PERSONALIZED EXPLAINABILITY REQUIREMENTS ANALYSIS FRAMEWORK FOR AI-ENABLED SYSTEMS

Jia Kai Quah[1], Yin Kia Chiam[2*], Nor Ashikin Md Sari[3]

[1,2]Department of Software Engineering, Faculty of Computer Science and Information Technology, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

[3]Department of Medicine, Faculty of Medicine, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

Email: kobekai1997@gmail.com[1], yinkia@um.edu.my[2*] (Corresponding Author), ashikin77@um.edu.my[3]

ABSTRACT

Artificial Intelligence (AI) has evolved into an indispensable technology that assists humans in making better decisions through predictive analysis and personalized recommendations in numerous sectors. However, complex machine learning (ML) models become less transparent and may recommend incorrect decisions, which leads to a loss of confidence and trust. Consequently, explainability is considered a key requirement of AI-enabled systems. Recent studies focus on implementing explainable AI (XAI) techniques to improve the transparency and trustworthiness of ML models. However, analyzing the explainability requirements of different stakeholders, especially non-technical stakeholders for AI-enabled systems, remains challenging. It lacks a comprehensive and personalized requirements analysis process that investigates the risk impact of outcomes produced by ML models and analyzes diverse stakeholder needs for explanations. This research proposes a framework with a requirement analysis that includes four key stages: (1) domain analysis, (2) stakeholder analysis, (3) explainability analysis, and (4) translation and prioritization, to analyze the personalized explainability needs of four types of stakeholders (i.e., development team, subject matter experts, decision makers and affected users) for AI-enabled systems. As demonstrated by the case study, it is feasible to apply the proposed framework to analyze diverse stakeholders' needs and define personalized explainability requirements for AI-enabled systems effectively.

Keywords: Requirements analysis; Explainable AI; User-centric; User story; Personalization.

1.0 INTRODUCTION

Artificial Intelligence (AI) is an umbrella term that includes machine learning (ML) or deep learning (DL) components that implement algorithms mimicking learning and problem-solving [1]. Based on Rzepka & Berger [2], AI-enabled systems are defined as systems that require a certain level of intelligence to carry out problem-solving, knowledge representation, reasoning, planning, learning, perceiving (including computer vision), acting (robotics), natural language processing, communicating, and decision making.

Explainability is identified as a key quality requirement of AI-enabled systems [3-5], and the requirements should encompass anticipated functionalities and explanations of the predictions, clarifying both its capacities and limitations [5]. The number of studies addressing transparency and explainability requirements has increased. Despite this increase, there is a lack of studies focused on practical methodologies for defining the explainability requirements to achieve the expected level of transparency of AI-enabled systems.

To obtain good explainability, it is highly related to the quality of explanations, and it should be presented in an appropriate format or combination of visualization techniques. The explanations that are incorporated with the predictions require careful consideration of the targeted audience, their educational background, AI knowledge, and cognitive limitations so that the principle of user-centric design can be achieved. Although the needs and goals of explainability are highly promising, the development of explainability does not necessarily need to be implemented in every AI-enabled system, as more resources have to be allocated. Therefore, the reason and timeliness of providing explanation in AI-enabled systems should be clarified precisely and systematically in practice [6].

The need for explainability arose due to the increasing use of complex ML and DL algorithms that are difficult to interpret. Even the AI developers who created them have difficulty explaining them well, which makes it more difficult to comprehend how they arrived at their prediction outcomes, eventually leading to a negative impact on decision-making by the users. When an AI-enabled system commits critical errors, it is difficult to identify the root cause and hard to determine who is responsible for the error, as it lacks traceability. Besides that, AI models may perpetuate biases in the data they were trained on, leading to unfair or discriminatory outcomes for certain groups of people. Therefore, the lack of interpretability and transparency of the models raised concerns by external stakeholders about the accountability and trustworthiness of the predictions, particularly in high-stakes domains such as healthcare [7-9], finance [10], and autonomous vehicles [3]. Most importantly, the XAI researchers raised the awareness that the black box models should explain themselves with supporting explanations rather than just providing the outcomes and making autonomous decisions [11, 12].

The emergence of explainability and the development of AI-enabled systems aim to address the growing need for transparency and trustworthiness of ML models. Even though the ML models have high prediction power, they have still been reported that they made incorrect decisions and lost the confidence and trust of humans. Meanwhile, the AI solutions generated by traditional AI-enabled systems are extremely difficult for lay users as they do not completely understand the technical terms and seldom utilize the prediction [13]. Although there are interpretable techniques incorporated to provide explanations that can illustrate the internal decision-making processes, an understandable explanation without considering various stakeholders' needs remains challenging [14]. Most stakeholders are not familiar with the terminologies, and explanations are given by the LIME [15] and Shapley values [16], they are not practically utilized by the non-technical stakeholders.

The current XAI techniques focus on generating explanations that follow only a One-Size-Fits-All approach, which undoubtedly opposes the principle of user-centric [17], and tends to overlook the inherent diversity among stakeholders and their specific explainability needs. For instance, the developers could require technical explanations of the overall behaviours of the models, but the non-technical users do not necessarily need to understand the underlying algorithms and inner workings of the models. Meanwhile, designing explanations that are genuinely comprehensible to stakeholders with varying levels of technical expertise and cognitive knowledge is difficult to capture [18]. While Wang et al. [19] studied the relationship between human/rational reasoning processes and the selection of XAI techniques, their design leans towards user-centric XAI. However, despite these efforts, justifying the effectiveness of explanations and establishing trustworthy or authoritative provenance remains challenging, as highlighted in the study.

While defining explainability as a non-functional requirement (NFR), the integration of explainability into AI-enabled systems continues to pose a formidable challenge. Chazette et al. [20] mentioned several challenges related to explainability during the requirement engineering, including the difficulty of extracting the existing knowledge, translating abstract notions, and miscommunication between the developers and various stakeholders. In short, there is still a lack of a comprehensive approach that addresses the impact of incorporating personalized explainability needs from various stakeholders during the requirement engineering phase.

Apart from the challenges of implementing explainability in AI-enabled systems, the current discourse lacks a formalized exploration of explainability goals and guidelines for determining when to utilize and how to identify the appropriateness of implementing explainability in AI-enabled systems. The challenge lies in the complicated nature of the requirement analysis process, which involves numerous navigating factors, including the inherent complexity of decision characteristics such as outcome criticality, time sensitivity, decision complexity, as well as diverse stakeholder perspectives and trade-offs between explainability and the other quality attributes [21].

In conclusion, the adoption of explainability in AI-enabled systems remains a complex task, requiring a holistic approach that addresses communication barriers, incorporates diverse stakeholder perspectives, and justifies the effectiveness of the explanations given. Overcoming these challenges during the beginning of the development cycle is crucial to ensure the implementation of transparency and trustworthiness in AI-enabled systems and to meet various users' expectations. At the same time, explainability requirements play a crucial role in guiding the design and development phases, including the selection of XAI techniques, focuses, and expectations of targeted groups.

This study proposes a personalized explainability requirements analysis framework for AI-enabled systems to address the specific explainability needs and goals of four main types of stakeholders (i.e., development team, subject matter experts, decision makers, and affected users). There are four main stages proposed in this framework to analyze the personalized explainability needs of various stakeholders, including (1) domain analysis, (2) stakeholder analysis, (3) explainability analysis, and (4) translation and prioritization. The step(s) in each stage illustrates how the requirement analysis is performed to obtain the user-centric explainability needs and goals from various stakeholders. A case study was conducted using an AI-enabled heart disease prediction system in the healthcare domain. Stakeholders represented the four types of stakeholders were invited to participate in this case study to evaluate the applicability and effectiveness of the proposed framework.

The remainder of this paper is organised as follows: Section 2 discusses work related to this research. Section 3 describes the proposed personalized explainability requirements analysis framework. Section 4 presents the context, data collection, data analysis, and results of a case study conducted to evaluate the proposed framework. Section 5 discusses the findings, limitations, and threats to validity of the case study. Section 5 presents a conclusion and discusses future research.

2.0 RELATED WORKS

This section reviews research studies that have adopted different approaches to elicit and analyze user requirements for explainability. Besides, there are question-based, scenario-based, human-centred, and user-story approaches.

There are several approaches that have been presented to analyze and discover explainability requirements. These approaches exhibit limitations that are addressed and resolved by the proposed framework.

Studies conducted by Chazette et al. [20, 22] aim to support software and requirements engineers in requirements analysis, design, and evaluation of explainable systems. The findings of a literature review and an interview study reported by Chazette et al. [20] show that requirements-related activities should be emphasized during the development of explainable systems. Stakeholder analysis is one of the important activities, and user-centric practices are most suitable for analysing explainability requirements. Chazette et al. [22] proposed four types of artifacts (i.e., a definition of explainability, a conceptual model, a knowledge catalogue, and a reference model) for explainable AI-enabled systems. Besides, the researchers also developed a quality framework for explainability to support the elicitation of explainability requirements. Software and requirements engineers can benefit from these artifacts to understand the explainability, define and refine high-level explainability requirements, make early design choices, identify appropriate methods, and metrics to evaluate the explainability requirements.

A DoReMi-approach [23] was developed in the study for the human-centered XAI design of a Clinical Decision Support System (CDSS). This approach includes three components: (1) Domain analysis, (2) Requirements elicitation and assessment, and (3) Multi-model Interaction design and evaluation. The domain analysis identifies what explanations and information should be provided by the CDSS. Next, use cases are used to elicit XAI requirements based on the explainability needs of clinicians. Lastly, UI design patterns are developed, which include an explanation of the design problem and UI design examples with information elements.

Liao et al. [24] place significant emphasis on adopting a question-driven user-centred approach, fostering collaboration between designers and AI engineers, and employing an end-to-end iterative design process. They also developed an XAI question bank that helps to understand the user needs for explainability [24]. This systematic procedure consists of four sequential steps: question elicitation, question analysis, mapping questions to solutions, and iterative design. Distinct stakeholders are engaged at each step, and a mapping guide is provided to facilitate the alignment of XAI techniques with user-generated questions. The process actively promotes stakeholders' collaboration, is user-centric, and continuous enhancement through user feedback. Nevertheless, the approach lacks standard guidelines on design and evaluation metrics for XAI. Despite there being a mapping guide on the XAI question bank, it may fall outside the scope of standard decision-support systems, and the methods for additional refinement to seek solutions are missing. In the framework proposed in this research, after the user stories are expressed by the stakeholders and the explanation prototypes are presented to the stakeholders, a refinement step is conducted to validate and refine the user stories to make sure their user stories and explanation prototypes are exactly what they expected from the end products.

Cirqueira et al. [25] propose a user-centric scenario-based requirements elicitation method that focuses on the socio and operational context of stakeholders. The study outlines that the fraud detection workflow in banking is complex, and fraud experts face challenges in understanding AI models, leading to reduced confidence in model-generated detection. In the context of XAI research for fraud detection, the study stresses the need to personalize explanations to meet the user requirements of domain experts, specifically fraud specialists in the realm of transaction fraud in banking. Fraud scenarios are created by identifying stakeholder settings, goals, tools, and capabilities. These scenarios are used to elicit XAI requirements in fraud detection information systems. It conducts problem-centred expert interviews with banking fraud experts, uncovering their operational environment, decision-making processes, and cognitive activities. However, the study highlights the limitation without consideration of other stakeholders, apart from the domain experts; the scenarios are not prepared for AI developers who are responsible for improving and developing the AI models. In contrast, our proposed solution comprehensively accounts for four key stakeholders, including the development team, subject matter experts, decision makers, and affected users. By incorporating these diverse perspectives into the framework, the requirement engineers could better understand the stakeholders' explainability needs and goals, yielding personalized, better, and more appropriate explanations that are provided to the targeted stakeholders.

In the context of explainability, Balasubramaniam et al. [4, 26] analyzed ethical guidelines from 16 organizations to understand the transparency and explainability guidelines. The researchers identified four components (i.e., addresses, aspects, contexts, and explainers) from the transparency guidelines that are important to represent individual explainability requirements for transparent and explainable AI systems. Balasubramaniam et al. [4, 26] modified and added more specific components tailored to the needs for explainability. A customized user story template is proposed using these four components: "As a <type of addressee>, I want to get explanation(s) on an <aspect> of a <system> from an <explainer> in a <context>". It is a concise and user-centric method for capturing explainability requirements from the viewpoint of end users or stakeholders [26]. They comply with this template that specifies the function of the user, the desired action, and the desired outcome, in order to assist the stakeholder in creating stories in a structured and consistent way.

In Ahmad et al. [27], a three-layered requirement engineering framework comprising six areas is proposed to elicit requirements for human-centred AI-enabled software systems. The six areas are user needs, model needs, data needs, feedback & control, explainability & trust, and error handling. The first layer is a reference model based on human-centred AI guidelines in requirements engineering. The second layer is a catalogue provided as a checklist to elicit requirements for AI-enabled systems. The last layer consists of a conceptual model to show the requirements visually.

These studies show that explainability is recognized as the main requirement for AI-enabled systems that are transparent, accountable, and trustworthy in supporting decision-making. User-centric or human-centred solutions are suitable to support the explainability requirements analysis of AI-enabled systems. However, these studies focus on specific domains or certain types of stakeholders only. For example, the DoReMi-approach [23] only focuses on the healthcare domain and one type of stakeholder (i.e., decision maker), which is clinicians, while the study by Cirqueira et al. [25] only captures the needs of subject matter experts in fraud detection. None of the studies considered AI risks during domain analysis. The level of explainability to be implemented could be linked to the identified potential AI risk outcomes within the AI-enabled systems. In our proposed framework, the degree of explainability can be personalized based on the discerned risk factors. A lower level of explainability could be implemented when negligible or low risk factors are identified, whereas more comprehensive and detailed explanations should be implemented if medium, high, or very high risk factors are distinguished.

In the explainability context, while the users are expressing their needs for explanations, the components in the user story templates mentioned earlier are insufficient to address the needs, and it is hard to implement the explainable components in the AI-enabled systems, as they lack the timeliness of the explanation that arrives when the user is interacting with the systems. In Balasubramaniam et al. [4, 26], although the user story was used to represent explainability requirements, the template is too complex to be used to analyze requirements from non-technical stakeholders who are unfamiliar with AI-enabled systems and difficult to express their explainability needs, especially the explainer element. Therefore, the proposed template of the user story in this study removed the explainer but included the more useful elements, such as timeliness and interaction. To further improve the user stories, the proposed work includes the translation of user stories to formal requirements, and the presentation of explanations is included in the formal requirements, which helps the developers to select more suitable explainability techniques to produce the explanations.

3.0 PERSONALIZED EXPLAINABILITY REQUIREMENTS ANALYSIS FRAMEWORK

This section describes the proposed framework. In this research, a personalized explainability requirements analysis framework for AI-enabled systems is proposed by breaking down the steps and categorizing them into four stages to support the requirement engineers in analyzing the explainability requirements in an organized and systematic manner. As depicted in Fig. 1, the proposed framework consists of four essential stages, namely (1) domain analysis, (2) stakeholder analysis, (3) explainability analysis, and (4) translation and prioritization. Sections 3.1 to 3.4 describe each stage in more detail.

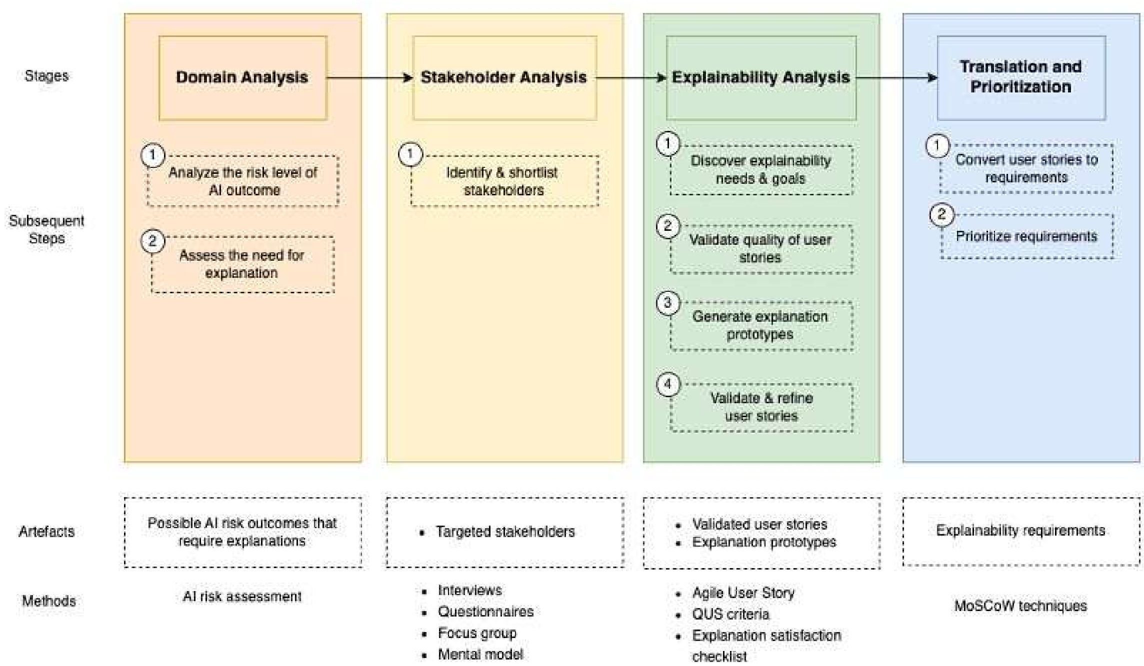


Fig. 1: Overview of the proposed FAT-BPSM.

3.1 Stage 1: Domain analysis

Domain analysis is the initial stage in requirement analysis, focusing on gaining a deep understanding of the specific domain information to justify the need for explainability in AI-enabled systems [23]. To justify the need for explainability, this stage involves analyzing the system's decision-making characteristics [21] based on the risk level of AI outcomes.

3.1.1 Step 1: Analyze the risk level of AI outcomes

Before the decision or outcomes of the AI-enabled systems were analyzed, the scale of both risk ratings, followed by the respective descriptions, must be predefined by the requirement engineers and relevant stakeholders based on the necessity or suitability of specific systems and use cases. Table 1 and Table 2 illustrate examples of definitions of probability and severity scale for AI-enabled healthcare systems.

After defining the probability and severity related to the domain and risk outcomes, a criticality matrix and the respective risk level are tabulated according to the criticality rating (see Table 3). In the example, the criticality ratings are divided into five risk levels as shown in Table 4. Ratings from 1 to 4 are categorized as 'Negligible' risk, while ratings from 5 to 9 are considered 'Low' risk. Criticality ratings from 10 to 14 fall under the 'Medium' risk, and ratings ranging from 15 to 19 are classified as 'High' risk. Any rating of 20 and above is considered a 'Very high' risk. This classification enables us to effectively assess and prioritize the level of risk associated with each criticality rating, allowing the need for explainability to be analyzed and determined.

By engaging with the relevant stakeholders, such as subject matter experts or decision makers, a list of potential risks of prediction outcomes generated by AI-enabled systems is identified. Next, the relevant stakeholders give ratings of severity and probability for each risk accordingly. If there is more than one stakeholder within a stakeholder group, they should engage in discussions to review and finalize their rating score, incorporating diverse insights and concerns, and ensuring that a consensus is reached. Finally, the criticality score of each risk is then calculated by multiplying the severity score and the probability to determine the risk level.

Table 1: Example of the probability scale for rating risks based on likelihood

Level	Probability *Likelihood of failure*	Description
5	Highly probable	Happens between 80% and above of all times
4	Probable	Happens between 50% to 79% of the time
3	Could occur	Happens between 30% to 49% of the time
2	May happen	Happens between 10% to 29% of the time
1	Unlikely	Happens below 10% of the time

Table 2: Example of the severity scale for rating risks based on consequence

Level	Severity *Consequence of failure*	Description
5	Major	Major health impact and may cause death
4	Significant	Severe health impact and hospitalized
3	Moderate	Moderate health impact and hospitalized
2	Minor	Mild health impact but not hospitalized
1	Very minor	No or very mild health impact

Table 3: Example of the risk rating matrix to assess criticality level

		Severity				
Probability		1	2	3	4	5
	1	1	2	3	4	5
	2	2	4	6	8	10
	3	3	6	9	12	15
	4	4	8	12	16	20
	5	5	10	15	20	25

Table 4: Example of the risk level based on criticality rating

Criticality rating	Risk level
1 – 4	Negligible
5 – 9	Low
10 – 14	Medium
15 – 19	High
20 – 25	Very high

3.1.2 Step 2: Assess the need for explanation

The requirement engineers analyze the system's decision characteristics by carrying out an AI risk assessment. Fig. 2 illustrates a decision workflow from a higher-level perspective to assess the need for explanations for an AI-enabled system based on the risk level. If a decision will be made based on the AI outcomes, the risks associated with the AI-enabled system need to be identified and assessed [28] to ensure the criticality of the outcomes and to justify the need for explainability and explanation. Based on the risk assessment, if the decision-making includes medium, high, and very high-risk factors, explainability is required to clearly explain how a decision or outcome has been generated by the AI-enabled system. At a higher level of explainability, it involves more detailed breakdowns, transparent, and accessible explanations about its decision-making processes.

On the other hand, if the outcomes are only used for reporting or informing purposes, or the criticality of the risk factors is low or negligible, a lower level of explainability is required to be implemented in the AI-enabled system, such as a movie recommendation system. At a lower level of explainability, the system provides minimal or general explanations about its decisions. The explanations are straightforward, focusing on the important aspects without including complex details. By identifying the levels of explainability based on the risk factors, the system provides the appropriate amount of information to the users, meanwhile balancing transparency, user understanding, and the complexity of the system's functions.

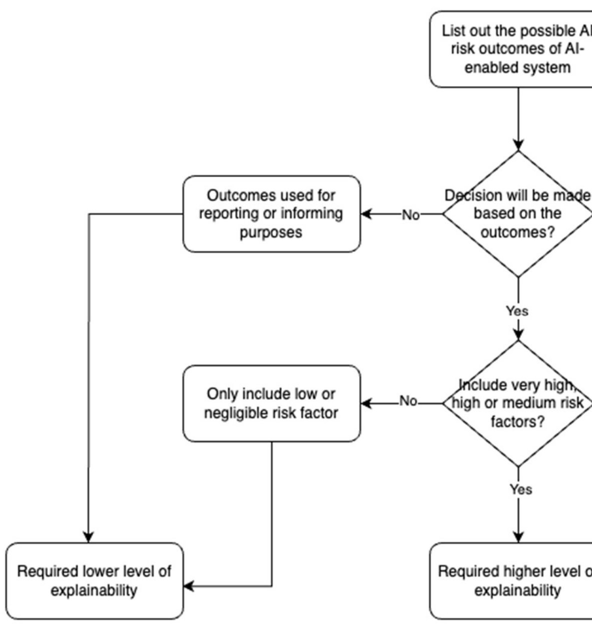


Fig. 2: Analyzing the system's decision characteristics in domain analysis

After the criticality of the AI outcomes is evaluated, critical outcomes require a higher level of attention and understanding from stakeholders, and personalising the explanations to their specific explainability needs can lead to more effective decision-making. In a nutshell, the higher criticality of risk factors indicates that more detailed explanations are required to support the decisions made and engender trust in humans. Once analyzed, explainability is needed for the AI-enabled systems, and the requirements engineers can proceed to the next stage to analyze the type of stakeholders.

3.2 Stage 2: Stakeholder analysis

Stakeholder analysis is the following stage that involves identifying and analyzing the various stakeholders who will receive the user-centric explanation provided by the AI-enabled systems. The proposed stakeholder analysis differs from traditional systems as this analysis in the end includes only four main groups of stakeholders because the selected groups of stakeholders provide cutting-edge insights, rather than a broad range of stakeholders that can be determined in traditional systems. In addition, the analysis emphasizes building trust from each of the key stakeholders through transparency and explainability, ensuring that the system aligns closely with the explainability needs and goals of the selected groups of stakeholders and does not diversify or generalize the explanations provided.

3.2.1 Step 1: Identify and shortlist stakeholders

Developing a good explainable AI-enabled system depends on the satisfaction of different types of stakeholders towards the appropriate and comprehensible explanations provided by the system [29]. It is crucial to identify the group of stakeholders, and this framework applies the Basic Stakeholder Analysis Technique [30] as it could quickly identify the potential stakeholders through interviews, questionnaires, scenarios, and focus groups with a checklist of questions as shown in Table 5.

In this study, four main groups of stakeholders are adopted and adapted from Langer et al. [31], Gerlings et al. [14], and Preece et al. [32], which include the development team, subject matter experts, decision makers, and affected users who will require explanations from the AI-enabled system's outcomes. In this context, the development team includes data scientists, software engineers, and AI developers. They are responsible for designing and developing AI-enabled systems. The subject matter experts refer to the experts who provide deep knowledge and insights in a certain domain to assist the development team in building explanations. Besides that, the system users include the decision makers and affected users. Some prototypical examples of decision makers include medical doctors, loan officers, judges, and hiring managers [31]. They utilized the predictions or outcomes of the AI-enabled systems to make critical decisions. In addition, the affected users are people who will be impacted by the system's decision; usually, they are referring to the patient, loan applicant, or job applicant. Table 5 provides a list of questions to discover the potential stakeholders that may require personalized explanations from an AI-enabled system.

Table 5: A checklist of questions to identify the stakeholders

Stakeholder	Questions
Development team	• Who is developing and verifying the accuracy of the AI models? • Who is responsible for cleaning data? • Who is the quality assurance team responsible for evaluating the AI system's performance and compliance? • Who is maintaining the AI system after deployment?
Subject matter experts	• Who are the domain experts who provide guidance or insights for the AI system? • Who evaluates/approves whether the AI system is safe to use?
Decision makers	• Happens between 30% to 49% of the time? • Who is the key decision maker for the AI system? • Who are the end users of the AI system's outcomes or recommendations? • Are there any customer support or helpdesk teams responsible for handling user inquiries or issues related to the AI system?
Affected users	• Who might be affected positively or negatively by the system's outcome? • Who are the consumers of the AI system's outcomes or recommendations?

Although there are more stakeholders to be identified, these four types of stakeholders could be commonly seen in real-world settings; meanwhile, they are also consistently highlighted in previous studies. The requirement engineers can shortlist the targeted stakeholders to draw boundaries and responsibilities between stakeholders so that the explanations can be provided effectively. In short, understanding the perspectives of different stakeholders is vital for capturing diverse explainability requirements and ensuring the system can generate tailored explanations that meet their expectations.

3.3 Stage 3: Explainability analysis

Explainability analysis focuses on assessing the interpretability and transparency requirements of the AI-enabled systems. This stage involves discovering the characteristics and type of explanations expected based on the explainability needs and goals of the shortlisted stakeholders. In this stage, the Agile User Story [33], which is a user-centric approach, is used to capture personalized explainability requirements from the standpoints of the stakeholders.

3.3.1 Step 1: Discover Explainability Needs & Goals

In general, the user stories are expressed from the end user's perspective as persona + need + goals. In the explainability context, the timeliness of the explanation is relatively important for different types of stakeholders while they are interacting with the AI-enabled systems to receive the explanation.

The enhanced user stories, as shown in Fig. 3, highlighted the perspectives of different groups of stakeholders, all of whom could express themselves with a guided user story template. At the same time, the requirement engineers could better understand their expectations of personalized explanations from the stakeholders and translate the user stories into all-rounded software requirements. The enhanced template of a user story is built up by the persona, explainability needs, explainability goal, and other new components (i.e., type of question, timeliness, and interaction).

> As a **<persona>**, I want to understand **<type of question>** on the **<explainability need>** by **<timeliness>** on the **<interaction>** with the system, so that **<explainability goal>**.

Fig. 3: Template of enhanced user stories to collect explainability requirements

The terminologies stated in the template are expressed as below:

- **Persona** acts as the user/role who writes their "story" to reflect their requirements and expectations. The persona must be a person or a role, but not a group of users. The user also refers to the specific types of stakeholders discussed in Section 3.2.1, which makes the user story unique and tailored to each type of stakeholder, as they have distinct explainability needs and goals. This diversity of users emphasized the neediness for personalized explainability requirements, ensuring that the system provides tailored explanations that align with the varying requirements and priorities of each stakeholder.
- **The type of question** is classified according to the type of explanation (i.e., plain-fact, causal, contrastive, transfactual, or counterfactual) by stating what, why, why not, what if, or how to in the user stories as depicted in Table 6.
- **Explainability need** refers to the need for an understandable explanation that should be provided by the systems to the persona.
- **Timeliness** defines the best or correct time to receive the explanation from the systems, for instance, before, when, and after using the prediction system.
- **Interaction** reflects how the persona interacts with the system; it can be expressed as building, evaluating, monitoring, or using the prediction models.
- **Explainability goals** specified in user stories provide a clear illustration of the purpose behind receiving explanations for the AI outcomes. Ideally, these goals may focus on obtaining an explanation of the possible outcomes that have been identified during the *Domain Analysis* as medium, high, or very high risks.

By providing explanations for the prediction outcomes generated by AI-enabled systems, stakeholders can make more informed decisions and take the appropriate actions. The template shown in Table 6 can be used to assist the shortlisted stakeholders in writing their requirements for explainability easily in simple and precise sentences or paragraphs, and then the requirement engineers could capture their explainability needs and goals systematically and turn them into a proper requirement in the following stage. Although the user story approach is a non-formal method to elicit requirements, the perspective of all the stakeholders could express their requirements and expectations in their own words instead of technical terms that some of them are unfamiliar with in real-world settings.

3.3.2 Step 2: Validate Quality of User Stories

Since the user stories are written by various stakeholders, the user stories would be duplicated or ambiguous; therefore, the user stories must be validated and refined to ensure their clarity and usefulness in extracting the key requirements. This step adopted the Quality User Story (QUS) framework proposed by Lucassen et al. [34], which consists of 13 quality criteria as shown in Table 7 to determine the quality and ensure the integrity of user stories in terms of syntax, semantics, and pragmatics. For instance, as reported by the tool, a user story written as "As a doctor or nurse, I want to..." is not atomic because the persona of a user story should comprise only one user.

Table 6: The templates of user stories for five different types of explanation

Type of Explanation	User Story			
	Persona	**Type of Question & Explainability Needs**	**Timeliness & Interaction**	**Explainability Goals**
Plain-fact (What/ How)	As a [user],	I want to understand **what/how** [explainability needs]	[when the user is interacting with] the system	so that [goal].
Causal (Why)	As a [user],	I want to understand **why** [explainability needs]	[when the user is interacting with] the system	so that [goal].
Contrastive (Why not)	As a [user],	I want to understand **why not** [explainability needs] predicted as [outcome A] but [outcome B]	[when the user is interacting with] the system	so that [goal].
Transfactual (What if)	As a [user],	I want to [explainability needs] **what if** [factor A] is changed to [factor B]	[when the user is interacting with] the system	so that [goal].
Counterfactual (How to)	As a [user],	I want to understand **how to** [explainability needs] to change the outcome as requested	[when the user is interacting with] the system	so that [goal].

Table 7: Thirteen (13) criteria used for user story quality validation (adopted from Lucassen et al., 2016)

Criteria	**Description**	**Individual/ Set**
Syntactic		
Well-formed	A user story includes at least a role and a means	Individual
Atomic	A user story expresses a requirement for exactly one feature	Individual
Minimal	A user story contains nothing more than role, means, and ends	Individual
Semantic		
Conceptually sound	The means express a feature, and the ends express a rationale	Individual
Problem-oriented	A user story only specifies the problem, not the solution to it	Individual
Unambiguous	A user story avoids terms or abstractions that lead to multiple interpretations.	Individual
Conflict-free	A user story should not be inconsistent with any other user story	Set
Pragmatic		
Full sentence	A user story is a well-formed full sentence	Individual
Estimatable	A story does not denote a coarse-grained requirement that is difficult to plan and prioritize	Individual
Unique	Every user story is unique, duplicates are avoided	Set
Uniform	All user stories in a specification employ the same template	Set
Independent	The user story is self-contained and has no inherent dependencies on other stories.	Set
Complete	Implementing a set of user stories creates a feature-complete application, no steps are missing.	Set

In Table 7, there are several keywords, such as role, means, and ends, that have different meanings described by the authors. A role defines the stakeholder or persona who presents the needs. A mean refers to various types of requirements and may have different structures. An end explains the reason for the means, and it may have one or multiple parts. In addition, the criteria are classified into individual or set. An individual criterion can be evaluated independently, with a sole emphasis on its qualities. In contrast, when a criterion is classified as set, the process of validation requires validating numerous interconnected user stories as an integrated whole. Based on the criteria mentioned, the requirement engineers could identify mistakes and amend the user stories accordingly. After the requirement engineers validate the quality of user stories, they can prepare to generate the explanation prototypes based on the predefined user stories.

3.3.3 Step 3: Generate Explanation Prototypes

After obtaining the predefined user stories using the template presented in Table 6 (Step 1 in Section 3.3.1), the type of explanations is analyzed to understand the explainability needs and goals for each group of stakeholders. The purpose of this research is to define personalized explainability requirements. As a result, the prototype technique is proposed in Step 3 of Stage 3 to support the validation and refinement of user stories in Step 4 (Section 3.3.4). The personalized explanation prototypes are generated in this step using the explainability techniques commonly used to generate explanations for these five types of explanations (Plain-fact, Causal, Contrastive, Transfactual, Counterfactual) and sample datasets of the AI-enabled systems.

The explanation prototypes generated in this step are not working prototypes but are personalized prototypes generated to demonstrate the examples and visualization formats of explanations based on the predefined user stories for each group of stakeholders. In practice, the requirement engineers could prepare a user interface to collect the templatized user stories and mirror the respective explanation prototype. This hands-on interaction allows stakeholders to obtain a concrete understanding of how the proposed features will operate and determine whether they meet the stakeholders' expectations. These explanation prototypes could be presented as charts, tables, text, or a combination of other visualization techniques to assess whether the desired explainability requirements by all stakeholders are met.

The personalized explanation prototypes focus on demonstrating the predefined explainability user stories and the feasibility of explainability techniques for each type of explanation. Once the explanation prototypes are created, the requirement engineers will show the explanation prototypes to each group of stakeholders, including the developer team, subject matter experts, decision makers, and affected users, in the next step to validate and refine the user stories. This allows the stakeholders, especially non-technical stakeholders, to validate the correctness of the explainability needs and goals defined in the user stories by looking at the sample descriptions of explanations and examples of data visualization generated using appropriate explainability techniques.

3.3.4 Step 4: Validate and Refine User Stories

By validating user stories through prototyping, the requirement engineers can close potential gaps early in the development process, minimize rework, and ensure that the delivery of high-quality explanation models can be achieved. During the validation phase, stakeholders provide feedback and a satisfactory level of usability and functionality of the explanation prototypes with certain acceptance criteria checklists, such as explanation satisfaction checklists provided by Hoffman et al. [35]. The checklists are categorized into two major perspectives: (i) the development team's perspective and the subject matter expert's perspective, and (ii) the decision maker's perspective and the affected user's perspective. For each checklist item, each participant can give a rating using a 5-point Likert scale (5: I agree strongly, 4: I agree somewhat, 3: I'm neutral about it, 2: I disagree somewhat, 1: I disagree strongly).

Explanation satisfaction checklist (Development team's perspective and subject matter expert's perspective) (adapted from [35]):

- From the explanation, I know how the [software, algorithm, tool] works.
- This explanation of how the [software, algorithm, tool] works is satisfying.
- This explanation of how the [software, algorithm, tool] works has sufficient detail.
- This explanation of how the [software, algorithm, tool] works seems complete.
- This explanation of how the [software, algorithm, tool] works is useful to my goals.
- This explanation of the [software, algorithm, tool] shows me how accurate the [software, algorithm, tool] is.

Explanation satisfaction checklist (Decision maker's perspective and affected user's perspective) (adapted from Hoffman et al., 2023):

- From the explanation, I know how the [software, algorithm, tool] works.
- This explanation of how the [software, algorithm, tool] works is satisfying.
- This explanation of how the [software, algorithm, tool] works has sufficient detail.
- This explanation of how the [software, algorithm, tool] works seems complete.
- This explanation of how the [software, algorithm, tool] works tells me how to make decisions.

Stakeholders evaluate the intuitiveness of the prototype and the justification for the needs of any additional requirements. Afterwards, the average rating is calculated, and the validation passing mark should be scored at least 4 or above; otherwise, refinement or abolishment of the requirement is needed. In short, the requirement engineer iterates on the enhancement of user stories, incorporating the feedback, and fine-tunes the user stories to better align with their final needs, and any additional requirements would assist in eliciting the effective user stories so that the requirement engineer could convert the well-defined user story to proper software requirements in the following stage. If some of the explanations presented in the prototypes are not required by the relevant stakeholders, the user stories may be removed.

If there are any major conflicts due to different perspectives and results in the stage of validating the quality of user stories, validating, and refining user stories, a brainstorming session will be held to resolve the conflicts. The conflicts will be analyzed by inviting all the relevant stakeholders to exchange viewpoints, information, and justifications with each other. The negotiation or voting technique can be used to get a common and agreeable solution to resolve the conflicts. The conflict resolution will be documented to finalize the validation.

The prototype validation feedback collected from stakeholders in this step also contributes to defining personalized explainability requirements using the enhanced version of the explainability requirement template presented in Fig. 5 (Section 3.4.1). The expected data visualization, such as text, chart, and table, can be included as part of the explainability requirements. This is crucial to support the system designers and developers in designing and developing user-centric explanations using the expected data visualization in the design and development phases.

3.4 Stage 4: Translation and Prioritization

Translation and prioritization are the last stage of the proposed explainability requirement analysis. This stage involves converting collected user stories to official explainability requirements statements and prioritizing the requirements.

3.4.1 Step 1: Convert user stories to requirements

As mentioned earlier, the user stories are not official software requirements. Köhl et al. [36] formulate explainability requirements based on a customized template that is shown in Fig. 4.

A system *S* must be explainable for target group *G* in context *C*
with respect to aspect *Y* of explanandum *X*.

Fig. 4: Explainability requirement template [36]

The terminologies stated in the template are expressed as below:

- **System** (S) refers to the particular AI-enabled system that is going to be developed.
- Explainable means that the system should have the capability to provide an understandable explanation for its decisions or outcomes.
- **Target group** (G) is the relevant group of end users with a specific background who need to understand the behaviours or require an explanation from the system.
- **Context** (C) refers to the circumstances or environment in which the system operates. It considers the specific conditions, constraints, and factors that may influence the system's behaviour.
- **Aspect** (Y) specifies a particular facet or feature of the system's behaviour or output that requires an explanation. It narrows down the focus to a specific aspect that needs clarity.
- **Explanandum** (X) is part of the system's behaviour, decision, or outcome that the target group needs to understand.

Fig. 5 shows the enhanced version of the explainability requirement template proposed in this study, where enhancements are possible in several areas, including scope, the timeliness of explanations arrived, and the visual presentation for the specific stakeholders in order to bridge the gap of misalignment between the requirements engineers and stakeholders. Meanwhile, context is removed from the original template as it is hard to express. Based on this template, it can significantly enhance the subsequent phases of design and implementation by selecting more appropriate techniques, hence ultimately leading to improved software development.

A system *S* must be explainable for target group *G* from
scope P with timeliness T with respect to aspect *Y* of explanandum *X*
by presentation V.

Fig. 5: Enhanced explainability requirement template

The new terminologies stated in the enhanced template are expressed as below:

- **Scope (P)** refers to the extent and depth of information provided within an explanation (i.e., individual, or overall model), where it can be obtained from the user story template (type of explanation).
- **Timeliness (T)** emphasizes the arriving time of the explanation provided by the systems (i.e., before, during, and after using the system), where it can be obtained from the user story template (timeliness).
- **Presentation (V)** expresses the expected data visualization, such as text, chart, and table, which can be obtained during the validation of prototyping.

Based on the template in Figure 5, the requirement engineers map and convert the validated user stories in the previous stage to standard statements of non-functional requirements.

3.4.2 Step 2: Prioritize requirements

After formulating the requirements, the requirement engineers should review and prioritize the requirements according to their urgency, user impact, and technical constraints. The requirement engineers should collaborate with relevant stakeholders, such as subject matter experts and decision makers, to discuss and align their prioritization of explainability requirements. The technique applied in this step is known as MoSCoW prioritization [37], which provides a straightforward categorization based on the relative need of requirements into 4 priority groups, which are: Must-Have (M), Should-Have (S), Could-Have (C), and Won't-Have (W). This classification helps the development team quickly understand the importance of each requirement and allocate the time and cost budget more efficiently. For instance, the focus on Must-Have requirements would ensure the critical and essential explanations are delivered early in the implementation phase. In contrast, the Won't Have requirements cannot be implemented in the current iteration as they have a lower priority. Furthermore, the MoSCoW technique ensures that business and user needs and expectations are prioritized by focusing on Must-Have and Should-Have requirements.

This user-centric approach helps deliver explanations that satisfy the explainability needs of all users and address their explainability goals. With the flexibility of deprioritization (Could-Have and Won't-Have), it helps to accommodate the rapidly changing requirements meanwhile maintaining the focus on delivering the essential functionality. Since the Agile User Story was adopted in the previous phase, the synergy between the Agile User Story and the MoSCow technique may iteratively refine the requirement and reprioritize whenever necessary in every development loop. This incremental and iterative approach allows for flexibility, adaptability, and early delivery of explainability to users. If any conflicts are identified during requirements prioritization, a brainstorming session will be conducted to resolve the conflicts among all relevant stakeholders. They can exchange their opinions, viewpoints, and perspectives to analyze the conflicts and negotiate agreeable solutions to finalize the prioritisation of requirements. The agreeable solution will be documented.

Once the requirements have been prioritized and validated, the requirement engineers should unambiguously document them. If there is any amendment throughout the validation iteration, the requirement engineers should update the requirement documentation to reflect the refined and validated set of requirements. This documentation serves as a reference for the development team and helps to ensure that the explanations are developed and aligned with the validated requirements in the subsequent phase of development, including the design and implementation phases.

4.0 CASE STUDY

This section presents the evaluation of the proposed personalized explainability requirements analysis for AI-enabled systems using a case study. In this research, the objective of conducting this case study is to evaluate the applicability and effectiveness of the proposed framework in a real-world setting using an AI-enabled heart disease risk prediction system with the engagement of relevant stakeholders. The guidelines by Runeson and Höst [38] recommended practices and checklists to conduct and report a case study in software engineering. This case study was designed and conducted based on the guidelines. Section 4.1 outlines the case study context and the process of participant recruitment. Section 4.2 demonstrates the data collection and analysis procedures. Moreover, this study was conducted in full compliance with the ethical standards outlined by the Universiti Malaya Research Ethics Committee (UMREC), under the reference number UM.TNC2/UMREC_2248. All participants were informed of the study's objective, procedures, potential risks, and advantages before their participation.

4.1 Case Study Context

Despite the analysis being conducted in any domain, recent studies in the healthcare domain raised several challenges in XAI (Al Kuwaiti [39-42]. Al Kuwaiti et al. [39] stated ethical and social issues, governance problems, and technical challenges. Healthcare professionals have a poor understanding of the AI-enabled clinical decision support systems as they lack identified risks, trust issues, biases, accountability, and explainability from the AI-enabled system, yielding a low adoption of AI solutions. Furthermore, the roles of AI in the medical field are listed in [43], including clinical decision-making, treatment optimization and patient care, disease diagnosis, and precision medicine. Kapa [43] and Lötsch et al. [44] also mentioned that transparency and explainability in AI-enabled healthcare systems should be enhanced to illustrate that transparent decisions are made with available scientific theory to explain the conclusion arrived, not just for developers but also for non-developers. Therefore, an AI-enabled heart disease risk prediction system that focuses on predicting heart disease risk of the general population patients was selected as the case study context because it uses AI models to make risk predictions of patients and assist physicians in decision-making, yielding personalized explanations that are needed for different stakeholders.

Given that the case study demands insights from numerous experts with diverse backgrounds, invitations were extended via emails to four participants well-acquainted with the AI-enabled heart disease risk prediction system. These individuals were invited to play the roles of various stakeholders and contribute their insightful opinions and advice. One of the participants is a board-certified cardiologist who will play two roles: decision maker and subject matter expert (i.e., physician and senior consultant cardiologist) in the case study. The other participant is an expert in biomedical engineering, specialising in cardiac magnetic resonance imaging and computer-aided diagnosis research. She has experience in the development of machine learning models for predicting heart disease and played the role of the development team (i.e., AI developer). The last two participants have a background in machine learning, predictive analytics, and cognitive science. They played the role of the affected user (i.e., patient). Each participant possesses a robust and specialized background, and they are aware of the importance of explainability in AI, enabling them to provide invaluable advice as their expertise aligns precisely with the requirements to play the roles in the system as intended.

4.2 Data collection and analysis

The data collection procedures were performed to collect various data for the evaluation of the applicability and effectiveness of the proposed framework in analyzing explainability requirements for AI-enabled systems. The procedures of the data collection and analysis are visualized in a flowchart as shown in Fig. 6. A total of three online meetings and two follow-up meetings were conducted with participants separately in an accelerated manner to facilitate the comprehensive and personalized analysis of explainability requirements for AI-enabled heart disease risk prediction systems. Interviews and survey questionnaires were used to collect data in this case study.

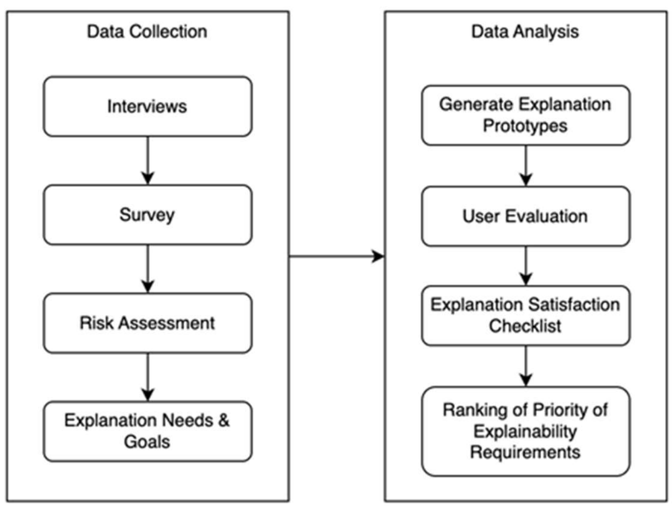


Fig. 6: Flowchart of data collection and analysis procedures

In the first meeting with stakeholders, the proposed framework was presented, and data collection for stages 1 and 2 ensued. A tailored risk assessment form through a Word file, addressing potential risks specific to heart disease prediction systems, was utilized to gather risk assessment scores from stakeholders, particularly subject matter experts and decision makers. Simultaneously, the participants were categorized based on their roles through the stakeholder type assessment. After that, the explanation needs and goals of stakeholders were diligently captured using the customized user stories template through an online Google Sheet during the second round of meetings. To provide clarity, examples of user stories were showcased, and participants articulated their needs using the provided template.

Eventually, the analysis of explainability needs and goals from stakeholders paved the way for progression to stage 3 and stage 4. In a subsequent meeting, prototypes were generated and presented to stakeholders for user evaluation, wherein participants provided valuable feedback. A set of explanation satisfaction checklists was prepared via online Google Forms to systematically collect evaluation and satisfaction scores for the explanation prototypes. For stage 4, a priority form was prepared using a Word file for stakeholders to rank the priority of explainability requirements.

The analytical procedures involved a comprehensive examination of the collected scores. A detailed analysis was conducted stage by stage for explainability requirements analysis throughout the case study, thereby ensuring a thorough understanding of the stakeholders' explainability requirements as well as the effectiveness and impact of the proposed framework.

4.3 Results

This section presents the results of the case study based on the step-by-step procedures for analyzing the explainability requirements for the AI-enabled heart disease risk prediction system and evaluations from the participants.

4.3.1 Stage 1: Domain analysis

During the domain analysis stage, the nine potential AI risks were initially identified by analysing the studies related to AI-enabled heart disease systems. In the first online meeting, the AI risks were assessed by the senior consultant cardiologist using the probability scale, severity scale, risk rating matrix, and risk level defined in Tables 1 to 4. As presented in Table 8, there are two medium risks (i.e., "Misdiagnosis or False Negatives" and "Decision Authority") and two high risks (i.e., "Dependency on Data Quality" and "Lack of Transparency").

This indicates that explainability requirements and more detailed explanations are required for this AI-enabled heart disease prediction system. On the other hand, "Ethical Concerns" were not selected as a potential risk, "Limited Expertise" and "Model Drift" were rated as 'Low', "Bias and Fairness", and "Algorithm Complexity" were rated as 'Negligible' from the perspective of a subject matter expert because the AI-enabled systems must always serve as an assisting tool for clinical decision-making. It should never override the final decisions made by physicians or patients, which may raise ethical concerns.

Table 8: High and medium risk assessment for an AI-enabled heart disease risk prediction system

AI Risk	Severity (1-5)	Probability (1-5)	Criticality	Risk Level
Misdiagnosis or False Negatives: The heart disease system may not accurately identify individuals at risk for heart disease, resulting in the possibility of false negatives. This may prevent patients from receiving necessary medical treatment [45].	4	3	12	Medium
Dependency on Data Quality: The heart disease predictive system's accuracy is highly dependent on the quality and relevance of the input data. Inaccurate or insufficient data may result in inaccurate predictions and recommendations [46].	5	3	15	High
Lack of Transparency: Users, including physicians and patients, may be hesitant to trust the system's predictions if they cannot perceive how it makes decisions, or if explainability is absent [47].	4	4	16	High
Decision Authority: The system should assist in clinical decision-making only, but not overriding any party (e.g., physician, patient) to make a final decision [48].	4	3	12	Medium

4.3.2 Stage 2: Stakeholder analysis

Stakeholder analysis is the second stage that involves identifying and analyzing the various stakeholders who will receive the user-centric explanation provided by the AI-enabled systems. Building a good explainable AI-enabled system depends on the satisfaction of different types of stakeholders with the appropriate and comprehensible explanations provided by the system. A list of questions listed in Section 3.2.1 was utilized to identify the stakeholders for each stakeholder group that requires personalized explanation, and they are shortlisted as shown in Table 9.

Table 9: Shortlisted stakeholders and their relative roles and responsibilities that require personalized explanations in a heart disease prediction system

Stakeholder	Participant Role	Stakeholder responsibility
Development team	AI developer	Engineers in designing and developing AI-enabled systems. They utilized the explanation to improve the acceptability and accuracy of the systems.
Subject matter experts	Senior consultant cardiologist	Experts in providing insights and deep knowledge of heart disease to assist the development team in building explanations.
Decision makers	Physician	General medical doctors utilize the prediction of AI-enabled systems in making decisions and recommendations for patients.
Affected users	Patient	General population patients who may use or be impacted by the system's decision and seek recommendations.

4.3.3 Stage 3: Explainability analysis

Explainability analysis focuses on assessing the interpretability and transparency requirements of the AI-enabled system. This stage involves discovering the characteristics and type of explanations expected based on the explainability needs and goals of the shortlisted stakeholders.

a) Step 1: Discover Explainability Needs & Goals

In this study, the Agile User Story is a user-centric approach that can be conducted to capture explainability requirements from the standpoints of the stakeholders, as the explanation required is different. The shortlisted stakeholders expressed their explainability needs and goals with the customized template of the user story. Although there are five distinct types of explanation given, stakeholders seldom expressed their user stories in terms of "What if" and "Why not", they are better at expressing their needs and goals with the other three types of questions: plain fact (what/how) and causal (why), and counterfactual (How to). In certain user stories, the "why not" questions express exactly opposition to the corresponding 'why' questions. For example, "Why was my application rejected?" holds the same meaning as "Why was my application not accepted?" This parallel structure emphasizes that the user stories should be validated and filtered in order to finalize the complete set of user stories so that all the user stories are unique and can be fulfilled without duplication.

b) Step 2: Validate Quality of User Stories

After the user stories are validated based on QUS criteria, the violated or duplicated user stories are excluded to reduce confusion or reclarification before the prototypes are generated. The validated user stories and the excluded user stories collected from various stakeholders can be retrieved from this link1. There are 26 user stories collected from four stakeholders (i.e., 12 user stories from the AI developer, 3 user stories from the senior consultant cardiologist, 4 user stories from the physician, and 7 user stories from the patients).

c) Step 3: Generate Explanation Prototypes

Based on the user stories, explanation prototypes are produced accordingly to meet the explainability needs and goals. Post-hoc explanation prototypes are generated so that diverse stakeholders can visualize and interact with the potential explanations to validate the user stories before proceeding to the design and development phases. In order to better address the post-hoc analysis, the selection of ML algorithms includes decision tree classification, support vector machine, histogram-based gradient boosting classification tree, and multi-layer neural network [45].

[1] https://figshare.com/s/337a4b9c6944ae9e019b

To prevent data consent and data privacy issues, an open dataset of cardiovascular disease is obtained from Kaggle2 to produce post-hoc explanations on the trained model. The dataset claimed that all of the dataset values were collected at the moment of medical examination. The dataset consists of 70,000 records of patient data and 11 features. However, the data is preprocessed to 8 features, 1 target variable, and reduced to 20,000 records for illustration purposes. Table 10 shows the list of features of the dataset used in this case study. There are three types of features:

- Objective: factual information;
- Examination: results of medical examination; and
- Subjective: information given by the patient.

Table 10: The features used in the heart disease prediction model

No.	Feature Name	Type of Feature	Feature Data Type
1	Age	Objective	float (years)
2	Gender	Objective	categorical code
3	BMI	Objective	float
4	Systolic blood pressure	Examination	integer
5	Diastolic blood pressure	Examination	integer
6	Cholesterol	Examination	categorical code
7	Glucose	Examination	categorical code
8	Smoking	Subjective	categorical code
9	Presence or absence of cardiovascular disease	Target variable	categorical code

While producing the explanation prototypes, a noteworthy discovery emerged as certain explanations could be seamlessly incorporated and mapped to different user requirements. For example, the transfactual explanation, where user stories are expressed with a "what if" keyword, could be synergistically integrated with counterfactual explanations. This integration allows for the provision of detailed explanations on how to attain improved or desired results by implementing recommended changes. The five types of explanation identified can be developed interchangeably, creating a dynamic and flexible framework to fulfill diverse user requirements and establish an enhanced representation of explanations. This interchangeability not only enriches the versatility of the explanations but also maximizes their effectiveness in catering to a wider array of user needs. Numerous explainability techniques and libraries were utilized and summarized in Table 11. The example of personalized explanation prototypes developed based on four user stories collected from the decision maker (i.e., physician) is presented in the Appendix (see Tables 13 to 17).

Table 11: The explainability techniques used to generate respective explanation prototypes

Type of explanation	Explainability techniques	Library
Plain-fact	Generalized Linear Rule Model (GLRM) & SHAP	XAI360, Alibi Explain
Causal	LIME, SHAP	XAI360, Alibi Explain
Contrastive	Comparison of LIME	XAI360
Transfactual	LIME	XAI360
Counterfactual	DiCE, CounterfactualProto	Alibi Explain

d) Step 4: Validate and Refine User Stories

During the validation and refinement of user stories, the personalized explanation prototypes are generated based on user stories provided by the AI developer (Development team), Senior consultant cardiologist (Subject matter expert), Physician (Decision maker), and Patients (Affected users). The user stories and explanation prototypes were presented to the 4 participants who represented different stakeholder roles. The stakeholders evaluated the prototypes by giving feedback and satisfaction scores for the explanation prototypes using the explanation satisfaction checklists as shown in Section 3.3.4. The average score of each type of evaluation given by each stakeholder is tabulated in Table 12. The average score above 4 will be accepted; otherwise, the explanation is rejected and removed from the list of user stories as it is not relevant or helpful from the perspectives of the specific stakeholder.

[2] https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset/data

When excluding certain explanations from stakeholders in AI-enabled systems, it is essential to differentiate between core requirements and user stories that do not directly contribute to the personalized explainability feature. Although the stakeholders provided a user story for contrastive explanation, the resulting explanation prototype was rated below 4 (refer to Table 12). This low rating indicates that the prototype did not align with the actual needs and goals of the users. Based on this evaluation, developers may choose to remove or refine this explanation requirement to avoid producing irrelevant or ineffective personalized explanations. By excluding underperforming prototypes at this stage, the team can reduce unnecessary complexity and focus on explanations that resonate better with users.

Beyond the point of evaluation score of the prototypes, the feedback obtained from the stakeholders could be further analyzed based on the quality attributes they are concentrating on and looking forward to in the given explanations. Some other attributes, such as usability and HCI, are discovered, and it is requested to be implemented in the explanations.

Table 12: Evaluation score of each type of explanation given to the stakeholders

Type of explanation	AI developer (Development team)	Senior consultant Cardiologist (Subject matter expert)	Physician (Decision maker)	Patient (Affected User)
Plain-fact (Global)	4.2 (Acceptable)	4.3 (Acceptable)	4.4 (Acceptable)	4.2 (Acceptable)
Causal (Local)	4.2 (Acceptable)	4.3 (Acceptable)	4.2 (Acceptable)	4.2 (Acceptable)
Contrastive (Local)	2.3 (Rejected)	4.0 (Acceptable)	-	-
Transfactual (Local)	-	-	4.2 (Acceptable)	5.0 (Acceptable)
Counterfactual (Local)	4.0 (Acceptable)	-	4.6 (Acceptable)	5.0 (Acceptable)

4.3.4 Stage 4: Translation and Prioritization

After confirming and finalizing all the user stories, they are translated into formal requirements based on the requirements template presented in Section 3.4.1. The translation of the user story and prioritization of requirements can be retrieved from this link[3].

5.0 DISCUSSION

This section presents the evaluation of the proposed personalized explainability requirements analysis for AI-enabled systems using a case study. In this research, the objective of conducting this case study is to evaluate the applicability and effectiveness of the proposed framework. This section discusses the results presented in the previous section. Section 5.1 discusses the applicability and effectiveness of the proposed framework. Section 5.2 and Section 5.3 present the implications for research and practice, and comparison with related work. Limitations and threats to validity are discussed in Section 5.4 and Section 5.5.

5.1 Applicability and effectiveness of the proposed framework

The case study demonstrates that the four-stages proposed framework can be applied systematically in a real-world setting. The results of surveys and interviews show that the applicability and effectiveness of the proposed framework are highly acceptable by the requirements engineers and the stakeholders, with and without the knowledge of AI. The proposed framework emerges as an effective collaboration between requirements engineers and the four distinct types of stakeholders who may require personalized explanations in AI-enabled systems. The structured framework, associated with the proposed stages and steps, proves to be not only suitable but also highly effective. Every stage of the proposed framework highlights the responsibilities of requirements engineers and involves the relevant stakeholders to ensure the execution of the framework is appropriate and necessary.

Requirements engineers are entrusted with performing careful steps, starting from analysing the possible outcomes generated by AI-enabled systems through risk assessment and discovering the stakeholders' perspective in earlier stages. Furthermore, the enhanced user story template, prototype evaluation form, and prioritization

[3] https://figshare.com/s/79aa44b10adac574ba85

method are designed to be user-friendly, enabling stakeholders with or without AI backgrounds to comprehend and apply these steps with clarity. This is proven by positive feedback mentioned by one of the participants who does not have experience in writing user stories during the interview: "The customized user story template is easier to understand and written according to the elements stated in the template, especially needs and timeliness, which could express what and when I need the explanation …". The existing user story template, while inherently flexible, has been identified as presenting challenges for users in expressing their explainability requirements. In response, the proposed steps for each stage in the requirements analysis framework offer an approach where the explainability requirements can be expressed in layman's terms, with specific elements stated explicitly in the personalization of user stories.

Furthermore, the process of producing explanation prototypes by requirements engineers and subsequently evaluating these prototypes with stakeholders serves as a tangible common ground for both parties to gain a comprehensive understanding of each other's perspectives regarding the desired end product and its visual representation. The invaluable feedback garnered during the prototype evaluation not only refines the explanation visualization but also emphasizes that quality attributes should be potentially associated with explanations given. For instance, the explanations provided to affected users extend the consideration of quality attributes to the user experience, focusing on usability in Human-Computer Interaction (HCI). Elements such as texts, fonts, colours, and the flexibility to expand on detailed explanations can be meticulously designed and developed. This approach ensures that the explanation provided to affected users aligns with HCI principles, making the system easy to learn, easy to remember, easy to read, and easy to interpret. Moreover, the quality of the user stories and explanations was investigated through careful evaluations. It contributes to a user-centric approach to identifying the explainability requirements that enhance the overall effectiveness and correctness of explanations provided by the AI-enabled systems.

After the user stories are collected and validated accordingly, it is crucial to recognize that the translation of these user stories into formal explainability requirements constitutes a refinement that we cannot overlook from the angle of software engineering, given its significant contribution to the overall process. This translation and prioritization stage acts as a transformative bridge, distilling the insights extracted from user stories into precise and formal requirements for explainability and illustrating the prioritization of the requirements from the perspective of stakeholders. This refinement not only empowers requirements engineers to gather more straightforward details related to explainability from different stakeholders' perspectives but also serves as a catalyst for enhancing the selection of explainability techniques, thereby contributing to the generation of more fitting and effective explanations in subsequent development phases.

5.2 Implications for research and practice

According to the evaluation results, it becomes evident that they show significant implications for both XAI and software engineering (SE) development practices with the active involvement of various stakeholders for the development of AI-enabled systems. The impact of the proposed framework extends beyond its current effectiveness, influencing the broader landscape of XAI and contributing to the enhancement of software engineering development practices in building explainable components in AI-enabled systems, particularly in personalized and user-centric requirements analysis.

Assessing the possible AI risks of the AI-enabled systems could help the requirement engineers to better understand the impact of the prediction outcomes and focus on improving transparency by incorporating the right degree of explainability in the AI-enabled systems. This insightful analysis not only allows them to comprehend the impact of domain possible risks from predictions but also facilitates more effective planning for the allocation of time and resources. Specifically, it allows for strategic resource allocation toward the development of tailored explanations, particularly for events classified as higher risk during the earlier development process; otherwise, implementing explainability could be excluded. This proactive analysis ensures that the explanations are implemented whenever necessary to address and mitigate potential risks, enhancing the overall transparency and trustworthiness of the AI-enabled systems.

One notable implication is the potential improvement in the design and development of explainable components and features within AI-enabled systems. Well-defined explainability requirements are documented by effectively investigating and validating the stakeholders' explainability needs and goals. It provides more information on the explainability requirements in terms of scope, timeliness, and visualization from various stakeholders so that a better selection of the future XAI techniques and presentation of the explanations could be achieved during the design and development phases.

Moreover, the evaluation results underscore the framework's role in fortifying the responsible AI aspect of AI-enabled systems. By systematically addressing explainability requirements, the framework contributes to the cultivation of transparency and trust considerations within the requirement process. This aligns with the growing emphasis on responsible AI practices and ensures that AI-enabled systems meet ethical standards and user expectations.

With the proper implementation of appropriate explanations, users can gain a better understanding of the workings and behaviour of AI-enabled systems. Beyond mere comprehension, these explanations serve as a built-in alert mechanism for users, prompting them to inspect the alignment between their decisions and the system's prediction. If there is any misalignment, users are delegated to initiate a re-evaluation process, cross-checking and verifying results to ensure congruence. For instance, during the interview, the participant who plays the physician role mentioned that in cases where the AI prediction deviates from the physician's decision, it acts as a trigger for the physician to instinctively reverify the supporting evidence for diagnosing a patient with heart disease. The explanation provided by the AI-enabled system can be likened to having a junior medical doctor justifying their decision. This highlights the role of explanations as valuable supporting evidence that not only explains the internal AI decision-making process or recommendation for improvement, but also solidifies the users' decisions through informed justification.

In a broader context, the proposed framework offers systematic requirements analysis procedures and process practices to SE and AI practitioners. The structured approach employed in the framework not only streamlines the identification and incorporation of explainability requirements but also provides step-by-step procedural best practices. This has the potential to elevate the overall quality and efficiency of requirements analysis at the intersection of SE and AI.

5.3 Limitations

The research evaluation is comprehensive in its approach, but it is not without its limitations. First and foremost, it is crucial to acknowledge that the dataset employed in this research is an open dataset. While it serves the specific purpose of addressing user interface and user experience, it does not encompass the entirety of potential datasets that may be encountered in real-world scenarios. Additionally, the dataset comprises only structured data, but excludes text or image data. Therefore, the explanations were not created for texts and images, and they were not able to carry out the evaluation accordingly. During the evaluation, the participants prioritized the accuracy and relevance of the prediction outcomes; however, it is important to remind the participants to concentrate on explanation designs and presentations instead of the correctness of prediction outcomes, so that the process of collecting explainability requirements could be conducted more effectively. Despite the applicability and effectiveness of the proposed framework discussed previously, this framework primarily focuses on post-hoc explanations, lacking in exploring ad-hoc scenarios and dynamic explanations.

5.4 Threats to validity

This section discusses the countermeasures taken to mitigate the external, internal, and construct threats that may arise when conducting the case study on personalized explainability requirement analysis in AI-enabled systems that could undermine the reliability and credibility of research findings. Regarding threats to internal validity, it is possible to introduce bias through our qualitative analysis, wherein the chosen sample of stakeholder types may not fully represent the diverse range of perspectives and needs within the broader stakeholder community. In addition, the dynamic nature of AI technologies may introduce temporal threats to validity, as the effectiveness and the accessibility of the explainability tools to generate the explanation prototypes may evolve or depreciate over time.

From another point of view, the threats to external validity arise from our data preparation and sampling methodology, particularly when we are utilizing online open-source data for our case study. Despite analyzing a substantial volume of data, the predictive outcomes may deviate from real-world scenarios due to the lack of geographical specificity in the open-source data. Variations in geolocation can significantly impact the characteristics of the data, potentially leading to discrepancies in prediction accuracy. For instance, while the target variable for prediction may be balanced, such as in the case of heart disease risk, the distribution of certain attributes within the data, such as smoking status, may not accurately reflect real-world scenarios. In the dataset, the ratio of smokers to non-smokers is approximately 3:7, with non-smokers comprising 70% of the data. This imbalance could introduce bias, potentially resulting in inaccurate predictions, such as misclassifying non-smokers as high-risk heart disease patients. Most importantly, these discrepancies underscore once again the need for explanation provided and the framework proposed when we are utilizing the prediction of AI-enabled systems to make decisions.

Construct validity concerns the operational measures that are interpreted the same way by researchers and case study participants. Two researchers were involved throughout the case study to assist participants in understanding the proposed method and questionnaires used in this study. An online session was arranged to explain the proposed method and questionnaires to the participants before sending the materials to them. They asked questions during the presentation to clarify their doubts. This is to mitigate the construct validity threats and ensure that both participants have the same understanding of the proposed method and questionnaires used in this study. Both participants were free to give their feedback on the proposed method and case study results without interruptions from the researchers.

The causal relations and design of the case study may affect the internal validity to reach the right conclusion for this study. To limit the threats to internal validity, questionnaires and an interview were used as data collection methods. Two researchers conducted the interview together, and the interview session was recorded. The data collected using questionnaires and interviews were analyzed by both researchers. The selection results of the two projects were analyzed by one of the researchers and cross-checked by another researcher. More than one project was used in this case study for evaluation to mitigate the threats to the internal validity of the proposed method.

6.0 CONCLUSION

Explainability is one of the important ethical principles to support transparency that aims to justify, control, improve, and discover more information about the prediction outcomes of AI-enabled systems. It allows the AI-enabled systems to explain themselves and enhance the human understanding of the outcomes produced by the systems, and this can improve the acceptability of humans to utilize the predictions for subsequent actions, including decision-making and system improvements. Apart from interpretable AI, explainability focuses on the level of understandable components provided by the systems. Undoubtedly, we should consider internal and external stakeholders with and without an AI background and cognitive limitations, so that personalized explanations could be generated.

Systematic requirement analysis for explainability is important in establishing a clear understanding of objectives and requirements within an AI-enabled system. The objective of explainability is to make the system's behaviour transparent, fostering user comprehension and trust in the system's decisions. Recognizing various types of explanations becomes pivotal, allowing personalization of detail and context to meet the diverse needs of stakeholders, thereby enhancing their confidence in AI solutions. The goodness of explanations is crucial to investigating which claims are accurate and correct through user evaluation. The formal documentation of explainability requirements serves as a reference for the development team, ensuring that the explanations align with the validated requirements, hence reinforcing the integrity of the development process.

Shifting from a theoretical framework, the applicability and effectiveness of the personalized explainability requirement analysis framework were evaluated with a case study in the healthcare domain. The AI risks of the AI-enabled heart disease risk prediction system are associated with medium, high, or very high risks during the domain analysis, and detailed explanations are required for this AI-enabled healthcare system. In fact, the explanations were identified as relatively significant because they impact human lifestyles and health conditions. Developing the AI-enabled heart disease risk prediction system involves four main stakeholders, namely the AI developer (Development team), Senior consultant cardiologist (Subject matter expert), Physician (Decision maker), and Patient (Affected users), who may be directly or indirectly affected by the outcomes of risk prediction. Their unique needs and goals from explanations are important so that the user-centric explanations can assist them in decision-making and provide relevant recommendations. Bridging the gap for non-technical stakeholders through customized user stories, along with translating these stories into formal requirements for technical stakeholders, the proposed framework ensures user-centric explanations in both layman's and technical terms. Nonetheless, strengthening the design and development phases with well-defined documentation minimizes misalignment and the need for reclarification in subsequent stages.

In conclusion, the misalignment and misunderstanding between requirement engineers and stakeholders when defining the explainability requirements were able to be resolved by the proposed solution. The solution incorporates expert input, user stories, explanation prototypes, user evaluations, and feedback to effectively address the problem. By following the outlined step-by-step stages, the solution ensures that the explanation needs and goals of all stakeholders are thoroughly understood and integrated into the system in the initial phases. Finally, this proposed framework ensures that the solution is both user-centric and aware of requirement priority, ultimately leading to a more effective and satisfactory outcome for all stakeholders involved.

Future research work could explore the quantitative measurement of the degree of explainability needed in an AI-enabled system. Ad-hoc scenarios and dynamic explanations can be explored in future research. Furthermore, the specific explainability needs and goals of regulators could be further investigated as they involve unclear rules and regulations in a particular domain.

ACKNOWLEDGEMENT

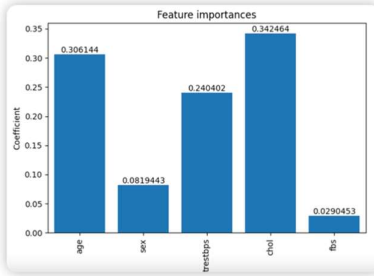
This research was funded by the Ministry of Higher Education Malaysia (MOHE)'s Fundamental Research Grant Scheme (FRGS), Project Code: FRGS/1/2020/ICT01/UM/02/3.
We want to thank all the case study participants in this study.

Appendix: Personalized explanation prototypes for decision makers (Physicians)

This appendix includes the personalized explanation prototypes and requirements generated based on the four user stories collected from the decision maker, the physician, in Stage 3 Explainability analysis. These prototypes were presented to the physician to give feedback and satisfaction scores (refer to Table 12). Each requirement was prioritized using one of the MoSCoW priority groups, which are: Must-Have (M), Should-Have (S), Could-Have (C), and Won't-Have (W).

Table 13: Plain-fact explanation prototype and requirement for user story 1

<table>
<tr><td>User Story 1</td><td>As a physician, I want to understand what are the top 2 features or risk factors utilized by the system before using the system, so that I can better integrate the system's outcomes into my clinical decisions.</td></tr>
<tr><td>Plain-fact explanation prototype</td><td>



A bar chart plotted to show the feature importance used by the prediction model. The highest coefficient of the features: Cholesterol level & Age, as shown
i. Cholesterol (chol) - 0.3425
ii. Age (age) – 0.3061
</td></tr>
<tr><td>Personalized explainability requirement</td><td>An AI-enabled heart disease system must be explainable physician from an overall perspective before using the system with respect to better integrate the system's outcomes into physician's clinical decisions regarding to top 2 features or risk factors utilized by the system by providing textual and table explanation.</td></tr>
<tr><td>Priority</td><td>Must-Have (M)</td></tr>
</table>

Table 14: Causal explanation prototype and requirement for user story 2

<table>
<tr><td>User Story 2</td><td>As a physician, I want to understand why the heart disease prediction system classified a patient as high risk after using the system, so that I can identify the underlying factors that contributed to the risk assessment.</td></tr>
<tr><td>Causal explanation prototype</td><td>
Patient ID: 64
Predict: High risk
Actual: High risk
The model is 83.0% confident this patient is high risk.
The actual outcome and predicted outcome is the same.
+ coefficient: contribute positively to the prediction outcome
- coefficient: contribute negatively to the prediction outcome
<table>
<tr><th>Feature</th><th>Patient Data</th><th>Range</th><th>Coefficient (+/-)</th></tr>
<tr><td>Systolic Blood Pressure</td><td>160.0</td><td>systolic_bp > 140.00</td><td>+0.3187</td></tr>
<tr><td>Diastolic Blood Pressure</td><td>110.0</td><td>diastolic_bp > 90.00</td><td>+0.1232</td></tr>
<tr><td>Age</td><td>60.17</td><td>age > 58.45</td><td>+0.1221</td></tr>
<tr><td>BMI</td><td>50.0</td><td>bmi > 30.35</td><td>+0.0758</td></tr>
<tr><td>Smoke</td><td>Non-smoker</td><td>smoke <= 0.00</td><td>-0.0053</td></tr>
<tr><td>Gender</td><td>Female</td><td>gender <= 1.00</td><td>-0.0095</td></tr>
</table>
</td></tr>
<tr><td>Personalized explainability requirement</td><td>An AI-enabled heart disease system must be explainable to physician from an individual perspective after using the system with respect to identify the underlying factors that contributed to the risk assessment regarding to classification of a patient as high risk by providing textual and table explanation.</td></tr>
<tr><td>Priority</td><td>Must-Have (M)</td></tr>
</table>

Table 15: Transfactual explanation prototype and requirement for user story 3

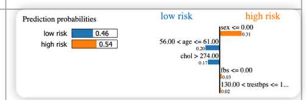
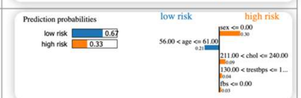
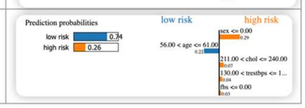
<table>
<tr><td>User Story 3</td><td>As a physician, I want to understand what if a patient's cholesterol level was lower after using the system, so that I can assess the potential impact of lifestyle changes on the heart disease prediction and advise the patient accordingly.</td></tr>
<tr><td>Transfactual explanation prototype</td><td>
<u>Rationale.</u> Example of 3 similar risk profile (i.e. age, resting blood pressure, and fasting blood sugar level) of the same patient, by lowering the cholesterol level leads to change of prediction result (High to Low risk).
<table>
<tr><th>PID</th><th>Sex</th><th>Age</th><th>Resting blood pressure</th><th>Cholesterol level</th><th>Fasting blood sugar level</th><th>Result</th><th>Confidence level</th><th>LIME technique</th></tr>
<tr><td>84</td><td>Female</td><td>57</td><td>135</td><td>280</td><td><120</td><td>High risk</td><td>54.0%</td><td>
</td></tr>
<tr><td>84</td><td>Female</td><td>57</td><td>135</td><td>240</td><td><120</td><td>Low risk</td><td>67.0%</td><td>
</td></tr>
<tr><td>84</td><td>Female</td><td>57</td><td>135</td><td>235</td><td><120</td><td>Low risk</td><td>74.0%</td><td>
</td></tr>
</table>
</td></tr>
<tr><td>Personalized explainability requirement</td><td>An AI-enabled heart disease system must be explainable physician from an individual perspective after using the system with respect to assess and advise the potential impact of lifestyle changes regarding to the changes of patient’s health metrics by providing table explanation.</td></tr>
<tr><td>Priority</td><td>Should-Have (S)</td></tr>
</table>

Table 16: Counterfactual explanation prototype and requirement for user story 4 (Example 1)

<table>
<tr><td>User Story 4</td><td>As a physician, I want to understand how to adjust a patient’s modifiable risk factors after using the system, so that I can provide practical recommendations to lower the heart disease risk.</td></tr>
<tr><td>Counterfactual explanation prototype (Example 1)</td><td>
Example 1:
Original prediction: 1 (High risk)
Counterfactual prediction: 0 (Low risk)
Age: 51
Sex: Male
Resting Blood Pressure: 125
Cholesterol level: 213
Fasting blood sugar level: < 120 mg/dl
You may suggest your patient
1. Reduce the Cholesterol level from 213 to 170 mmHg or lower
Recommendation
According the explanation model, your cholesterol level are at high risk.
Since you are above 50 years old, you may reduce your cholesterol level by
1. Minimum 45 min/day, 5 days/week of moderate intensity PA (i.e. 150 min/week) or
2. 15 min/day, 5 days/week of vigorous intensity PA (75min/week) or
3. a combination of both
based on the handbook <u>Primary & Secondary Prevention of Cardiovascular Disease 2017</u> by the Ministry of Health
The AI-enabled system incorporates practical recommendations from the Ministry of Health, Malaysia to provide patients with guidance on managing their health metrics effectively. By aligning with official health guidelines, the AI-enabled system ensures that the advice it offers is consistent with well-established and evidence-based measures for maintaining good health.
</td></tr>
<tr><td>Personalized explainability requirement</td><td>An AI-enabled heart disease system must be explainable physician from an individual perspective after using the system with respect to provide practical recommendations to lower the heart disease risk regarding to adjust a patient’s modifiable risk factors by providing textual explanation.</td></tr>
<tr><td>Priority</td><td>Should-Have (S)</td></tr>
</table>

Table 17: Counterfactual explanation prototype and requirement for user story 4 (Example 2)

User Story 4	As a **physician**, I want to understand **how to** adjust a patient's modifiable risk factors **after using** the system, so that I can provide practical recommendations to lower the heart disease risk.
Counterfactual explanation prototype (Example 2)	**Example 1:** **Age**: 51 **Sex**: Male **Resting Blood Pressure**: 125 **Cholesterol level**: 213 **Fasting blood sugar level**: < 120 mg/dl **Original prediction**: 1 (High risk) **Counterfactual prediction**: 0 (Low risk) You may suggest your patient 1. Reduce the **Cholesterol level** from **213** to **170 mmHg or lower** **Recommendation** According the explanation model, your cholesterol level are at high risk. Since you are above 50 years old, you may reduce your cholesterol level by 1. Minimum 45 min/day, 5 days/week of moderate intensity PA (i.e. 150 min/week) or 2. 15 min/day, 5 days/week of vigorous intensity PA (75min/week) or 3. a combination of both based on the handbook Primary & Secondary Prevention of Cardiovascular Disease 2017 by the **Ministry of Health** The AI-enabled system incorporates practical recommendations from the Ministry of Health, Malaysia to provide patients with guidance on managing their health metrics effectively. By aligning with official health guidelines, the AI-enabled system ensures that the advice it offers is consistent with well-established and evidence-based measures for maintaining good health.
Personalized explainability requirement	An AI-enabled heart disease system must be explainable **physician** from an **individual** perspective **after using** the system with respect to provide practical recommendations to lower the heart disease risk regarding to adjust a patient's modifiable risk factors by providing **textual** explanation.
Priority	Should-Have (S)

REFERENCES

[1] I. Ozkaya, "What Is Really Different in Engineering AI-Enabled Systems? ", *IEEE Software*, Vol. 37, No. 4, 2020, pp. 3-6.

[2] C. Rzepka, and B. Berger, "User interaction with AI-enabled systems: A systematic review of IS research", in *Thirty Ninth International Conference on Information Systems, San Francisco*, 2018, pp. 1-17.

[3] L. M. Cysneiros, M. Raffi, and J. C. S. do Prado Leite, "Software Transparency as a Key Requirement for Self-Driving Cars", in *IEEE 26th International Requirements Engineering Conference, Canada*, August 2018, pp. 382-387.

[4] N. Balasubramaniam et al., "Transparency and Explainability of AI Systems: Ethical Guidelines in Practice", in *Requirements Engineering: Foundation for Software Quality*, Cham, Springer International Publishing, 2022, pp. 3-18.

[5] K. Ahmad et al., "Requirements engineering framework for human-centered artificial intelligence software systems", *Applied Soft Computing,* Vol. 143, 2023, p. 110455.

[6] A. Adadi, and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)", *IEEE Access*, Vol. 6, 2018, pp. 52138-52160.

[7] A. F. Markus, J. A. Kors, and P. R. Rijnbeek, "The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies", *Journal of Biomedical Informatics*, Vol. 113, 2021, p. 103655.

[8] J. Amann et al., "Explainability for artificial intelligence in healthcare: a multidisciplinary perspective", *BMC Medical Informatics and Decision Making*, Vol. 20, No. 310, 2020, pp. 1-9.

[9] N. A. Ghani et al., “Methodical evaluation of healthcare intelligence for human life disease detection”, *Malaysian Journal of Computer Science*, Vol. 36, No. 3, 2023, pp. 208-222.

[10] P. Weber, K. V. Carl, and O. Hinz, “Applications of Explainable Artificial Intelligence in Finance—a systematic review of Finance, Information Systems, and Computer Science literature”, *Management Review Quarterly*, Vol. 74, No. 2, 2024, pp. 867-907.

[11] L. H. Gilpin et al., “Explaining Explanations: An Overview of Interpretability of Machine Learning”, in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Italy*, 2018, pp. 80-89.

[12] R. Guizzardi et al., “Ethical Requirements for AI Systems”, in *Canadian Conference on Artificial Intelligence*, Cham, Springer International Publishing, 2020, pp. 251-256.

[13] A. Olsson, and G. Joelsson, “Requirements Analysis For AI Solutions : A Study On How Requirements Analysis Is Executed When Developing AI Solutions”, *Bachelor's Thesis in Informatics, University of Boras*, 2019.

[14] J. Gerlings, M. S. Jensen, and A. Shollo, “Explainable AI, But Explainable to Whom? An Exploratory Case Study of xAI in Healthcare”, in *Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects*, Cham, Springer International Publishing, 2022, pp. 169-198.

[15] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why Should I Trust You?": Explaining the Predictions of Any Classifier”, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California*, 2016, pp. 1135-1144.

[16] S. M. Lundberg, and S. I. Lee, “A unified approach to interpreting model predictions”, in *Proceedings of the 31st International Conference on Neural Information Processing Systems, California*, 2017, pp. 4765-4774.

[17] F. Sovrano, F. Vitali, and M. Palmirani, “Making Things Explainable vs Explaining: Requirements and Challenges Under the GDPR”, in *International Workshop on AI Approaches to the Complexity of Legal Systems, Frankfurt*, 2021, pp. 169-182.

[18] U. Ehsan et al., “The Who in XAI: How AI Background Shapes Perceptions of AI Explanations”, in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, Honolulu*, 2024, pp.1-32.

[19] D. Wang et al., “Designing Theory-Driven User-Centric Explainable AI”, in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow*, 2019, pp. 1-15.

[20] L. Chazette et al., “How Can We Develop Explainable Systems? Insights from a Literature Review and an Interview Study”, in *Proceedings of the International Conference on Software and System Processes and International Conference on Global Software Engineering, Pittsburgh*, 2022, pp. 1-12.

[21] T. Clement et al., “XAIR: A Systematic Metareview of Explainable AI (XAI) Aligned to the Software Development Process”, *Machine Learning and Knowledge Extraction,* Vol. 5, No. 1, 2023, pp. 78-108.

[22] L. Chazette, W. Brunotte, and T. Speith, “Explainable software systems: from requirements analysis to system evaluation”, *Requirements Engineering,* Vol. 27, No. 4, 2022, pp. 457-487.

[23] T. A. J. Schoonderwoerd et al., “Human-centered XAI: Developing design patterns for explanations of clinical decision support systems”, *International Journal of Human-Computer Studies,* Vol. 154, 2021, p. 102684.

[24] Q. V. Liao et al., “Question-driven design process for explainable AI user experiences”, *arXiv preprint arXiv:2104.03483*, 2021.

[25] D. Cirqueira et al., “Scenario-Based Requirements Elicitation for User-Centric Explainable AI”, in *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*, Cham, Springer International Publishing, 2020, pp. 321-341.

[26] N. Balasubramaniam et al., “Transparency and explainability of AI systems: From ethical guidelines to requirements”, *Information and Software Technology,* Vol. 159, 2023, p. 107197.

[27] K. Ahmad et al., “Requirements practices and gaps when engineering human-centered Artificial Intelligence systems”, *Applied Soft Computing,* Vol. 143, 2023, p. 110421.

[28] Y. Xia et al., “Systemic Immune Inflammation Index (SII), System Inflammation Response Index (SIRI) and Risk of All-Cause Mortality and Cardiovascular Mortality: A 20-Year Follow-Up Cohort Study of 42,875 US Adults”, *Journal of Clinical Medicine*, Vol. 12, No. 3, 2023, p. 1128.

[29] R. Confalonieri et al., “A historical perspective of explainable Artificial Intelligence”, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 11, No. 1, 2021, p. e1391.

[30] J. M. Bryson, “What to do when Stakeholders matter”, *Public Management Review*, Vol. 6, No. 1, 2004, pp. 21-53.

[31] M. Langer et al., “What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research”, *Artificial Intelligence*, Vol. 296, 2021, p. 103473.

[32] A. Preece et al., “Stakeholders in explainable AI”, *arXiv preprint arXiv:1810.00184,* 2018.

[33] M. Cohn, *Agile estimating and planning*, Pearson Education, 2005.

[34] G. Lucassen et al., “Improving agile requirements: the Quality User Story framework and tool”, *Requirements Engineering*, Vol. 21, No. 3, 2016, pp. 383-403.

[35] R. R. Hoffman et al., “Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance”, *Frontiers in Computer Science,* Review, Vol. 5, 2023, p. 1096257.

[36] M. A. Köhl et al., “Explainability as a Non-Functional Requirement”, in *IEEE 27th International Requirements Engineering Conference, Jeju*, 2019, pp. 363-368.

[37] E. Miranda, “Moscow Rules: A Quantitative Exposé”, in *Agile Processes in Software Engineering and Extreme Programming*, Cham, Springer International Publishing, 2022, pp. 19-34.

[38] P. Runeson, and M. Höst, “Guidelines for conducting and reporting case study research in software engineering”, *Empirical Software Engineering*, Vol. 14, No. 2, 2009, pp. 131-164.

[39] A. Al Kuwaiti et al., “A Review of the Role of Artificial Intelligence in Healthcare”, *Journal of Personalized Medicine*, Vol. 13, No. 6, 2023, p. 951.

[40] A. T. Mohamed, S. Santhoshkumar, and V. Varadarajan, “Intelligent deep learning based predictive model for coronary heart disease and chronic kidney disease on people with diabetes mellitus”, *Malaysian Journal of Computer Science,* 2022, pp. 88-101.

[41] D. Lee, and S. N. Yoon, “Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges”, *International Journal of Environmental Research and Public Health*, Vol. 18, No. 1, p. 271, 2021.

[42] T. R. Ramesh et al., “Predictive analysis of heart diseases with machine learning approaches”, *Malaysian Journal of Computer Science*, pp. 132-148.

[43] S. Kapa, “The Role of Artificial Intelligence in the Medical Field”, *Journal of Computer and Communications*, Vol. 11, No. 11, 2023, pp. 1-16.

[44] J. Lötsch, D. Kringel, and A. Ultsch, “Explainable Artificial Intelligence (XAI) in Biomedicine: Making AI Decisions Trustworthy for Physicians and Patients”, *BioMedInformatics*, Vol. 2, No. 1, 2022, pp. 1-17.

[45] A. Barredo Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI", *Information Fusion,* Vol. 58, 2020, pp. 82-115.

[46] A. S. Albahri et al., "A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion", *Information Fusion*, Vol. 96, 2023, pp. 156-191.

[47] A. Kiseleva, D. Kotzinos, and P. De Hert, "Transparency of AI in healthcare as a multilayered system of accountabilities: between legal requirements and technical limitations", *Frontiers in Artificial Intelligence*, Vol. 5, No. 879603, 2022, pp. 1-21.

[48] Y. You et al., "The medical authority of AI: A study of AI-enabled consumer-facing health technology", in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems New York,* ACM, 2021, pp. 1-16.