain survey in the area of Explainable Artificial Intelligence (XAI) (Duque Anton et al., 2022). XAI focus on techniques that make the reasoning of decisions from AI models understandable to humans. In particular, when such decisions are interacting with non-experts, explanations have to be understandable. The study provides a literature survey of XAI, emphasizing the "ways of explaining decisions and reasons for explaining decisions". The authors build upon the explainability goals of Linardatos et al.(Linardatos et al., 2020) that are *mode of explanation*, *data types*, *purpose of explanation*, *specificity of explanation*, and *audience of explanation*. The investigated studies are classified into three main aspects of XAI: (1) solutions to enhance explainability of algorithms, (2) requirements, challenges and solutions of explainability in specific domains, and (3) papers that address the meaning of XAI for improved outcomes, stakeholder or user acceptance. In this study we use the characterization of the purpose and the audience of the explanation for the design of our template.

Köhl et al. postulate that often exact details are unclear when explainability is demanded (Köhl et al., 2019). Thus, they investigate the elicitation, specification, and verification of explainablity as a NFR. They propose a target- and context-aware sentence template to capture explainability requirements. The resulting requirements are related to softgoal interdependency graphs, showing relations and dependencies of different NFRs. Although the study provide a very broad view to the current state of explainability requirements, it neither propose a template to capture explainability requirements nor contribute a case study. We follow their principal ideas of sentence templates and target groups.

Into the same direction argue (Chazette and Schneider, 2020), who conducted a survey with 107 end users to understand their expectations of embedded explanations. The authors recommend strategies for the elicitation and analysis of explainability NFRs based on users' experiences with technical systems, and expectations of personal values and preferences. Explainability is impacted by domain aspects, cultural and corporate values, and project constraints. In general, explinability requirements are closely related with usability and User Centered Design (UCD). In a further paper of the same author (Chazette et al., 2021) it is argued that the RE community lacks guidance on how to consider explainability in a software design phase. We provide an approach for this gap to guide explaniability requirements elicitation by additional hints for analysts.

Habiba et al. (Habiba et al., 2022) recently proposed a framework for the user-centric approach to define explainability requirements for XAI. They argue that explanations help to support transparency, and thus increase stakeholder trust of a system. As a consequence they propose a five step framework for the definition of explainability requirements. The framework is in line with other requirements processes, such as the UCD-process and the foundations of the International Requirements Engineering Board (IREB) (Pohl, 2016). After (1) identifying stakeholders, the relevant (2) requirements are identified, (3) a common vocabulary is specified as a glossary, (4) requirements are validated and the finally (5) requirements are classified. Step (3) is very interesting, as it focus on the reasons for explainability and its impact of single explanations on the to-be system. In step (5), the authors reflect on the adequateness of each explainability requirement to individual stakeholders.

[1]https://ieeexplore.ieee.org/Xplore/home.jsp
[2]https://dl.acm.org
[3]https://link.springer.com
[4]https://www.sciencedirect.com

Table 1: Overview of related work.

Citation	Topic	*XAI*	*RE*	*Elicitation*	*Document.*
(Afzal et al., 2021)	Data-Debugging with visual explanations	✓	–	✓	✓
(Ahmad et al., 2023)	Mapping study of RE for AI	✓	✓	–	✓
(Alonso et al., 2020)	Explanations for fuzzy decision trees	✓	–	–	✓
(Balasubramaniam et al., 2023)	Transparency & explainability of AI syst.	✓	–	✓	✓
(Brunotte et al., 2022)	Research road-map for explainability in RE	✓	✓	–	–
(Calegari and Sabbatini, 2022)	Trustworthy AI	✓	–	✓	–
(Cepeda Zapata et al., 2022)	Adoption of AI in medical software	✓	–	✓	✓
(Chazette et al., 2021)	Lack of guidance to consider explainability	–	✓	✓	✓
(Chazette and Schneider, 2020)	Elicitation and analysis of explainability NFRs	–	✓	✓	✓
(Cirqueira et al., 2020)	Scenarios for Elicitation of XAI	✓	–	✓	✓
(Duque Anton et al., 2022)	Survey of Explainability in AI-solutions	✓	✓	✓	–
(Habiba et al., 2022)	UCD based framework for explainability in XAI	✓	✓	✓	✓
(Habibullah et al., 2022)	Exploration of NFRs for machine learning	✓	–	✓	–
(Langer et al., 2021)	Stakeholder perspective on XAI	✓	–	✓	✓
(Köhl et al., 2019)	Study on explainablity as a NFR	–	✓	✓	✓
(Vermeire et al., 2021)	Method to provide explainability in context of XAI	✓	✓	✓	✓
Total	16 studies	12	8	13	12

We follow this process in a similar way.

The work of Vermeire et al. (Vermeire et al., 2021) investigates on methods to provide explainability in context of XAI. The authors argue that there is a gap between stakeholder needs and explanation methods for machine-learning models. Thus, they present a methodology to support data scientists in providing explainability to stakeholders. In their methodology, they map explainability methods for data with user requirements in an analysis step leading to a recommendation and explanation of user goals. In the framework the authors propose to use (1) explanation method properties (such as different types of explanations, process properties), (2) typical stakeholder explainability needs, and (3) a questionnaire to reveal users' needs. We follow their idea of a questionnaire to investigate user needs and guidance for analysts.

The road-map of Brunotte et al. documents the results of a 2022 workshop in explainability engineering (Brunotte et al., 2022). They summarized three fundamental RQs in the area of explainability: (1) defining and measuring explainability, (2) stakeholders and contexts, and (3) goals and desiderata. Regarding the definition of explainability (1) they conclude that there is still no common definition of what it means to be explainable and how exactly explainability differs from other related concepts. We follow the direction of the latter two RQs through a practical application of the elicitation and documentation of explainability requirements in the medical domain and base our template on goals and desiderata.

Habibullah et al. investigate the role of NFR in machine learning (ML) (Habibullah et al., 2022). Based on a literature review, they cluster the NFR categories based on their shared characteristics. Explainability (altough not yet part of the ISO 25010 standard) builds an own category with Interpretability, Justifiability and Transparency. In context of a ML system, explainability targets the ML algorithm, the ML model, and its results. Hence, such requirements must be defined precisely.

The study of Langer et al. (Langer et al., 2021) introduces the concept of relating explainability approaches to stakeholders' desiderata. With this concept, the authors describe the stakeholders' expectation on explainability in context of XAI, documented in an explainability-model. They emphasize that the explanation process and the desiderata must be motivated and guided. The authors propose to consider the concepts of explainability with its information, ability to understand stakeholders, satisfying their desiderata and the given context. They argue that desiderata satisfaction requires the consideration of human-computer interaction principles in form of scenario techniques to document and understand the stakeholders' context. This work influenced our guidance and enhanced explainability requirements in sentence templates in combination with scenarios.

Alonso et al. describe a method to create explanations for decision trees (Alonso et al., 2020). They

argue that users require both, graphical visualizations and textual explanations. Hence, the authors introduce a tool that automatically creates such explanations learned from data. We use their ideas of an *explainer* that generates graphical and textual explanations associated with decision trees users are faced in the system. The tools uses a decision tree structure to visualize the *why* aspect to users. Based on if/then rules, they visualize the decisions made.

Cirqueira et al. propose a user-centric perspective on human-AI interfaces (Cirqueira et al., 2020). The authors propose to specify user requirements in the given context and problem domain with a scenario-based analysis. The scenarios are build upon stakeholder goals, captured by interviews. The user's sociological and company context are included in the scenarios that tell a consequence of events and steps of actors to reach a goal. Based on this information user interfaces (UIs) and system functions can be defined. We follow their proposal of scenarios to describe explainability in a given stakeholder context.

Balasubramaniam et al. examined several organizations to understand their motivation for explainability (Balasubramaniam et al., 2023). They argue that transparency is essential for users trust. Transparency improves understandability of a system and thus its decisions. The authors propose to document aspects of decisions in a table based form to support the understandability of the decisions made. We follow their model of explainability components, consisting of Addressees, Aspects, Contexts, and Explainers.

Cepeda Zapata et al. estimate transparency as a main challenge in medical AI applications (Cepeda Zapata et al., 2022). To ensure this, they use a comprehensive documentation to create traceability. This increases the probability for a correct implementation of user needs. The documentation contributes to the systems' transparency, provides aspects of explainability, documented as written natural language and in visual form. This leads to improved trust between patients and medical practitioners.

Ahmad et al. performed a encompassing systematic mapping study in the area of RE for AI (Ahmad et al., 2023). Beside other aspects, they also investigate the role of explainability, which is evident in literature after 2019. The authors state a need for future empirical research on RE for XAI. Their results indicate that graphical representations on the behavior of the systems requirements and relations.

The highlighted literature validates the need for elicitation and documentation of explainabaility requirements in various contexts. However, there is no study that shows the elicitation and documentation of explainability requirements in practice.

Figure 1: Used design science approach (Wieringa, 2014).

3 METHODS

This section describes the methods used in this study.

3.1 Design Science Research

The methodical approach of this study is based on design science research for information systems as reported by Wieringa (Wieringa, 2014) and Hevner et al. (Hevner et al., 2004). Design Science is a systematical and iterative approach to design and investigate artifacts in a given context. An artifact is something artificial created by someone for a practical purpose. Artifacts may interact with its context to improve a situation. The artifacts context might be anything, such as the design, development, and use of information systems, where humans are part of this context.

Fig.1 shows the phases of the design science research used in this article. In the *problem investigation* we analyze the related work to gain an overview of current approaches to elicit and document explainability requirements. Based on a synthesis of the approaches, we define a template and guidance as *treatment design*. This template and the according guidance were then applied to an ongoing research project as *validation*. The project focuses on the development of a medical situation recognition system, which gathers sensor data from ongoing surgeries in an intelligent operating room (Junger et al., 2022; Junger et al., 2024). Based on this data, the surgical phases are estimated to provide these context information to other systems, such as the OR-Pad, which can display phase-related clinical information during surgery. The first application of the template and guidance has revealed deficits that were improved in an iteration of the *treatment design* and *validation*. Discussions about the artifacts shaped the scope of the sketched approach. Comments indicated that the process to elicit and document explainability requirements must be simplified. In consequence, artifacts were rearranged and unnecessary details were discarded. The *implementation* in the above mentioned project has been done in form of revised requirements and additional NFRs in the SRS. The access of additional research material is explained in Sec. 6.

The scope of this study is defined by its **Research Questions** (RQs) in Tab. 2. RQ1 is answered by the synthesis of the literature in related work and RQ2 by

Table 2: Research Questions and explanation.

ID RQ	Explanation
RQ1: How are explainability requirements elicited and documented in projects?	Shows the used methods to elicit explainability requirements. Various documentation techniques such as templates, structured sentences, models are investigated to understand use of explainability in RE.
RQ2: What is the design for a template to document explainability requirements?	Proposes a template and guidance to elicit and document explainability requirements in the application in the selected research project.

the design of an explainability requirements template.

3.2 Development of the Template

Based on the reviewed literature, we developed a template to document explainability requirements. The template is supported by tables 3 and 4, as well as a one-page instruction (provided in the Appendix) that delivers definitions, key aspects and examples of possible instances of each template field. Thus, the template and the guidance assist requirements engineers in the elicitation and documentation of explainabulity requirements. We applied the template and guidance in *validation* and improved the template based on evaluation results.

We followed following principles for the design of the template: The choice of documentation is crucial for effectively conveying information to the individuals involved in a project. Ebert et al. (Ebert, 2022) advocate the use of a sentence template to establish a clear structure, making it easier to maintain consistency and ensuring that individual elements remain testable. These attributes are particularly valuable for developers in the system implementation process, enhancing readability and comprehension. Schön et al. (Schön et al., 2017) recommend employing scenarios when engaging with users or stakeholders, as they offer better understandability and an overview of the system's situation. Potts et al. (Potts et al., 1994) highlight the positive aspect that scenarios allow tailoring the level of detail based on project needs, providing a specific representation. Scenarios are also beneficial for discussing the system and conveying a general understanding through specific situations.

3.3 Requirements Analysis and Application in Project

We performed an analysis of existing requirements in an ongoing research project that aims at the development of a situation recognition system (SRS) for surgical procedures. Using various sensors, the system automatically infers the current activity in an operating room (OR), based on rules. The SRS provides this information to different context-aware system (CAS). The context information of the current activity should be provided to the actor without any further interaction. The goal is to obtain a flexible and medical intervention-independent assessment of the current situation in order to accurately convey context and details (Junger et al., 2022). In a test-scenario the SRS is connected to an operating room (OR) tablet called OR-Pad. The OR-Pad provides planning- and in-situ information before, during, and after surgery, and reflects relevant information at the current time based on the surgical phase of the operation (Ryniak et al., 2022). The main goal of the combination of SRS and OR-Pad is to support the medical actors in the OR. Failure to understand the information received can lead to a lack of trust in the system by the actor, as there is uncertainty to the extent of correct information provided for the given situation. This can lead to disruptions in the OR, which can result in actor hesitation. These disruptions should be reduced with the help of explainability. Thus, the project demands correctness of the system, represented by transparency to provide a comprehensive understanding of the system. Trust can be gained from understanding what is happening in the system (Chazette et al., 2019). Explainability leads to better overall quality of software systems (Chazette et al., 2022). Within the project, the software architecture and a prototype for this adaptive SRS is developed. During the project, a set of requirements (Junger et al., 2022) have been elicited and documented as natural language with some use case diagrams in a system specification document (based on word). We apply the proposed template for explainability requirements in this context and measure its effectiveness with the evaluation shown in Fig. 2.

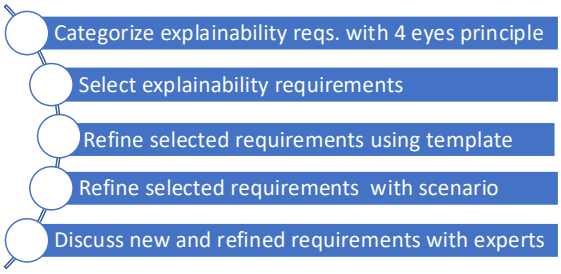

Figure 2: Evaluation Process.

We analyzed the existing quality requirements for the SRS (Junger et al., 2022) and the OR-Pad (Ryniak et al., 2022) to understand the extent they describe a reason for the systems behavior or results.

For that, we used the question: *Does the function need to share information?* and *Does the information give an answer to the why?*. If both questions can be answered with yes, we classified the requirements as *candidate explainability req*. Because the intention with explainability is that the software should enable the user to make the system's decisions comprehensible by providing a clear rationale for the decision, the procedure and the result.

4 RESULTS

The results of each design science phase are given in the following subsections.

4.1 Problem Investigation - Literature Synthesis

The problem leading to RQ1 how explainability requirements are elicited and documented is answered based on literature. The related work revealed various approaches for eliciting and documenting explainability requirements. The key findings emphasize two essential methods for documentation: employing a sentence template (Balasubramaniam et al., 2023), illustrated in Fig. 3, and utilizing a scenario (Cirqueira et al., 2020). These approaches play a crucial role in simplifying information extraction and fostering a deeper understanding of the requirements. The literature also shows pertinent questions based on Langer et al. (Langer et al., 2021) for refining information identification individually and ensuring completeness in the required elements. Furthermore, a finite set of example applications for individual elements was generated, with room for additional supplements.

An interview, following a questionnaire, was done by the authors with one developer of the SRS and ORPad. First of all, the meaning of explainability was clarified. Then, the current state within the systems was discussed regarding explanability. The aim was to identify to what extend explainability was already addressed in the requirements and system architecture. Furthermore, the relevance of explainability on both systems was elaborated. Afterwards, it was discussed how the system could be optimized regarding explainability and which requirements were therefore needed to be specified in general.

As **<type of addressee>** I want to receive explanations about an **<aspect>** from an **<explainer>** in a **<context>**.

Figure 3: Sentence template for explainability requirement elicitation. Variables are filled by insertion procedures.

4.2 Treatment Design: Design of the Template

Based on the findings during the problem investigation, we followed a template-based and scenario-based approach. The template was inspired from Younas et al.'s (Younas et al., 2017) guideline for non-functional requirement elicitation in agile methods.

Fig. 4 illustrates the agile NFR elicitation process tailored specifically for Explainability, encompassing both the sentence template and the scenario. Information retrieval elements are essential in both methods, with the exception that the explanatory unit is explicitly required only in the sentence template. In some cases analysts might have an existing set of requirements, which need completion wrt. explainability. Thus, in the first step analysts might decide to revise existing requirements or add new requirements following elicitation activities. To aid the elicitation, questions and usage examples are listed for each element, providing guidance and a comprehensive overview. For this purpose, Table 3 summarizes the key questions for all elements of the sentence template. Table 4 identifies the individual implementation steps for the creation of a scenario for explainability requirements. Tables in the supplementary material were defined for the individual elements (<type of address>, <aspect>, <context>, <explainer>) of the sentence template with possible settings [5] and the associated identification questions for the possible settings, which have been added for a better overview. Following this systematic process of collecting information step by step, a robust set of explainability requirements can be effectively elicited.

In the following, the elements of the template are described: **Type of addressee** describes the person to be addressed, which is crucial for the explanation. The target audience allows the identification of the knowledge needed and the existing prior knowledge to create an appropriate representation that ensures understanding of the environment and context of the stakeholder group. The **Aspect** enriches the determination of the explanatory purpose with information and specifies when and how the explanatory system should be used to clarify structural aspects. The **Context** determine the representation of an explanation with information. It specifies the type of explanation and the data to be transmitted in different representations. The **Explainer** describes the unit that reflects an explanation of the system and helps to determine

[5] These are just examples to make it easier to find the right way to fill in the sentence template element. If the possibilities do not fit into the use case to be defined, the corresponding question should help to define the possibility.

Table 3: Key-questions to identify the elements of explainability requirements.

Element	Key questions
<type of addressee>	Who are relevant target groups and what are their specific characteristics? What are the relevant desiderata in a particular context and are they being met? Is the information really useful/helpful for the target group?
<aspect>	Has the target group acquired a sufficient level and the right understanding to judge whether wishes are fulfilled and to facilitate their fulfillment? Is the information really useful for the target group?
<context>	Are there contextual influences that hinder or promote the satisfaction of the desiderata by the explanatory process? Is the explanatory approach able to provide the right type and amount of information in an acceptable presentation? Is the information really useful/helpful for the target group?
<explainer>	Is the explanation approach able to pass on the information through a suitable source that promotes the understanding of the target group? Is the information really useful/helpful for the target group?

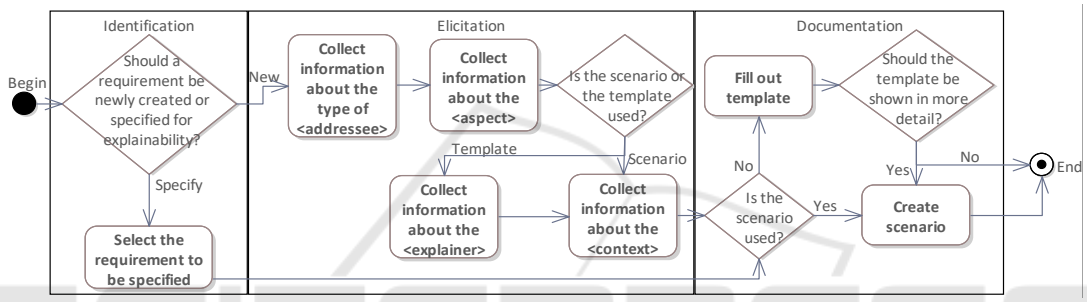

Figure 4: Specification of *issue identification* for explainability requirements elicitation and documentation according to Younas (Younas et al., 2017) agile NFR elicitation process.

that unit. The **Scenario** defines different schemes to elicit requirements.

4.3 Validation - Application in Project

We categorized a total of 67 NFRs in Junger et al. (Junger et al., 2022; Ryniak et al., 2022). Overall, 5 NFR were identified as explainability requirements using the 4-eyes principle during the review. One requirement was also identified that requires an additional explainability requirement to be added. We see that the definition of a non-functional explainability requirement led to the existing of functional requirements with explainability related content.

4.3.1 Explainability Application Example

To apply the documentataion of expl. req, both, the sentence template and the scenario has instanciated to show a practical example. This is done using the same requirement for both paths to present a good comparison. The original elicited requirement N11 is shown in fig 5. An explainability requirement is now to be created as an example on the basis of requirement N11. This is to be done in connection with the

```
N11:  The SRS must catch errors and log
meaningful error messages in case of a
misbehavior or sensor data that cannot
be processed.
```

Figure 5: Original requirement N11 (Junger et al., 2022).

created concept in order to test its usability. The error message is chosen as an exemplary representation because it can cause a lack of understanding on the part of the users if the representation of the error message is not suitably explained. Operators do not understand why an error occurs at that exact moment. The question of why raises that Explainability can help. This requirement is already raised, but based on the review it is determined by mutual agreement under the 4 eyes principle that this is an Explainability requirement and therefore needs to be enriched with Explainability. The information that is presented in the information gathering has been created based on interview one and two. The OR-pad represents the clinical- and the debugging context. The clinical context is the user interface of the currently existing OR-Pad, which is characterized by a simple design that can be quickly understood and is only used as a support when this is of interest to the operator. The debugging context has

Table 4: Process steps to create a scenario for explainability requirements.

Element	Description
Identify <type of addressee>	Representation to gain understanding of the environment and context of the stakeholder group.
Identification of the <aspect>	Diagram to identify the goals and sub-goals of the stakeholder group and the cognitive tasks that are performed in their environment, sorted in a chronological order.
Identify the <context>	Map to identify features of existing tools used by stakeholders.
Create scenario	Mapping to identify the scenarios that represent the processes and cognitive tasks of the stakeholder group, derived from the stakeholder group's attitudes, goals and tools.
Use scenario to identify requirements	Requirements generation through identification using interviews, prototypes, surveys and anthropomorphism studies

As a *surgeon*, I want to receive explanations of an *occurring error in a misbehavior or unprocessable sensor data within the SRS* from a *visual representation of a user interface associated with the SRS* in an *intuitive and understandable framework*.

Figure 6: Revised explainability req. N11 with template.

become of importance during interview three and can be an additional interface that runs concurrently with the clinical interface to reflect additional information and log a record of the entire operation. Based on the sentence template, a new structuring emerges. N11 is now to be reviewed in the clinical context, leading to a revised requirement 6. Unlike the sentence template, the scenario is a visual representation, whereby the level of detail must be selected independently depending on the project situation and the variant shown in 7 is only an example.

4.3.2 Requirements Evaluation Based on an Expert Interview

The elicited requirement on the way of the sentence template and the scenario is evaluated and verified. The expert initially receives an explanation of the concept, followed by the presentation of requirements without justification to ensure a neutral assessment. The expert then provides feedback, and additional questions are posed for more detailed insights. Analyzing the sentence template, it was generally viewed positively for being intuitive and logical, with the use of variables considered helpful. Using this template during requirements analysis would guarantee a structured and standardized way to define and document requirements. . In contrast to the previous requirements, improvements in the specificity and clarity of the new requirements were noted, but challenges in creating requirements spontaneously were highlighted instead of extending an existing requirement. Furthermore, concerns were raised about the extensive length. Regarding the scenario, presented as an activity diagram, it received positive evaluation for providing further information in an easily understandable form. Nevertheless, concerns were raised regarding the higher effort needed to prepare such scenarios. The expert suggested incorporating sketches in addition, particularly for communication with stakeholders that need further information on the requirement. For example, clinical stakeholders could benefit from a detailed, visualized requirement to understand the context and relevance of the depicted Explainability requirement. - Overall, different stakeholder groups will prefer specific representation forms, such as sentence templates, flowcharts, and activity diagrams. The choice between textual descriptions and mockups depends on the audience: clinicians favoring mockups and developers relying on textual descriptions. Thus, a combination of sentence templates and scenarios, with scenarios refining templates, was recommended, i.e. using the template for the technical specification and scenarios in addition for the communication with other stakeholders. The combined use of mockups and scenarios was deemed effective, emphasizing the importance of textual descriptions. Storyboards, combining visual scenes and descriptions, were suggested for effective communication, brainstorming, and verification of requirements.

4.4 Implementation - Evaluation of the Template

The focus of the implementation phase is to apply and evaluate the approach developed in *treatment design* in a real world case. As the case we use the *situation recognition system*-development project as explained in section 3.3. A technical expert in the role of an analyst applied the template and guidance to the SRS (Junger et al., 2022) and OR.pad (Ryniak et al., 2022) project in March 2024. Through an 0:45 min introduction to the idea of explainability requirements by the 1st author, the analyst gained a basic under-

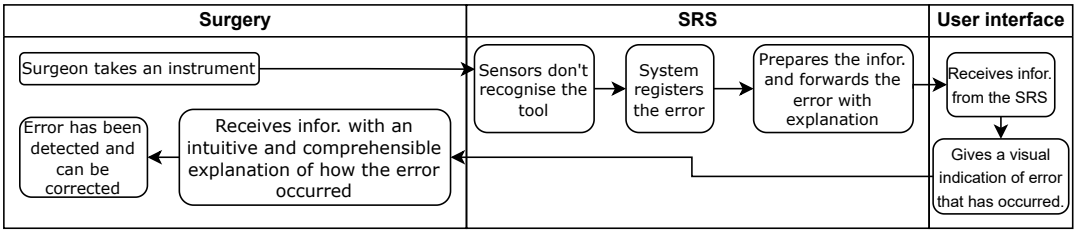

Figure 7: Revised explainability requirement N11 based on scenario.

standing of explainability and the assisitive template and guiding aids. The process was also discussed, depicted in the following.

Step 1: Identify Possible Explainability Requirements. Based on these instructions, the analyst took the list of available requirements (SRS & OR-Pad), provided as supplementary material in the analyst's former article [6]. The analyst furthermore was provided with a first assessment of the requirements regarding their potential to address explainability, consisting of a total of 16 requirements identified using the 4-eyes principle of the first two authors. The analyst had inspected each requirement against the definition for explainability requirements and criteria for explainability, as provided in section 1. A total of 18 requirements were identified as candidates. Each requirement was rated and it was depicted if (1) the req. is already adressing explainability to a certain extend, (2) a req. need to be used to derive a new expl. req, (3) the req. does not absolutely need explainability and need to be evaluated with stakeholders, and (4) the req. does not address or need explainability due to any resaon. As a result, four of the 18 requirements were neglected, as they describe *goals* or *visions* that have been already covered by other non-functional requirements, marked in orange. Nevertheless the rejected goals and vision led to two yet missing requirements, not identified as relevant by the first check of possible expl. req., resulting in 16 req. (red). Four requirements were clearly identified as explainability requirements that need to be specified (green), four requirements depicted as unsure if and for which stakeholder explainability would be beneficial (yellow), and the remaining four requirements were stated to be extendable to an explainability requirement. The resulting list is available as supplementary material.

Step 2: Application of Template and Guidance. The explainability template and the guidance aids presented in section 4.2 were applied to the 14 identified explainability requirements candidates. . According to the categorization depicted in the previous section, the analyst has 4 non-functional requirements specified and for further 10 additional non-functional requirements derived. The analyst used the sentence template and guidance to assess step-by-step each requirement, following the process in Fig. 4. The analyst use both, the sentence template in Fig. 3, the example in Fig. 6 as well as the scenario in Fig. 7 for reference. During the editing of the requirements, the analyst also created four scenarios to see the usefulness. However, the analyst was focusing on the sentence template as it provides enough information for the further use of the specification and complements the existing requirements analysis the best. Overall, the analyst referred to Table 3 and 4 to complete and detail the specification.

Step 3: Evaluation. The expert rated her experiences during the application of the template and gave qualitative feedback to the questions (eq_n) in Table 5.

Regarding *eq1*, the expert stated that a clear definition of explainability and how such requirements can be identified, is necessary and helpful. For each requirement in the SRS and OR-Pad, the expert considered if and how explainability would be beneficial. The expert noticed that some requirements contained more then one characteristic of the IS, leading to additional explainability requirements. Only few existing requirements were found to refer directly to explainability as this aspect was not in focus yet.

Regarding *eq2*, the initial discussion of the template's application to obtain a clear definition of explainability was very helpful. The use of Fig. 3 and the examples in the supplemental material (Tab.2 Type Of Addressee and Tab.4 Aspect) provided helpful guidance, combined with the definitions of the elements of the sentence template from the text. The use of the template was self-explaining and no further support was necessary. Although Fig. 4 was useful, the definition of the scenario was not the experts' first choice. Table 3 provided helpful information for the definitions of the elements in the template. Especially

[6]https://www.tandfonline.com/doi/suppl/10.1080/21681163.2021.2004446, and for the OR.Pad, available at https://link.springer.com/article/10.1007/s11548-022-02787-w#Sec19.

Table 5: Evaluation questions for implementation of treatment design.

Quest.	Description
eq1	What criteria had been used to identify explainability requirement candidates?
eq2	Which tools or methods supported the elicitation and documentation of explainability req.?
eq3	What difficulties in understanding have arisen?
eq4	What has been improved and how?
eq5	How clear and understandable was the template used?
eq6	How much time did the application take?
eq7	How do you rate the overall benefit of the approach?

in case of vague requirements that might be related to explainability, the expert could consider details and deepen possible explainability aspects.

Regarding *eq3*, there were three difficulties: (1) The expert was unsure what type of requirements should be considered to address explainability appropriately and decided to outsource explainability requirements into NFRs. (2) A compact explanation would have made the process more efficient. We addressed this issue by a one pager that is part of the supplementary material. (3) The workflow for defining new requirements or specifying existing ones is not shown. This led to confusion about the process. Furthermore, the criteria for choosing between template and scenario was not clear. The expert focused on the sentence template, rated it as more adequate for the purpose of the SRS and OR-Pad in its current form. Scenarios were only created as additional examples to support discussions with clinical partners.

Regarding *eq4* indicating possible improvements, the expert found the approach overall very useful, especially, but not limited to rule-based systems as the SRS. Once the expert realized the impact of explainability, the expert could highlight the system behavior and the basis of the estimations more transparent for users. Additional scenarios improve the understanding of the system functioning for some user groups, such as clinicians.

Regarding *eq5*, the clarity of the template, the expert found the template sufficiently clear, comprehensible and an appropriate scope. There were advantages through the training that has taken place by the definition of the first explainability requirements. Scenarios should only be used occasionally for discussion purposes as the creation was perceived as time-consuming but helpful to broaden the understanding of the system.

Regarding *eq6*, Table 6 shows the time required for the application in each step. In total, the time spent on the content alone was 4 hours and 30 minutes, not including notes and debriefing.

Regarding *eq7*, the expert rated the overall benefit of this approach for improving explainability requirements as positive. By deriving and specifying non-functional requirements from existing requirements, a more precise and clearer representation of the explicit explainability requirements is achieved. This process enables the targeted addressing of sub-aspects that improve the explainability of the system. Overall, the approach to improve explainability requirements appears promising: It offers a structured and systematic method to increase completeness of the requirements specification.

Table 6: Time consumption of each step from the analyst.

Step	activity	Duration
step 1	introduction	45 mins
step 1	determination of explainability	50 mins
step 2	familiarization with template	130 mins
step 2	documentation with template	30 mins
step 2	create example scenarios	60 mins
Total		5:15 hrs:min

5 THREATS TO VALIDITY

We follow the classification of Wohlin et al. (Wohlin et al., 2012). **Construct validity** considers whether the study measures what it claims. The study is based on the bachelor's thesis of the 2nd author, supervised by 1th author. The related work was not gained through a systematical approach, however elements of a systematic mapping study were used, such as reviews of the first and second author, and the results were discussed in an open university talk. The search term was validated by two authors. **Internal validity** determines the extent of conclusions and the bias of the study. The application of the template was limited to one research project only, with a limited size of requirements. The results were produced by two authors with a review of a third author. The evaluation was done in two meetings, including constructive feedback, where the template and guidance has been improved. **External validity** describes the generalization of the study results to other situations. The template and guidance is not domain specific and can be applied in other domains. The application in one project only limits the generality of the results.

6 CONCLUSION

This paper presents a four-phase design science research approach to explore how explainability requirements can be documented and elicited. In the related work 18 articles have been reviewed and analyzed, revealing the importance of detailed documentation and visual diagrams to cater to different stakeholder preferences. A sentence template for documenting these requirements is proposed, which is combined with a scenario method for elicitation. Expert interviews confirm the usefulness of this combined method. Additionally, the process includes refined elicitation techniques and provides a one-page information sheet to capture details. This research is specifically evaluated in the context of a surgery assistance system, showing it improves the completeness and detail of the SRS. Future work could focus on integrating this workflow into existing RE processes and creating a tutorial video for practitioners.

APPENDIX

Additional research data is available as supplementary material at EU's Zenodo portal[7].

It contains (i) The existing requirements and our classification for explainability requirements. (ii) Additional tables and figures and guidance information. (iii) The artifacts of the first design science iteration. (iv) The created scenarios.

REFERENCES

Afzal, S., Chaudhary, A., Gupta, N., Patel, H., Spina, C., and Wang, D. (2021). Data-debugging through interactive visual explanations. In *Trends a. Appl. in Knowl. Disc. a. Data Mining, Delhi, India*, pages 133–142. Springer.

Ahmad, K., Abdelrazek, M., Arora, C., Bano, M., and Grundy, J. (2023). Requirements engineering for artificial intelligence systems: A systematic mapping study. *Information a. Software Techn.*

Alonso, J. M., Ducange, P., Pecori, R., and Vilas, R. (2020). Building explanations for fuzzy decision trees with the expliclas software. In *2020 IEEE Int. Conf. on Fuzzy Systems*, pages 1–8. IEEE.

Balasubramaniam, N., Kauppinen, M., Rannisto, A., Hiekkanen, K., and Kujala, S. (2023). Transparency and explainability of ai systems: From ethical guidelines to requirements. *Information and Software Technology*, 159:107197.

Brunotte, W., Chazette, L., Klös, V., and Speith, T. (2022). Quo vadis, explainability?–a research roadmap for explainability engineering. In *International Working Conference on Requirements Engineering: Foundation for Software Quality*, pages 26–32. Springer.

Calegari, R. and Sabbatini, F. (2022). The psyke technology for trustworthy artificial intelligence. In *Int. Conf. of the Italian Assoc. for AI*, pages 3–16. Springer.

Cepeda Zapata, K. A., Ward, T., Loughran, R., and McCaffery, F. (2022). Challenges associated with the adoption of artificial intelligence in medical device software. In *Irish Conf. on AI and Cognitive Science*, pages 163–174. Springer.

Chazette, L., Brunotte, W., and Speith, T. (2021). Exploring explainability: A definition, a model, and a knowledge catalogue. In *2021 IEEE 29th International Requirements Engineering Conference (RE)*, pages 197–208.

Chazette, L., Karras, O., and Schneider, K. (2019). Do end-users want explanations? analyzing the role of explainability as an emerging aspect of non-functional requirements. In *27th RE Conf.*, pages 223–233. IEEE.

Chazette, L., Klös, V., Herzog, F., and Schneider, K. (2022). Requirements on explanations: A quality framework for explainability. In *30th Int.RE Conf.*, pages 140–152.

Chazette, L. and Schneider, K. (2020). Explainability as a non-functional requirement: challenges and recommendations. *Requirements Engineering*, 25:493–514.

Cirqueira, D., Nedbal, D., Helfert, M., and Bezbradica, M. (2020). Scenario-based requirements elicitation for user-centric explainable ai: A case in fraud detection. In *Int. cross-domain conf. for ML a. knowledge extract.*, pages 321–341. Springer.

Duque Anton, S. D., Schneider, D., and Schotten, H. D. (2022). On explainability in ai-solutions: A cross-domain survey. In *Computer Safety, Reliability, and Security. SAFECOMP Workshops*, pages 235–246, Cham. Springer Int. Pub.

Ebert, C. (2022). *Systematisches Requirements Engineering:Anforderungen ermitteln, dokumentieren, analysieren u.verwalten*. dpunkt verlag.

Habiba, U.-E., Bogner, J., and Wagner, S. (2022). Can requirements engineering support explainable artificial intelligence? towards a user-centric approach for explainability requirements. In *2022 IEEE 30th International Requirements Engineering Conference Workshops (REW)*, pages 162–165. IEEE.

Habibullah, K. M., Gay, G., and Horkoff, J. (2022). Non-functional requirements for machine learning: An exploration of system scope and interest. In *Proceedings of the 1st Workshop on Software Engineering for Responsible AI*, pages 29–36.

Hevner, A., Ram, S., March, S., and Park, J. (2004). Design science in information systems research. *MIS Quarterly*, 28:75–105.

Junger, D., Hirt, B., and Burgert, O. (2022). Concept and basic framework prototype for a flexible and intervention-independent situation recognition system in the or. *Computer Methods in Biomechanics and*

[7]https://zenodo.org/doi/10.5281/zenodo.10854274

Biomedical Engineering: Imaging & Visualization, 10(3):283–288.

Junger, D., Kücherer, C., Hirt, B., and Burgert, O. (2024). Transferable situation recognition system for scenario-independent context-aware surgical assistance systems: A proof of concept. *Int J CARS., lecture presentation at CARS 2024*.

Köhl, M. A., Baum, K., Langer, M., Oster, D., Speith, T., and Bohlender, D. (2019). Explainability as a non-functional requirement. In *2019 IEEE 27th Int.RE Conf.*, pages 363–368.

Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., and Baum, K. (2021). What do we want from explainable artificial intelligence (xai)?–a stakeholder perspective on xai and a conceptual model guiding interdisciplinary xai research. *Artificial Intelligence*, 296:103473.

Linardatos, P., Papastefanopoulos, V., and Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. *Entropy*, 23(1):18.

Pohl, K. (2016). *Requirements engineering fundamentals: a study guide for the certified professional for requirements engineering exam-foundation level-IREB compliant*. Rocky Nook.

Potts, C., Takahashi, K., and Anton, A. (1994). Inquiry-based requirements analysis. *IEEE Software*, 11(2):21–32.

Ryniak, C., Frommer, S., Junger, D., Lohmann, S., Stadelmaier, M., Schmutz, P., Stenzl, A., Hirt, B., and Burgert, O. (2022). A high-fidelity prototype of a sterile information system for the perioperative area: OR-pad. *ICARS Journ.*, 18.

Schön, E.-M., Thomaschewski, J., and Escalona, M. J. (2017). Agile requirements engineering: A systematic literature review. *Comp. stand. & interf.*, 49:79–91.

Vermeire, T., Laugel, T., Renard, X., Martens, D., and Detyniecki, M. (2021). How to choose an explainability method? towards a methodical implementation of xai in practice. In *Joint Europ. Conf.on Machine Learning and Knowledge Discovery*, pages 521–533. Springer.

Wieringa, R. J. (2014). *Design Science Methodology for Information Systems and Software Engineering*. Springer Berlin Heidelberg.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in software engineering*. Springer Sc.&Business Med.

Younas, M., Jawawi, D., Ghani, I., and Kazmi, R. (2017). Non-functional requirements elicitation guideline for agile methods. *J of Telec., Electr. a. Comp. Eng.*, 9:137–142.