

HUOM! Tämä on alkuperäisen artikkelin rinnakkaistallenne. Rinnakkaistallenne saattaa erota alkuperäisestä sivutukseltaan ja painoasultaan.

Käytä viittauksessa alkuperäistä lähdettä:

Balasubramaniam, N., Kauppinen, M., Rannisto, A., Hiekkänen, K. & Kujala, S. 2023. Transparency and explainability of AI systems: From ethical guidelines to requirements. *Information and Software Technology*, 159, 107197. <https://doi.org/10.1016/j.infsof.2023.107197>.

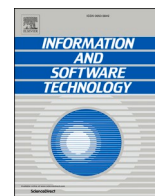
PLEASE NOTE! This is an electronic self-archived version of the original article. This reprint may differ from the original in pagination and typographic detail.

Please cite the original version:

Balasubramaniam, N., Kauppinen, M., Rannisto, A., Hiekkänen, K. & Kujala, S. 2023. Transparency and explainability of AI systems: From ethical guidelines to requirements. *Information and Software Technology*, 159, 107197. <https://doi.org/10.1016/j.infsof.2023.107197>.



© 2023 The authors. Licensed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license (<https://creativecommons.org/licenses/by/4.0/>).



Transparency and explainability of AI systems: From ethical guidelines to requirements

Nagadivya Balasubramaniam^{a,*}, Marjo Kauppinen^a, Antti Rannisto^{a,b}, Kari Hiekkanen^{a,c}, Sari Kujala^a

^a Department of Computer Science, Aalto University, Finland

^b Solita, Finland

^c Haaga-Helia University of Applied Science, Helsinki, Finland

ARTICLE INFO

Keywords:

Transparency
Explainability
Ethical guidelines
Quality requirements
Explainability requirements
AI systems

ABSTRACT

Context and Motivation: Recent studies have highlighted transparency and explainability as important quality requirements of AI systems. However, there are still relatively few case studies that describe the current state of defining these quality requirements in practice.

Objective: This study consisted of two phases. The first goal of our study was to explore what ethical guidelines organizations have defined for the development of transparent and explainable AI systems and then we investigated how explainability requirements can be defined in practice.

Methods: In the first phase, we analyzed the ethical guidelines in 16 organizations representing different industries and public sector. Then, we conducted an empirical study to evaluate the results of the first phase with practitioners.

Results: The analysis of the ethical guidelines revealed that the importance of transparency is highlighted by almost all of the organizations and explainability is considered as an integral part of transparency. To support the definition of explainability requirements, we propose a model of explainability components for identifying explainability needs and a template for representing explainability requirements. The paper also describes the lessons we learned from applying the model and the template in practice.

Contribution: For researchers, this paper provides insights into what organizations consider important in the transparency and, in particular, explainability of AI systems. For practitioners, this study suggests a systematic and structured way to define explainability requirements of AI systems. Furthermore, the results emphasize a set of good practices that help to define the explainability of AI systems.

1. Introduction

The use of artificial intelligence (AI) is changing the world we live in [23]. Algorithmic decision-making is becoming ubiquitous in daily life. Moreover, machine learning is utilized in the crucial decision-making processes, such as loan processing, criminal identification, and cancer detection. [1,18]. The number of organizations that are interested in developing AI systems are increasing. However, the black-box nature of AI systems has raised several ethical issues [3].

To handle the ethical issues of AI and to develop responsible AI systems, various interest groups across the world (e.g., IEEE, ACM) have defined comprehensive ethical guidelines and principles to ensure

responsible AI usage. The AI ethical guidelines developed by three established expert groups [16,20,25] emphasized transparency and explainability for developing AI systems. Frost et al. [31] and Jobin et al. [32] reviewed 47 and 84 AI ethical guidelines reports respectively. The results of these AI ethical guidelines reviews indicated transparency as the most covered ethical guideline and explainability was portrayed as the key scope of transparency. Moreover, organizations have defined their own AI ethical guidelines that encompass the ethical issues which are prominent to the organization [3].

Organizations utilize different machine learning models and algorithms in the decision-making processes. Moreover, the outputs and the decisions of AI systems are usually difficult to understand and lack

* Corresponding author.

E-mail addresses: nagadivya.balasubramaniam@aalto.fi (N. Balasubramaniam), marjo.kauppinen@aalto.fi (M. Kauppinen), antti.rannisto@aalto.fi (A. Rannisto), kari.hiekkanen@aalto.fi (K. Hiekkanen), sari.kujala@aalto.fi (S. Kujala).

<https://doi.org/10.1016/j.infsof.2023.107197>

Received 22 July 2022; Received in revised form 22 February 2023; Accepted 6 March 2023

Available online 8 March 2023

0950-5849/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

transparency [8]. Recent studies [6,8] have highlighted explainability as a key requirement of AI systems that improves transparency. In addition, a study [2] on RE techniques and an industry guideline for building AI systems emphasized that explanations of AI systems enforced trust and improved the decision making of users when using AI systems.

Transparency and explainability are identified as key quality requirements of AI systems [6,8,13] and are portrayed as quality requirements that need more focus in the machine learning context [18]. Explainability can impact user needs, cultural values, laws, corporate values, and other quality aspects of AI systems [6]. The number of papers that deal with transparency and explainability requirements have recently increased. However, studies on how to define explainability and transparency requirements of AI systems in practice are still rare and at their early stage.

This study consisted of two phases. The goal of Phase 1 was to explore **what ethical guidelines organizations have defined for the development of transparent and explainable AI systems**. In this study, we analyzed the ethical guidelines of AI published by 16 organizations to understand what quality requirements these organizations have highlighted in their ethical guidelines. Then, we performed a detailed study focusing especially on transparency and explainability guidelines to delineate the different components of explainability requirements of AI systems. The main contributions of Phase 1 are the model of explainability components and the template for representing individual explainability requirements. The results of Phase 1 were originally published as a conference paper [29]. The main contributions of Phase 1 of this study are:

- The analysis of the AI ethical guidelines of 16 organizations provides insights into important quality requirements of AI systems. The analysis also reveals the reasons why the organizations consider transparency and explainability as critical quality requirements of AI systems.
- Based on the analysis of the transparency and explainability ethical guidelines and the definition of the explainability requirement [6], we propose a model of explainability components. The model contains examples of what these explainability components can be in practice. The purpose of the model is to support practitioners in identifying the explainability needs of different stakeholders.
- We suggest a template for representing individual explainability requirements. This template is a requirements engineering practice that can assist practitioners in defining individual explainability requirements in a structured way.

Phase 2 was an extension of our work done in Phase 1. The goal of this extension was to evaluate the results of Phase 1 and investigate **how explainability requirements can be defined in practice**. To answer this question, we conducted an empirical study. In this study, we first used the model of explainability components with practitioners in a workshop to identify explainability needs for a recruitment game system. Further, we used the results of the workshop and the template proposed in Phase 1 to define the explainability requirements for the recruitment system. Finally, we analyzed what can be learned from applying the model of explainability components and the template for explainability requirements in practice. The main contributions of Phase 2 are:

- This paper summarizes the key lessons we learnt from the evaluation of the model of explainability components and the template for defining explainability requirements of an AI system. These lessons highlight the important prerequisites for the definition of explainability requirements.
- To make the lessons learned concrete for practitioners, we also reformulated them into good practices. The purpose of these good practices is to help practitioners to define explainability requirements of AI systems.

This paper is organized as follows. Section 2 describes the related work on transparency and explainability as quality requirements of AI systems. In Section 3, we present the research method used in the Phase 1 and Phase 2 of this study. Section 4 describes the results from the analysis of the ethical guidelines and introduces the model and template for defining explainability requirements of AI systems. In Section 5, we first present how the model and template were used to define the explainability requirements, and then we summarize the lessons learned from the empirical study. We discuss our results and the validity threats and limitations of our study in Section 6 and Section 7 respectively. Finally, Section 8 concludes the paper.

2. Related work

In what follows, we first emphasize the definition of ethical requirements of AI systems and the close association of ethical guidelines to requirement definition. Next, we focus on transparency and explainability which are emerging quality requirements of AI systems.

2.1. Ethical requirements of AI systems

Guizzardi et al. [17] introduced and defined ethical requirements of AI systems as *'Ethical requirements are requirements for AI systems derived from ethical principles or ethical codes (norms)'*. Besides, the authors highlighted that defining the ethical requirements at the beginning of AI system development helps in considering the ethical issues during the early phases of development. Generally, ethical requirements of AI constitute both functional and quality requirements derived from the stakeholder needs in accordance with ethical principles [17,24]. The studies on ethical requirements have depicted the close association of ethical guidelines with the definition of requirements.

2.2. Transparency as a quality requirement

Cysneiros [11] and Leite and Capelli [14]'s studies classified transparency as an impactful non-functional requirement (NFR) of the software system. Further, the authors delineated the interrelationship of transparency with other NFRs, such as trust, privacy, security, accuracy, etc. through softgoal interdependence graphs (SIGs).

In addition, the dependency between transparency and trust is a salient facet that needs to be considered in system development, such as self-driving cars [5,13]. Kwan et al. [21] developed an NFR catalog for trust, and the study reported that transparency positively impacted in achieving users' trust, which was portrayed as the key corporate social responsibility (CSR) principle.

Recent studies [12,13,18,19] have discussed transparency as a key NFR in machine learning and autonomous systems. Transparency in AI systems has been identified as quintessential, but the black box nature of AI systems makes the definition of transparency requirements challenging [13,19]. Horkoff [19] emphasized the real-world impact of machine learning and the crucial question *'how these results are derived?'*. Likewise, Chazette et al. [7] highlighted that transparency as an NFR is abstract and requires better understanding and supporting mechanisms to incorporate them into the system. Explanations of machine learning and AI results have been proposed to mitigate the issues of transparency [7,19]. The studies [7,8] on the relationship between explanations and transparency of AI systems proposed explainability as an NFR.

Explainability suggested as an NFR had been linked to other NFRs such as transparency and trust by [6]. As Köhl et al. [22] link explainability to transparency, and Chazette et al. [7,8] also report that explainability aims in achieving better transparency. Moreover, explanations of AI systems had been identified to contribute higher system transparency. For instance, receiving explanations about a system, its processes and decisions impact both understandability and transparency NFRs [6].

2.3. Explainability as a quality requirement

Köhl et al. [22] addressed the gap in ensuring explainability in system development and performed a conceptual analysis of systems that needs explanations (e.g., automated hiring system). The analysis aimed to elicit and specify the explainability requirements of the system. The authors proposed definitions for three questions: 1) to **who** are the ‘explanations for’ focusing on understandability, context, and target of the system, 2) **when** the system is considered explainable, and 3) **how** to define explainability requirements.

Köhl et al. [22] and Chazette et al. [6] proposed definitions to help understand what explainability means from a software engineering perspective (Table 1). The definition of the explainability requirement by Chazette et al. [6] is based on the definition proposed by Köhl et al. [22]. Both of these definitions have the following variables: a system, an addressee (i.e., target group), an aspect, and a context. In addition to these variables, Chazette et al. [6] have also included an explainer in their definition of explainability.

Chazette et al. [7,8] discussed explainability as an NFR and interlinked it with transparency. Further, explainability supports in defining the transparency requirements which impacts software quality. The authors also identified that end-users are more interested to get explanations during adverse situations, and they are least interested to know the inner working of the system i.e., how the system worked [7,8]. In addition, [6,8,22] highlighted the tradeoffs between the explainability and other NFRs. Consequently, [6] indicated that when eliciting the explainability requirements, consideration of positive and negative impacts of explanations to the users could avoid conflict with transparency and understandability NFRs.

Subsequently, Chazette et al. [6] featured explainability as an emerging NFR and evaluated how explainability impacts other NFRs and qualities. Their study revealed that transparency, reliability, accountability, fairness, trustworthiness, etc. are positively impacted by explainability. However, the authors acknowledged that studies on incorporating explainability in the software development process are in its early stage and need more research [6].

3. Research methodology

3.1. Phase 1: analysis of ethical guidelines of sixteen organizations

The goal of Phase 1 of this study was to investigate *what ethical guidelines organizations have defined for the development of transparent and explainable AI systems*. In the analysis of the AI ethical guidelines, we used the following research questions:

- *What quality requirements do organizations highlight in their AI ethical guidelines?*
- *Which components of the explainability requirements can be found in the AI ethical guidelines of organizations?*
- *How do transparency and explainability relate to other quality requirements in the AI ethical guidelines of organizations?*

Our selection criterion was to find organizations that have defined and published their ethical guidelines for using AI. In late 2018, AI

Table 1
Definitions of explainability requirement and explainability.

Köhl et al. [22]	Chazette et al. [6]
A system <i>S</i> must be explainable for target group <i>G</i> in context <i>C</i> with respect to aspect <i>Y</i> of explanandum <i>X</i> .	A system <i>S</i> is explainable with respect to an aspect <i>X</i> of <i>S</i> relative to an addressee <i>A</i> in context <i>C</i> if and only if there is an entity <i>E</i> (the explainer) who, by giving a corpus of information <i>I</i> (the explanation of <i>X</i>), enables <i>A</i> to understand <i>X</i> of <i>S</i> in <i>C</i> .

Finland, which is a steering group in-charge of AI program, organized the ‘Ethics Challenge’. The challenge invited enterprises in Finland to develop ethical guidelines of AI as a way to promote the ethical use of AI. We identified 16 organizations that have published their ethical guidelines. We gathered the documents from the organizations’ websites and those documents contained data such as AI ethical guidelines and their explanations as simple texts, detailed PowerPoint slides set, and videos explaining the guidelines.

First, we classified the organizations that have published the ethical guidelines of AI into three categories: professional services and software, business-to-consumer (B2C), and public sector. Table 2 summarizes these categories.

Category A includes seven professional services organizations that provide a broad range of services from consulting to service design, software development, and AI & analytics. The two software companies in Category A develop a large range of enterprise solutions and digital services. The five B2C organizations represent different domains: two telecommunication companies, a retailer, a banking group, and an electricity transmission operator. The public sector organizations represent tax administration and social security services. The six companies of Category A are Finnish and the other three are global. Furthermore, all the organizations of Category B and C are Finnish.

We started the data analysis process by conceptual ordering [10] where the ethical guidelines of AI in 16 organizations were ordered based on their category name. Then, the categories which were also quality requirements of AI were identified by line-by-line coding process [4]. This process was performed by the first author and was reviewed by the second author. Next, we performed the word-by-word coding technique and we focused on transparency and explainability guidelines in this step. We used Charmaz’s [4] grounded theory techniques on coding and code-comparison for the purpose of data analysis only.

The first two authors of this paper performed separately the initial word-by-word coding. The analysis was based on the variables used in the definition of explainability by Chazette et al. [6]. These variables were addressees of explanations, aspects of explanations, contexts of explanations, and explainers. We also analyzed reasons for transparency. Discrepancies in the codes were discussed and resolved during our multiple iterative meetings, and missing codes were added. Table 3 shows examples of ethical guidelines and codes from the initial word-by-word coding process. Next, in the axial coding process, the sub-categories from the initial coding process were combined or added under the relevant high-level categories. The quality requirements that are related to transparency and explainability were combined and the second author reviewed the axial coding process.

3.2. Phase 2: evaluation of the results of Phase 1 in practice

This section first summarizes the research process of Phase 2. We then justify the case and AI system selection. Finally, we describe the data collection and data analysis processes.

3.2.1. Research process of the empirical study

The goal of Phase 2 was to evaluate the results of Phase 1 with practitioners and investigate **how explainability requirements can be defined in practice**. The following two research questions were used to guide our empirical evaluation:

Table 2
Overview of the organizations of the Phase 1 study.

Category	No. of Organizations	Identifications
Category A: Professional services and software	9	O1-O9
Category B: Business-to-Consumer (B2C)	5	O10-O14
Category C: Public sector	2	O15 and O16

Table 3
Example codes of the initial word-by-word coding process.

Example lines of ethical guidelines	Examples of codes
We tell our customers in a clear and understandable way where, why, and how AI has been utilized.	Addressees – Customers Relationships – Understandability
Their input, capabilities, intended purpose, and limitations will be communicated clearly to our customers.	Addressees – Customers Aspects – Input, Capabilities, Purpose, and Limitations
Ensure AI transparency. To build trust among employees and customers, develop explainable AI that is transparent across processes and functions.	Reasons for transparency – Trust Addressees– Employees and customers

- How can the model of explainability components be used for defining explainability requirements of an AI system?
- How can the template be used for representing explainability requirements of an AI system?

Wohlin and Runeson [36] propose three research methodologies for industry–academia collaborative research in software engineering and, especially, for empirical evaluations of research solutions. These methodologies are Action Research, Design Science, and the Technology Transfer Research Methodology. All three align to the general and iterative engineering research process of describe-solve-practice, where practice refers to the evaluation and practical use of the solution [36]. In this empirical study, we selected the Technology Transfer Research Methodology for two main reasons. First, this research methodology emphasizes gradual empirical evaluations of solutions. Secondly, there are also concrete guidelines on how these evaluations can be done as industry-academia collaboration [34,35,36].

The Technology Transfer Research Methodology is an adaptation of the Technology Transfer Model formulated by Gorcheck et al. [34]. This model consists of seven steps with a strong focus on academia-industry collaboration, where the research problem originates from industry needs and challenges, and the developed solution go through several evaluation steps to ensure that it can be used in practice with low risk [34,36]. Wohlin and Runeson [36] reformulate the seven steps of the Technology Transfer Model and they propose the following six activities to be used as the research process of the Technology Transfer Research Methodology (TTRM):

- 1 Identify the industrial challenge
- 2 Assess practice and formulate a research objective
- 3 Study the state-of-the-art
- 4 Develop one or more candidate solution(s)
- 5 Evaluate the solution(s)
 - a In academic setting – “To minimize risk before transferring a solution to industry, it can be validated in an academic setting, for example through experimentation or simulation” [36]
 - b Static evaluation – “The static validation is done through seminars and discussions with the key stakeholders to anchor the proposed solution” [36]
 - c Pilot evaluation – “The pilot should be as representative as possible of the regular context in which the solution is typically used to ensure that the solution fits well with the current practices” [36]
- 6 Move the chosen solution into practice and evaluate

We conducted the first three activities of TTRM iteratively. First, we analyzed the AI ethical guidelines of three companies and interviewed three practitioners in order to understand what kind of ethical guidelines companies have defined for solving potential ethical issues of AI and for developing AI systems. As part of this case study, we also conducted a literature review on potential ethical issues of AI systems and compared it to the AI ethical guidelines of three established expert

groups [16,20,25]. The key finding of our case study was that the companies focus on solving potential ethical issues such as accountability, explainability, fairness, privacy, and transparency [3].

In the second iteration, we focused on transparency and explainability of AI systems because the results of the first iteration indicated that transparency and explainability are critical quality requirements of AI systems [3]. To understand the current state of the art, we conducted a literature review on transparency and explainability requirements. In particular, the research results of Chazette et al. [6] served as a source of inspiration for us. We used their definition of the explainability requirement when we analyzed the AI ethical guidelines of 16 organizations as described in Phase 1 of this paper.

The **development of a candidate solution**, which is the fourth research activity of TTRM, started when we analyzed the AI ethical guidelines of 16 organizations and discovered concrete examples of explainability components from these guidelines. We realized that these concrete examples can help practitioners define explainability requirements. Based on the literature review and analysis of the ethical guidelines of 16 organizations, we developed two candidate solutions, which are the model of explainability components and the template for representing explainability requirements. The first versions of them were published in the conference paper [29].

We started the evaluation of the model of explainability components and template for explainability requirements from **static evaluation**. We did not evaluate them in an academic setting because practitioners showed initial interest in them. We were invited to the seminar of a research project that focused on AI governance and auditing (AIGA). In the seminar, Author 1 introduced the first versions of the model and template. Author 2 also participated in the seminar and discussion about the model and template. After the seminar, a data scientist from Solita (the organization of this empirical study) contacted us. We organized an informal discussion with the data scientist to understand the current state of practice.

Next, we decided to conduct a **small-scale pilot evaluation**. We approached Solita again and identified a person (Author 3) who helped us to select an AI-based system and important stakeholders that can participate in the pilot. Then, we conducted three interviews to understand the purpose and the role of the selected AI system. After the interviews, we (Authors 1–3) planned and organized a workshop. We organized a workshop with a multi-disciplinary team to elicit the explainability needs of stakeholders for the selected AI system. The model of explainability components guided the work of the multidisciplinary team. After the workshop, we analyzed the data collected from it. Based on this analysis, we defined a set of explainability requirements using the template. The results of this research activity are reported in Section 5.2.

Finally, we analyzed how the model of explainability components, and the template assisted in defining the explainability requirements of the AI system. We described the results of this analysis as lessons learned as recommended by Wohlin and Runeson [36]. According to them, the learning gained from the research is documented as lessons learned in TTRM. The lessons we learned from the empirical evaluation are described in Section 5.3. and, based on them, we suggest six good practices that can support practitioners in defining explainability requirements.

Three authors of this paper participated actively in Phase 2 and conducted the research activities together. Authors 1 and 2 are researchers and Author 3 is a practitioner from the case organization. Here, we summarize their roles in the evaluation phase:

- Author 1 was responsible for planning the interviews and workshop. She was the interviewer of the three interviews and one of the facilitators of the workshop. The author was also in charge of transcribing the interviews and the workshop. Then, she analyzed the interviews and workshop transcript and organized review meetings

with Author 2 to clarify the codes and categorizations. In addition, she reported the results of Phase 2.

- Author 2 reviewed the interview questions and participated in two out of the three interviews. The researcher also reviewed the workshop plan and participated in the workshop as an observer and asked follow-up questions when needed. She analyzed the interviews and workshop transcripts separately and participated in multiple review meetings to discuss the findings. Further, she reviewed the reported results and provided detailed feedback.
- Author 3 identified potential workshop participants with Author 1 and invited them that have different roles in the recruitment and the development of the recruitment game system. He reviewed the workshop plan and was one of the facilitators of the workshop. He also analyzed the transcript of the workshop and wrote an analytical reflection of the most important observations. Author 3 also discussed the most important observations with Author 2. Finally, Author 3 reviewed the results of Phase 2 that are reported in this paper and provided overall feedback.

3.2.2. Case and AI system selection

Phase 2 of our study was conducted with Solita. The reason for selecting this organization was that they were interested in investigating the explainability of their AI systems. Solita is Organization O7 from Phase 1, and it defines itself as a technology, data, and design company based in Finland. The company operates in six European countries and comprises of over 1500 experts. In addition, it is one of the biggest data consulting houses in the Nordic countries with around 600 data scientists and data engineers actively involved in developing applications. Further, the organization’s interest to explore the possibilities of incorporating transparency and explainability in AI systems development aligned well with our research goal.

Initially, when the discussion to collaborate started the company representative (Author 3) suggested three potential AI applications. Two of the suggested AI applications were from healthcare domain. The AI system that the company representatives selected for our study was the recruitment game system. The key reason for selecting the recruitment game system was that the company representatives had identified the opportunity to develop it further and possibly use it as part of the recruitment process. In addition, the system is in its early prototype stage, which indicates the opportunity to perform requirements elicitation to ensure explainability of AI systems. Hence, we decided to apply the model of explainability components and the template for defining explainability requirements of the recruitment game system.

3.2.3. Description of the recruitment game

The two potential goals of the recruitment game system are as follows: 1) for job applicants, the system proposes suitable open job positions based on their scores from playing the game and their educational background and work experience, and 2) for recruiters, the system helps in selecting suitable candidates based on their skills.

The starting point of the recruitment game system consisted of a game where the players are given four different programming tasks. When the player has completed the tasks, they get points for their answers. Then, they are directed to a form to fill in their educational background and work experience. Next, an AI uses the game scores and the background of the applicants to suggest the open positions available at Solita that match their skills and interests. For recruiters, they can view the game scores and the background of the applicants and see what the suggested job positions for the applicants are. Then, the recruiters can review the applicants and select suitable persons for the open positions at Solita.

3.2.4. Data collection

The core of the research collaboration with Solita consisted of three interviews and a workshop session. Table 4 summarizes the practitioners that participated in Phase 2. The goal of the three interviews was to

Table 4
Practitioners of Phase 2.

Data Source	Role	Experience
Interview 1	Software Developer	3 years in software development
Interview 2	Lead Ethnographer	7 years in consumer research 6 years in service design
Interview 3	Head of Research	About 30 years in the ICT sector
Workshop	Director of Talent Acquisition	16 years in recruitment
	Recruiter in Belgium	10 years in recruitment
	Recruiter in Finland and	2 years in recruitment
	Software Developer	3 years in software development
	People Lead and	15 years in software development
	Senior Software Designer	Several months in recruitment
	Head of People Lead	16 years in software development 16 years in recruitment
	Software Designer	<1 year in software development
	Software Designer	2 years in software development

obtain an overview of the organization’s viewpoint on transparency and explainability and to get an understanding of the recruitment game.

The key criterion for selecting the interview participants was that they had knowledge of the recruitment game system. Moreover, it was important that the participants represented different viewpoints on the system, such as system and strategy planning, service and system design, and system development. The interviews were aimed to gather details on the following aspects of the recruitment game system: why the recruitment game was developed, who developed the game, how the game works, who developed the AI part, and how the AI works. In addition, we asked what explainability means to the interviewees.

The interviews were open-ended and semi-structured. The interviews lasted from 45 to 90 mins. One of the three interviews was held face-to-face, while the others were organized virtually using Zoom. Two researchers (Author 1 and Author 2) conducted two interviews together. One interview was conducted by one researcher, because the other researcher (Author 2) supervised the interviewee’s master’s thesis, and the thesis related the development of recruitment game. The interviews were recorded.

The workshop session was conducted after the interviews. The purpose of the workshop was to elaborate on the explainability perspective of the AI system and to create discussion among the practitioners and between the practitioners and researchers. The workshop session was planned by the first author and was reviewed by one senior researcher (Author 2) and one practitioner from Solita (Author 3). Fig. 1 illustrates the outline of the workshop session.

The number of practitioners in the workshop session was seven, and the workshop lasted two hours. The workshop participants formed a multi-disciplinary team that provided diversified viewpoints for the recruitment game system. Five participants were recruiters and two were the developers of the recruitment game. One researcher and one practitioner from Solita acted as facilitators, and one researcher was an observer who also asked follow-up questions when needed.

During the session, we utilized the Miro board application where the participants jotted down their ideas which were then elaborated during the discussion. First, the participants general view on transparency and explainability of AI system was discussed for warm-up purpose. Then, the facilitators introduced the system and provided an illustration about the working of the recruitment game system. Next, the positive and negative aspects of using the AI system in the recruitment were discussed. Subsequently, the model of explainability components was introduced and explained to the participants. Then, the participants spent an intensive 80 mins dedicated to defining the addressees, aspects, contexts, and explainers of explanations. The workshop session was recorded and transcribed.

3.2.5. Data analysis

All the interviews in Phase 2 were transcribed based on the recordings. The interview transcripts were analyzed by the first author.

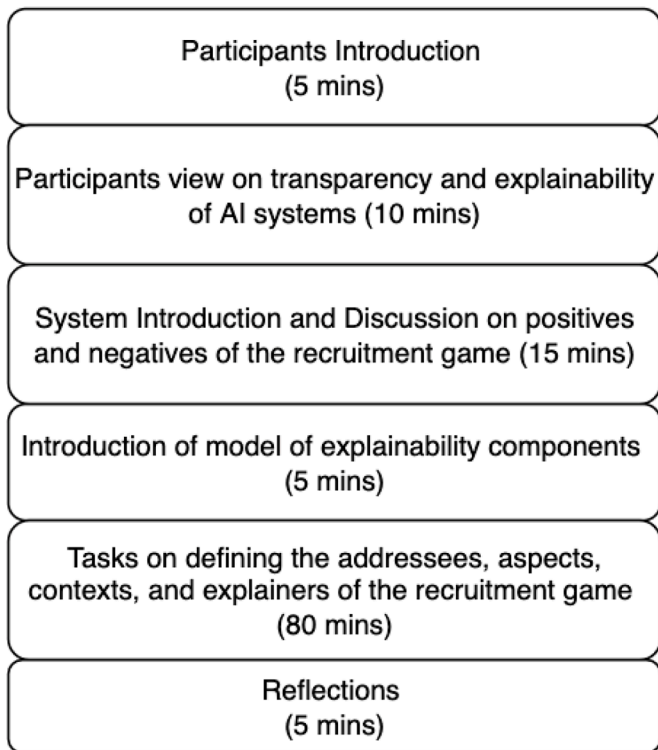


Fig. 1. Structure of the workshop (2 h).

The key themes that were identified when analyzing the transcripts were: potential goals of the recruitment game, main functionalities of the recruitment game, people involved in game development, behavior of the system, role of AI, data used by AI, and AI functionality. In addition, the analysis helped in understanding the interviewees viewpoint on transparency and explainability of AI systems. The results of this analysis were used when planning the workshop.

The workshop transcript was coded by two researchers (Author 1 and 2). In addition, the authors did word-by-word coding of the contents on the Miro board application as well. For Phase 2, we followed the same analysis process as Phase 1. The first two authors performed word-by-word coding separately. We used the model of explainability components when we coded concrete examples of addressees, aspects, contexts, and explainers. The first two authors met six times to compare their coding and during the meetings, any differences in the codes were discussed and the missing codes were added. During the axial coding, we first clustered the concrete examples of addressees, aspects, contexts, and explainers and then identified the relationships between them. Next, the third author reviewed the results of the paper to make sure the data are interpreted right during the analysis.

Finally, the first two authors defined a set of explainability requirements of recruitment game using the template. Thereafter, the third author reviewed the defined explainability requirements. The explainability requirements defined for the recruitment game system by using the template are presented in Section 5.2.

4. Results of analysis of ethical guidelines (Phase 1)

The Phase 1 of our study presents the results from the analysis of ethical AI guidelines of the sixteen organizations. First, we summarize what quality requirements the organizations have raised in their ethical guidelines of AI systems. In Section 4.2, we report the results of the analysis of transparency and explainability guidelines and describe the components for defining explainability requirements. We also propose a template for representing individual explainability requirements. In Section 4.3, we summarize the quality requirements that relate to

transparency and explainability.

4.1. Overview of ethical guidelines of AI systems

This section gives an overview of what quality requirements the organizations refer to in their ethical guidelines. In Table 5 and 6, we summarize the quality requirements of AI systems that have been emphasized in the ethical guidelines of the sixteen organizations.

In this study, 14 out of the 16 organizations have defined **transparency** ethical guidelines, and all the professional services and software companies have defined the transparency guidelines for developing AI systems. The key focus on the transparency guidelines encompassed the utilization of AI i.e., how the AI is used in the organizations (O2, O5, O6, O13). Moreover, openness or communicating openly (O4, O5, O11, O12, O14, O15) on how and where the AI is used in the system are indicated in the guidelines. Interestingly, **explainability** was always defined as a part of transparency guidelines in 13 out of the 14 organizations. The only exception was the organization O7 that did not cover explainability in their ethical guidelines of AI systems. A more detailed analysis of transparency and explainability guidelines is described in the following section.

Privacy ethical guidelines in organizations focused to protect and to avoid unethical usage of personal and sensitive data (O1, O2, O6). Moreover, compliance with privacy guidelines and the GDPR were emphasized in the privacy guidelines of the two organizations (O3, O4). Furthermore, Organization O6 highlighted that it is important to communicate how, why, when, and where user data is anonymized. Confidentiality of personal data and privacy of their customers are prioritized (O11, O16) and adherence to data protection practices (O11, O12, O13, O14, O15) are covered in the privacy guidelines of B2C and public sector organizations.

Few of the professional services and software organizations (O1, O5, O6, O9) and B2C (O11, O13) organizations defined their **security** and privacy guidelines together. Ensuring the **safety** of the AI system and user data by preventing misuse and reducing risks, and compliance to safety principles were also highlighted in privacy and security guidelines (O4, O6, O8, O11, O16). The security guidelines portrayed the need to develop secure AI systems (O5, O6, O8) and to follow data security practices (O1, O10, O11, O13, O16).

Professional services and software organizations and B2C organizations developed ethical guidelines for **fairness** that aim to avoid bias and discrimination. According to the B2C organizations, AI and machine learning utilization should eliminate discrimination and prejudices when making decisions and should function equally and fairly to everyone (O10-O13). In professional services and software organizations, fairness is advocated by fostering equality, diversity, and inclusiveness. The algorithms and underlying data should be unbiased and are as representative and inclusive as possible (O1, O4, O6, O8). From the organizations' viewpoint, developing unbiased AI contributes to responsible AI development.

Accountability ethical guidelines focused on assigning humans who will be responsible for monitoring AI operations, such as AI learning, AI decision-making (O5, O11, O16). The objective of the organizations was

Table 5
Quality requirements in ethical guidelines of Category A.

Quality Requirements	Professional services and software								
	O1	O2	O3	O4	O5	O6	O7	O8	O9
Transparency	x	x	x	x	x	x	x	x	x
Explainability	x	x	x	x	x	x	x	x	x
Privacy	x	x	x	x	x	x	x		x
Security	x				x	x		x	x
Safety				x		x		x	
Fairness	x		x	x	x	x		x	x
Accountability			x	x	x		x		x
Reliability					x	x			

Table 6
Quality requirements in ethical guidelines of Category B and C.

Quality Requirements	B2C						Public Sector	
	O10	O11	O12	O13	O14	O15	O16	
Transparency		x	x	x	x	x		
Explainability		x	x	x	x	x		
Privacy		x		x	x	x	x	
Security	x	x	x	x			x	
Safety		x					x	
Fairness	x	x	x	x				
Accountability		x			x		x	
Reliability							x	

to assign owners or parties who will be responsible for their AI operations and algorithms. The respective owners or parties will be contacted when concerns arise in the AI system, such as ethical questions and issues, harms, and risks (O4, O3, O11, O14, O16). Further, a couple of professional services organizations recommended establishing audit certifications, human oversight forums, or ethics communities to ensure accountability mechanisms throughout the system lifecycle and to support project teams (O7, O9). In organizations, the accountability guidelines are reckoned to closely relate to responsibility i.e., humans being responsible for the decisions and operations of the AI system.

Professional services and public sector organizations provide contrasting perspectives about **reliability** in AI development. For professional services and software organizations, reliability is coupled with safety and quality standards that help in assessing the risks, harms, and purpose of AI before its deployment (O5, O6). Whereas reliability in the public sector organization centered on the use of reliable data in AI. When the data or algorithms are unreliable or faulty, the organization corrects them to match the purpose of the AI system (O16).

4.2. From ethical guidelines to explainability requirements

In this section, we first report why the organizations emphasized transparency and explainability in their ethical guidelines. Then, we present the examples of explainability components. We identified these examples from the transparency guidelines of the 14 organizations. These components are based on the explainability definition proposed by Chazette et al. [6]. Finally, we suggest a template for representing individual explainability requirements.

Reasons to be transparent: The ethical guidelines of 10 organizations contained reasons why to incorporate transparency in AI systems. Five organizations (O1, O4, O5, O6, O11) portrayed building and maintaining users' *trust* as a prominent reason. Moreover, two organizations (O12, O13) highlighted that transparency supports *security* in AI systems. Organization O2 emphasized that being transparent helps in differentiating when AI makes the actual decision and when AI makes the recommendations that support people in making decisions. Furthermore, Organization O5 mentioned that transparency paves the way to mitigate *unfairness* and to gain more users' trust. The other reasons to develop transparent AI systems were to assess the *impact* of AI systems on society and to make AI systems available for assessment and scrutiny (O7, O14).

Fig. 2 illustrates the components of explainability that can be used when defining explainability requirements of AI systems. The purpose of these components is to give a structured overview of what explainability can mean. The four components can also be summarized with the following questions:

- Addressees - To whom to explain?
- Aspects - What to explain?
- Contexts - In what kind of situation to explain?
- Explainers - Who explains?

Fig. 2 also contains concrete examples of what these explainability

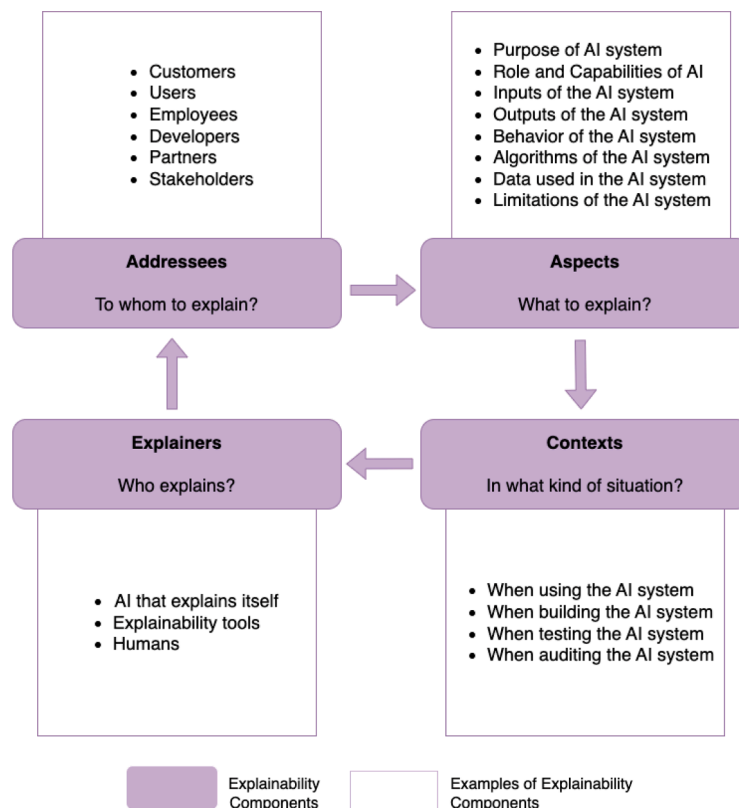


Fig. 2. A model of explainability components.

components can be in practice. These examples have been identified from the ethical guidelines of the organizations.

Addressees: The transparency guidelines covered a wide range of addressees to whom the AI or the different aspects of AI should be explained. Seven organizations (O1, O2, O6, O7, O13, O14, O15) highlighted that their AI should be explained and clearly communicated to their *customers*. Likewise, the explanations of AI systems were targeted to their *users* in O3, O5, O6, O11. According to the transparency guidelines of the organization O14, *partners and stakeholders* are also addressees of their AI systems. Besides, Organization O1 mentioned *employees* as their addressees, and Organization O5 narrowed the addressees down to *developers* of the AI systems.

Aspects: The key aspect that needs to be explainable is the *purpose* of AI systems (O6, O11). The intended purpose of the system should be communicated to the people who could be directly or indirectly impacted by the system (O11). Particularly, the addressee(s) should know how and why the organization is utilizing AI (O5, O13). Further, the *role and capabilities of AI* (O2, O3, O6, O11) need to be explained, so that addressees can see when AI makes the actual decision and when it only supports people in making decisions with recommendations.

Further, four organizations (O4, O6, O11, O15) mentioned to explain the *inputs and outputs* of the systems, such as inputs and outputs of the algorithms, decisions of AI systems. The organization O5 indicated to explain the *behavior of the AI system* which encompasses the working principles of the system (O4). In addition, *algorithms* and the inner workings of AI models are explained to the target addressees (O3, O15).

Five organizations (O2, O3, O12, O13, O15) highlighted that it is vital to explain the *data* used in AI systems. Specifically, the data used for teaching, developing, and testing the AI models, and the information about where and how the data is utilized should be explainable. Nevertheless, the accuracy of the data on which the AI is based should be included when explaining the data. A couple of organizations (O5, O6) indicated that the *limitations* of the AI systems as an aspect that needs to be explained.

Contexts: Apart from what to explain (aspects) and to whom to explain (addressees), the guidelines also mentioned in what kind of situations to explain i.e., the contexts of explanations. First, the situation when explanations are needed is when addressees are *using* the AI system (O2, O13, O14, O15). Next, developers would need explanations in the context of *building* the AI system (O4) and *testing* the AI system (O15). According to the organization O4, the situation where the explanations could play a supporting role is when *auditing* the AI system.

Explainers: The guidelines of two organizations (O8, O9) referred to the explainer of the AI systems. Regarding the explainer (i.e., who explains), Organization O8 suggested developing AI that can explain itself. Moreover, developing explainability tools for providing explanations of AI systems was proposed by Organization O9. But they did not mention any concrete definition or examples of explainability tools.

The components of the explainability requirement can also be presented as a simple sentence (Fig. 3). The purpose of this template is to assist practitioners to represent individual explainability requirements in a structured and consistent way. This simple template is based on the template that is used for defining functional requirements as user stories in agile software development. The template suggested by Cohn [9] is as follows: As a (type of user), I want (capability) so that (business value).

Here we give two high-level examples of explainability requirements based on Fig. 3.

- “As a user, I want to get understandable explanation(s) on the behavior of the AI system from the system, when I’m using it”
- “As a developer, I want to get explanation(s) on the algorithms of the AI system from an explainability tool, when I’m testing it”

These high-level examples of explainability requirements aim to show that different addressees may need different types and levels of explanations. For example, when debugging the system, developers are likely to need more detailed explanations of AI behavior than users. Users do not necessarily want to understand the exact underlying algorithm and inner workings of the AI model.

In their conceptual analysis of explainability, Köhl et al. also suggest that different addressees need different, context-sensitive explanations to be able to understand the relevant aspects of a particular system [22]. They also remark that an explanation for an engineer may not explain anything to a user. Furthermore, they mention that the explainer could be even a *human expert*.

4.3. Quality requirements related to transparency and explainability

The analysis of the ethical guidelines exhibited that transparency and explainability are associated with several other quality requirements. Fig. 4 presents the nine quality requirements that are related to transparency and explainability.

According to the organizations, *understandability* contributes to the development of transparency and explainability of AI systems. The transparency guidelines covered three details when addressing the importance of understandability, they are 1) to assure that people understand the methods of using AI and the behavior of the AI system (O5, O12), 2) to communicate in a clear and understandable way on where, why, and how AI has been utilized (O15), and 3) to ensure people understand the difference between actual AI decisions and when AI only supports in making the decisions with recommendations (O2). Thus, understandability supports explainability and transparency by ensuring the utilization of AI is conveyed to people clearly and in necessary detail. *Traceability* in transparency guidelines accentuates the importance of tracing the decisions of the AI systems (O2, O12). Organization O12 also mentioned that it is important to trace the data used in the AI decision-making process to satisfy transparency.

The transparency and explainability of AI systems can also assist in building *trustworthiness* (O1, O4, O5, O11). Prioritizing transparency when designing and building AI systems and explaining the system to those who are directly or indirectly affected is crucial in building and maintaining trust. Furthermore, two organizations (O7, O13) highlighted *privacy* in their transparency guidelines. Ensuring transparency can also raise potential tensions with privacy (O7). Moreover, *auditability* in the transparency guideline suggested that it is vital to build AI systems that are ready for auditing (O4). Organization O5 indicated that transparency also assists in ensuring *fairness* in AI systems. In addition to the relationships shown in Fig. 4, we identified *security*, *integrity*, *interpretability*, *intelligibility*, and *accuracy* in the transparency guidelines, but their relationship with transparency and explainability is not clearly stated in the guidelines.

5. Results of the empirical study (Phase 2)

This section presents the results from the study conducted with the

As a <type of addressee>, I want to get explanation(s) on
an <aspect> of a <system> from an <explainer> in a <context>.

Fig. 3. A template for representing individual explainability requirements.

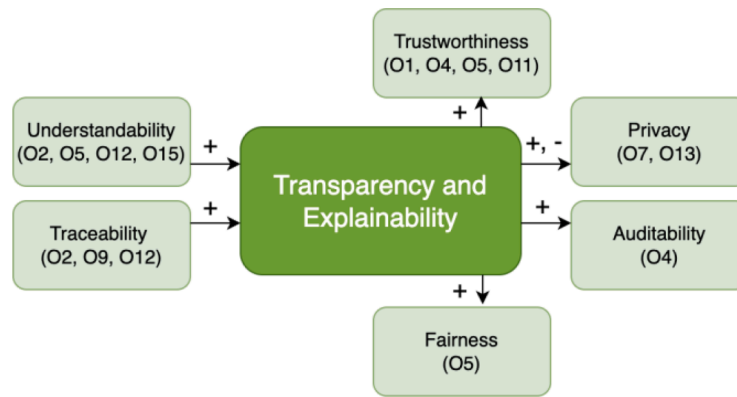


Fig. 4. Quality requirements related to transparency and explainability + Helps; – Conflicts.

representatives of Solita. First, we summarize the organization's perspective on the explainability of AI systems, their work relating to explainability, and the current state of the recruitment game. In Section 5.2, we report the results of the analysis of explainability components i. e., addressees, aspects, contexts, and explainers of the recruitment game. Using these results, we also defined a set of explainability requirements for the game. Section 5.3 describes the lessons learned in using the model of explainability components for defining the explainability requirements of the system.

5.1. Context of the study and current state of the recruitment game

This section provides the practitioners' viewpoint on transparency and explainability of AI systems and describes the current state of the system.

5.1.1. Background of explainability in the case organization

According to the interviewees, transparency and explainability are closely related concepts. The interviewees mentioned that developing understandable AI is a core part of achieving transparency and explainability of AI systems. In addition, two of them highlighted the black-box aspects of AI systems and emphasized the need to open the system to check how the results are formed and to identify the ways to explain them.

Moreover, one of the interviewees mentioned that the definition of the ethical guidelines of AI was one of their first steps in developing responsible AI, and the organization is continuously evolving in that domain. For instance, Solita is currently participating in the AIGA (Artificial Intelligence Governance and Auditing) research project, which aims to put responsible AI into practice [28]. The project was a collaboration of a team comprising academic and industry partners, and Solita was one of the industry partners of the AIGA project.

Our research collaboration with Solita was not part of the AIGA project, but it was closely related to the research activities of the AIGA project. We received an invitation to present our research in an AIGA seminar. We presented the results of Phase 1, after which there was a discussion session. Representatives from two organizations contacted us to discuss about the potential collaboration as the results we presented were interesting to them. We had an informal meeting with a representative from Solita who shared their work involving the use of "model cards" for explanations of AI systems [27]. The informal discussion served as a starting point for our collaboration with Solita.

5.1.2. Current state of the recruitment game

The recruitment game was developed in two parts. The game part was developed by seven students as a student project done for Solita and the recommendation part was added to the game by a master's thesis student. The recommendation part of the recruitment game systems was developed using a low-code development tool. Three recruiters from

Solita participated in the development of the recommendation part of the system. The system was still in its prototype phase and has not been used in the recruitment process yet, but the company representatives had identified the opportunity to develop it further and possibly use it as part of the recruitment process.

5.2. Explainability requirements of recruitment game

This section describes the results of using the model of explainability components and the template for representing explainability requirements. First, we summarize the addressees who needs explanations on the recruitment game system. Then, we report the aspects of the recruitment game that need to be explained to the addressees and contexts in which the addressees need explanations. In addition, we summarize the explainers who will be giving the explanations to the addressees. Finally, we give examples of explainability requirements of the recruitment game using the template.

Addressees: Fig. 5 presents an overview of the addressees that need explanations about the recruitment game. The two predominantly mentioned addressees of the recruitment game system are applicants who are looking for suitable positions at Solita and recruiters who are in-charge of selecting and skilled candidates. The workshop participants also mentioned three applicant groups that are summer trainees, junior software developers, and senior software developers.

During the workshop, the participants highlighted that both recruiters and interviewers work together during the recruitment process. Therefore, they both need explanations on how the system works and how it makes recommendations. Further, a participant categorized people leads and developers as people involved in interviewing the applicants. In addition, project managers require explanations about the recruitment game system, for example, the reason for selecting a particular applicant. One of the participants highlighted developers of the recruitment game as addressees who would need the explanations when they are developing the system.

Aspects: Table 7 summarizes the aspects of the recruitment game that have to be explained to applicants. The first aspect that needs to be explained is the purpose of the recruitment game. Further, the workshop participants pointed out that applicants need to understand the role of the system and the weight of the results in the recruitment process of Solita. The next aspect that requires explanation is the behavior of the system. Particularly, the participants highlighted that the overall working of the system and the procedure to use the system need to be explained to the applicants.

In the recruitment game system, the results made by the system is identified as an aspect that requires explanations. The participants discussed that the applicants want to know the ways to interpret the scores from the game and the details regarding people who would view those scores. Moreover, the reasons for not receiving a full score and what happens after playing the game need to be explained to the applicants.

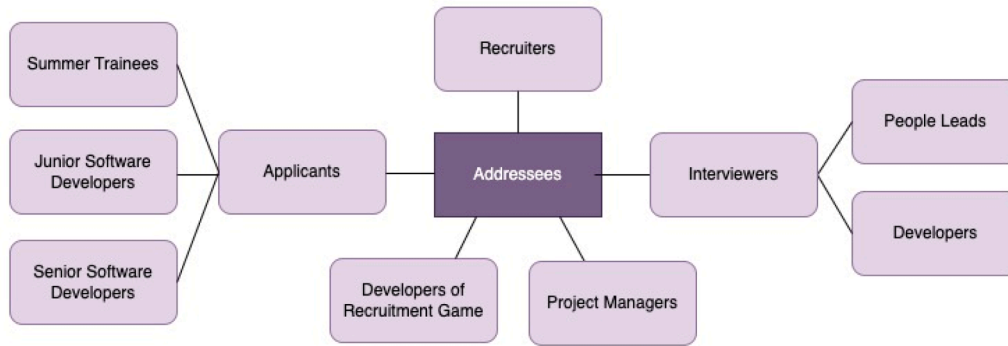


Fig. 5. Addressees that need explanations about the recruitment game.

Table 7
Aspects to be explained to **applicants**.

Aspect	Aspect as a question
Purpose of the system	- What is the purpose of the system?
Role of the system	- What is the role of the system in the recruitment process? - How will the system will be used exactly? - What kind of weight will the results will have in the recruitment process?
Behavior of the system	- How does the system work? - How to use the system? (How to play the game?)
Results of the system	- Who will see the scores? - How to interpret the scores? - Why didn't I get full score even though my answer was right? - What happens after the game? - Why was I suggested to apply for the job that I don't think is interesting or good fit?
Data used in the system	- What data is used in the recruitment game?

Apart from that, a participant summarized that the applicants should receive explanations when the system suggests job positions that are out of their interest area. Finally, the applicants require explanations about the data used in the recruitment game.

Table 8 summarizes the aspects of the recruitment game for which the recruiters and interviewers require explanations. The participants mentioned that recruiters and interviewers need explanations regarding the role and behavior of the recruitment game system. Particularly, explanations about the role of the system in the recruitment process and workings of the system. In addition, the addressees should receive explanations to understand the results. One of the participants indicated that in some situations, the recruiters and interviewers would need explanations specific to an applicant regarding why the job positions were suggested to them by the system.

Contexts: **Table 9** presents the different contexts of explanations for the addressees of the recruitment game system. First, the participants mentioned that they need explanations when planning the recruitment process. After the recruitment planning phase, the addressees would require explanations before using the game and after using the game. Next, recruiters and interviewers would be benefited from explanations

Table 8
Aspects to be explained to **recruiters and interviewers**.

Aspects	Aspects in detail
Role of the system	- What is the role of the system in the recruitment process?
Behavior of the system	- How does the system work?
Results of the system	- How to interpret the scores? - Why was applicant suggested to apply for jobs that we don't find suitable?

Table 9
Contexts of explanations.

Addressees	Contexts
Applicants	- Before using the game - After using the game - Job fairs
Recruiters and Interviewers	- When planning the recruitment process - Recruitment marketing - Job fairs - When using the game results in the evaluation process
Developers	- When applicant challenges the hiring decision - When recruiter challenges the hiring decision - When calibrating the AI

when they are interviewing the applicants and when they are using the game results in the evaluation process of hiring an applicant.

Furthermore, the workshop participants highlighted that explanations are helpful when the users of the system challenge the decision. Subsequently, developers can use the explanations about the inaccurate decisions when they are calibrating the AI or teaching the AI during the testing phase. The explanations about the system could also be useful when the system is used in job fairs and recruitment marketing.

Explainers: The participants shared that the recruitment game **system** should itself be an explainer. For example, by providing an introductory walkthrough before using the game and explaining the results after the game. The other explainers mentioned by the workshop participants are **humans**, such as people leads, recruiters, and game developers. For instance, the participants highlighted that people leads act as an explainer during the interviews when applicants ask questions regarding the game results. Likewise, game developers can explain how the system is used and limitations of the system to the recruiters and people leads.

5.2.1. *Examples of explainability requirements of the recruitment game*

After we defined the components of explainability (**Fig. 5** and **Tables 7-9**), we used these results and the template, and we defined a set of explainability requirements from the perspective of applicants and recruiters, who are the main user groups of the recruitment game. We defined two alternatives for representing the same explainability requirements. The following examples of explainability requirements were defined by the researchers of this study based on the analysis of the workshop data (**Fig. 5** and **Table 7-9**):

- **Alternative A:** As an applicant, I want to get explanation(s) from the system on the *behavior of the recruitment game* before using the game.
- **Alternative B:** As an applicant, I want to get explanation(s) from the system on *how the recruitment game works* before using the game.

- **Alternative A:** As an applicant, I want to get explanation(s) from the recruiters on the *meaning and consequences of the results of the game* after using the game.
- **Alternative B:** As an applicant, I want to get explanation(s) from the recruiters on *how to interpret the scores of the recruitment game* after using the game.
- **Alternative A:** As a recruiter, I want to get explanation(s) from the system about the *results of the recruitment game* when an applicant challenges the hiring decision.
- **Alternative B:** As a recruiter, I want to get explanation(s) from the system on *how to interpret the scores of the recruitment game* when an applicant challenges the hiring decision.
- **Alternative A:** As a recruiter, I want to get explanation(s) from the game developers on *behavior of the system* when planning the recruitment process.
- **Alternative B:** As a recruiter, I want to get explanation(s) from the game developers on *how the recruitment game works* when planning the recruitment process.

Alternative A is based strictly on the template presented in Fig. 3. In Alternative B, the aspects of the explainability are represented as questions because the participants of the workshop used these questions (Tables 7 and 8). It is possible that Alternative B is easier to understand. Therefore, these alternatives need to be tested with practitioners and compare which one is better in relation to understandability, usability, and usefulness.

5.3. Lessons learned

In this section, we share our reflections on using the model of explainability components and the template for defining explainability requirements for the recruitment game system. The following list gives an overview of the lesson learned from our study:

- The model of explainability components enables the systematic identification and analysis of the explainability needs of stakeholders.
- The template helps to represent individual explainability requirements in a structured and consistent way.
- Organizing a workshop with a multidisciplinary team supports in capturing different viewpoints on explainability of an AI system.
- When defining explainability requirements, it is important to have a good and shared understanding of the process that the AI system is supposed to support.
- It is important to define the purpose of the AI system clearly from the perspective of users and other stakeholders before defining the explainability requirements.
- It is important to consider potential risks and negative consequences of the AI system collectively with stakeholders and from various perspectives.

Lesson 1: The model of explainability components enables the systematic identification and analysis of the explainability needs of stakeholders.

At the beginning of the workshop, we asked participants to share their views on explainability and provided an overview of the model of explainability components. Then, the participants focused on all the four components of the explainability model separately. Their first task was to identify the addressees who need explanations about the recruitment game. During the tasks, the participants made notes on the Miro board and the notes were discussed before moving to the next tasks. Next, similar to the first task, the participants identified the aspects of the recruitment game that need to be explained. The process continued for the context and explainer components of the recruitment game, and the participants discussed their notes at the end of each task.

The tasks for eliciting the explainability of the recruitment game

helped in identifying the different stakeholders' needs of the system. Overall, the model supported in organizing the discussion to discover various perspectives about the explainability of the AI system. Moreover, during the analysis, the model guided the interpretation of workshop data. For example, the participants expressed the aspects to be explained as questions. In the analysis of the workshop data, we used the model to categorize the questions according to aspects as shown in Tables 7 and 8.

During the workshop, we did not define the aspects, contexts, and explainers for each addressee separately. This challenged the analysis as the views of explainability components expressed by our participants related to more than one addressee of the recruitment game. In addition, during the analysis it was challenging to relate the aspects or needs to the respective addressees or stakeholder groups. These challenges can be handled if the addressees of the system are prioritized and the aspects, contexts, and explainers are defined for each addressee separately.

Lesson 2: The template helps to represent individual explainability requirements in a structured and consistent way.

When we represented the explainability requirements of the recruitment game from the perspective of applicants, we used Table 7 as a starting point. This table summarizes the aspects that need to be explained to the applicants. After that we used Table 9 to specify the context of the explanations. Finally, we selected the explainers from the two main categorizations i.e., the system and humans.

One potential benefit is that these individual explainability requirements can be used in the product backlog of the Scrum methodology or on the Kanban board. In this way, explainability requirements guide the development of the AI system. In addition, these individual explainability requirements can be used in validating the explainability of the AI system during the testing phase.

The explainability template was not as easy to use as the template for user stories. The sentence structure of the template is quite complex because the template is a single sentence that contains all the four explainability components (addressee, aspect, explainer, and context). Therefore, the explainability components of the template can be reorganized to make the explainability requirements easier to understand. Organizing a follow-up workshop or a review meeting would have helped to validate the explainability requirements and to get feedback on the understandability of these explainability requirements.

Lesson 3: Organizing a workshop with a multidisciplinary team supports in capturing different viewpoints on explainability of an AI system.

The participants of the workshop were developers, and recruiters who are potential users and product owners of the recruitment game. The multidisciplinary team in the workshop enabled system-level and organizational-level perspectives of the system to be covered. For example, the developers highlighted the complexity of the programming task in the game and recruiters covered the role of the system during the recruitment process. The workshop discussion supported the exchange of ideas on the explainability of the recruitment game.

One of the limitations of our data collection process was that we did not invite applicants who are the users of the recruitment game. Instead, the recruiters brought up the applicants' viewpoints in the workshop. Conducting a workshop with potential applicants would help to validate the discovered explainability needs and to identify any missing needs.

Lesson 4: When defining explainability requirements, it is important to have a good and shared understanding of the process that the AI system is supposed to support.

The participants often referred to their recruitment process when the needs for explanations were discussed during the workshop. For example, they highlighted that the recruitment process of Solita is holistic. One of the participants described this in the following way: "... generally about recruitment, we don't usually recruit a person for any specific role. We just find software developers and their skills can be combinations of different skills. So, it is really challenging to think of all the qualities we are looking in the application." Another participant also mentioned that it is

not only technical aspects of candidates that are considered in the recruitment process. The same person also pointed out that the focus is to treat people as people and not as resources. The workshop discussion also revealed that the recruitment process can vary depending on whether the company is hiring summer trainees, junior software developers or senior software developers.

We as the facilitators of the workshop created a general level and limited view of the process based on the recruitment game before the workshop, but we did not interview representatives of recruiters. It would have been easier to ask follow-up questions in the workshop and analyze the data from the workshop if we had good knowledge about the recruitment processes. In addition, the discussion in the workshop could have been deepened if we had summarized the main activities of the recruitment process and discussed them together with the multidisciplinary team before the definition of explainability requirements.

Lesson 5: It is important to define the purpose of the AI system clearly from the perspective of users and other stakeholders before defining the explainability requirements.

As the development of the recruitment game started as a student project, the original idea of the recruitment game was that applicants can play the game and get scores from the four programming languages by doing different kind of tasks. After that, applicants can also get a pdf document containing their scores and they can attach with the application. When the recommendation part of the game was developed, the purpose was also to support recruiters in their process.

The purpose of the recruitment game was not clearly expressed in the workshop. When we introduced the game to the participants, we focused on describing the main functions of the game. The participants started to imagine what the purpose of the recruitment game could be when the positive aspects of the recruitment game were discussed in the workshop. The participants saw the following potential benefits:

- efficiency of the recruitment especially when number of applications is high
- a fun experience for applicants
- the possibility to attract more applicants
- a creative image of the company

The participants were able to see different ways how the recruitment game could be used in the actual recruitment process or before it, for example, in job fairs and recruitment marketing. Some participants saw that the game could assist in the recruitment process when there are a large number of applicants. Several participants also pointed that the usage of the game must be a fun experience for applicants. One of them described his view by saying that "... [the game] improves candidates experience in the best-case scenario". The participants also saw that the recruitment game can be a creative way to attract new applicants and it can also have a positive effect on the brand of the company. One of the recruiters summarized this in the following way: "...If we attract people with a game like this, it is much more creative [and] it shows much more of your brand [Solita] than a simple form you [the applicant] need to fill".

Since the purpose and role of the AI system in the recruitment process were not clearly defined before the participants started discussing the explainability components, it sometimes made the discussion of the explainability components fuzzy and challenging to interpret. It is also important to define the purpose and role of the AI system clearly, because they are aspects to be explained to addressees.

Lesson 6: It is important to consider potential risks and negative consequences of the AI system collectively with stakeholders and from various perspectives.

Many potential risks were raised by the participants during the workshop. When the participants were asked to discuss negative aspects of the recruitment game, they identified the following risks:

- poor applicant experience
- a narrow evaluation of applicants

- a bad image for the company
- the possibility to cheat

The participants perceived that the recruitment game can be stressful, and people may also be suspicious of the use of AI-based tools in recruitment, which can lead a bad applicant experience. One of the participants also pointed out that the recruitment game doesn't show the full potential of the applicant and coding tests with time limits do not imitate real situations where software developers have time to think before coding. The recruiters perceived that the recruitment game focuses on the results. They are also very interested in the thought process how the results are achieved. Furthermore, one of the participants pointed out that the game needs to work properly, otherwise it can create bad image for the company.

When we asked the participants to share their views on transparency and explainability, two of them raised the black-box problem of AI systems and possible biases that are really problematic especially regarding recruitment. One of the participants also described based on their earlier experience that the usage of AI in recruitment can be a sensitive and complex topic.

In summary, based on the lessons learned from our study, we suggest the following good practices to be used in the definition of explainability requirements of AI systems:

- Gain a good and shared understanding of the user and organizational processes that the AI system is supposed to support.
- Define clearly the purpose of the AI system and its role from the perspective of users and other stakeholders of the system.
- Analyze critically the potential risks and negative consequences of the AI system.
- Organize workshop(s) with a multidisciplinary team to capture different viewpoints on the explainability of the system.
- Use the model of explainability components to identify and analyze the explainability needs of stakeholders.
- Use a template to represent the explainability requirements of the AI system in a structured way.

In order to make the lessons learned concrete for practitioners, we reformulated them into good practices. The purpose of these practices is to support practitioners in defining explainability requirements of an AI system. The first two practices create the basis and context for the definition of explainability requirements. The third one is important because it helps different stakeholders share openly their views on the use of AI including their fears and negative perceptions. Based on the empirical study, workshops seem to be a good practice that supports the collaboration of a multidisciplinary team. During the workshop(s), solutions such as the model of explainability components can be used for identifying explainability needs. Based on these explainability needs, a template can be used for representing explainability requirements of an AI system.

6. Discussion

6.1. Transparency and explainability guidelines in practice

Nearly all the organizations of this study highlighted the importance of transparency and explainability in their ethical guidelines of AI systems. There were only two organizations out of sixteen that did not emphasize transparency. The results of this study support the findings of our previous analysis that suggested transparency and explainability as critical requirements of AI systems [3]. Three other papers [6,7,8] have also report transparency and explainability as the important quality requirements for developing AI systems.

Thirteen organizations of this study defined explainability as a key part of transparency in their ethical guidelines. Similarly, the studies of Chazette et al. [7] and Chazette and Schneider [8] reported that

integrating explanations in systems enhances transparency. According to Chazette et al. [7], it can, however, be difficult to define and understand the quality aspect of transparency [7]. The analysis of the ethical guidelines also indicates that it can be difficult to make a clear distinction between transparency and explainability in practice.

The prime goal of the organizations to incorporate transparency and explainability in AI systems was to build and maintain trustworthiness. Two studies [6,15] also report that explainability supports in developing transparent and trustworthy AI systems. Furthermore, Zieni and Heckel [26] suggest that implementing transparency requirements can support in gaining users' trust. According to Cysneiros et al. [13], and Habibullah and Horkoff [18], trust as a quality requirement plays a vital role in the development of autonomous systems [13] and machine learning systems [18].

Based on the definition of explainability proposed by Chazette et al. [6] and the analysis of the ethical guidelines, we suggest four important components to be covered in explainability requirements. These components of explainability are 1) to whom to explain (addressee), 2) what to explain (aspect), 3) in what kind of situation to explain (context), and 4) who explains (explainer). The ethical guidelines of the organizations included concrete examples of what these four components can mean in practice. We believe that concrete examples can support practitioners in understanding how to define explainability requirements in AI projects.

The analysis of the ethical guidelines revealed that the organizations consider customers and users as key *addressees* that need explanations. Developers, partners, and stakeholders were also mentioned as addressees who need explanations of AI systems. According to Chazette et al. [6], understanding the addressees of the system is as a key factor that impacts the success of explainability.

The ethical guidelines of the organizations contained a rather large number of *aspects* that need to be explained to addressees. For example, the explanations should cover the role and behavior of the AI system. Furthermore, the ethical guidelines of the organizations pointed out that it is important to describe the purpose and limitations of the AI system. Köhl et al. [22] state that explaining aspects of AI system is beneficial for their addressees to understand the system. Subsequently, Chazette et al. [6] highlighted aspects that need explanations are processes of reasoning, behavior, inner logic, decision, and intentions of the AI systems.

The results show that the different *contexts of explanations* (i.e., in what kinds of situations to explain) are: when using, building, testing, and auditing the AI system. Köhl et al. [22] and Chazette et al. [6] highlighted that the context-sensitive explanations support target groups receive intended explanations. In our study, the AI system itself was mentioned as the *explainer*. Similarly, Chazette et al. [6] reported that explainers can be a system or parts of the system.

One interesting result from the analysis of the ethical guidelines was the relationship of transparency and explainability with understandability, trust, traceability, auditability, and fairness. For instance, the *understandability* quality aspect focused on explaining the behavior of the system transparently to the addressees. Chazette et al. [6] also reported understandability as a crucial quality requirement that positively impacts explainability and transparency and enhances the user experience.

Further, the guidelines exhibited the association with *fairness*, where ensuring transparency and explainability helps in mitigating unfairness. Various studies [6,18,19] have identified fairness as important quality requirement of machine learning [18,19] and explainable systems [6]. In addition, quality requirements such as, *accuracy*, *traceability*, *privacy* and *security* were emphasized in the ethical guidelines. In the literature [6,18,19], all these four quality requirements are considered to be essential when building AI systems.

6.2. Defining explainability requirements

The results of the empirical study show that the model of

explainability components supports the systematic identification of explainability needs. During the workshop, we followed a step-by-step process to discover explainability needs of different stakeholders. First, the participants of the workshop identified stakeholders who need explanations and then they defined the aspects of the recruitment game that need to be explained. Finally, the workshop participants considered in what kind of situations explanations are needed and who gives these explanations.

One interesting finding from the workshop was that humans were highlighted as one of the important explainers of the recruitment game. Similarly, Köhl et al. [22] indicate that human experts can provide explanations of AI systems. Our original model of explainability components did not include humans as explainers. Therefore, we must update the model.

The results of our study also indicate that the explainability template helps representing explainability requirements in a structured way. Sommerville and Sawyer [33] also recommended the use of standard templates for describing requirements as a good practice. According to them, the key benefit is that standard templates make requirements easier to write, read and present them consistently.

The results of this study also suggest that organizing workshops with multidisciplinary teams is a good practice to define explainability requirements. Two organizations in our previous study also recommended to use multidisciplinary teams when developing AI systems [3]. The overall goal of using multidisciplinary teams is to bring out different viewpoints and support the ethical development of AI systems. Moreover, defining the purpose of the system is another good practice for developing ethical AI systems that was highlighted by the organizations of our previous study [3]. The results of this empirical study also emphasize that it is important to define the purpose of an AI system clearly before defining explainability requirements. The clear purpose of the AI system guides and simplifies the work of multidisciplinary team when they define explainability requirements.

This study also revealed that it is essential to consider potential risks and negative consequences of the AI system together with stakeholders and from various perspectives. We did not observe a direct impact of this good practice on the definition of explainability requirements, but we discovered that possible negative consequences can first affect the main functional requirements of the recruitment game and thereby also the explainability requirements.

The participants of this empirical study also brought up the black-box problem of AI systems and possible biases, which they perceived to be problematic, especially in recruitment. According to Dattner et al. [30], advances in AI have produced new tools that can be used in hiring. These tools can have both positive and negative implications, and the authors point out, in particular, the legal and ethical implications of using AI in hiring [30].

7. Threats to validity and limitations

7.1. Phase 1

Generalizability. Our study focused on the ethical guidelines of AI published by the 16 organizations. However, the ethical guidelines do not necessarily reflect what is happening in these organizations. Nevertheless, we think the guidelines contain important knowledge that should be considered when developing transparent and explainable AI systems. Therefore, we believe that organizations can utilize the results of this study to gain an overview and to understand the components that can help defining explainability in AI systems development.

Majority of the organizations of this study were Finnish or Finland-based international companies, and only three out of the sixteen organizations were global. When we compared the ethical guidelines of the global organizations with the ethical guidelines of the other organizations, there were no significant differences between them.

Reliability. Researcher bias might have influenced the data analysis

process. To avoid misinterpretation and bias, the coding process was done by two researchers separately. The high-level categorization of the organizations was also reviewed by a third senior researcher who is also one of the authors of this paper.

The organizations selection strategy resulted in some limitations. We selected organizations that have published their ethical guidelines of AI publicly in Finland. Hence, may be the smaller number of public sector organizations in our study. However, the focus of our study was on transparency and explainability, so we did not make conclusions based on the categories of the organizations.

7.2. Phase 2

There are five main limitations to our empirical study. First, the scope of the study was limited. We conducted one workshop to discover the explainability needs of different stakeholders and we did not to validate the explainability requirements of the recruitment game with the recruiters and developers. Therefore, it is not yet possible to report how useful the template is from the point of view of practitioners. In order to evaluate the usability and usefulness of the candidate solutions, it is important to continue empirical evaluations with users, developers, and other key stakeholders in a step-by-step manner as recommended by Wohlin and Runeson [36]. Our goal is to gradually expand the empirical evaluations of the candidate solutions proposed in this paper, and we aim at evaluations in which practitioners use the solutions in a real context and we researchers act as observers.

Another limitation related to the scope of the evaluation is that we researchers used the template to define the examples of explainability requirements. There were two reasons for this. The first reason was to minimize the risk that the recruiters would feel that they are spending time on a task that is not typically the responsibility of potential users. Secondly, we wanted to critically analyze the use of the template ourselves before asking practitioners to use it. This allowed us to identify two possible alternatives for representing explainability requirements. We report these alternatives A and B in Section 5.2.

The third limitation of our empirical evaluation is that we did not to involve applicants in the workshop, although they are one of the main user groups of the recruitment game. We made this decision consciously. In this first pilot evaluation, we decided to focus on the recruiters' point of view because one of the goals of the recruitment game was to support their processes. In addition, Author 3, who was the representative of the case company, succeeded in inviting professionals with wide and long experience on recruitments. Therefore, they were also able to describe explainability needs of applicants. The recruiters also talked very openly about their views on the use of AI in hiring. We assume that this would have not happened if applicants had participated in the workshop. Although we did not include applicants in the workshop, we consider their participation very important when identifying their explainability needs.

The fourth limitation relates to researcher bias which is one of the main validity threats associated with qualitative studies. To reduce this bias, we used researcher triangulation. Two researchers (Author 1 and Author 2) did the data analysis separately, compared the analysis results and resolved discrepancies in the analysis results iteratively in multiple meetings. To improve the validity of the results, one of the practitioners (Author 3) from Solita reviewed the results of the pilot study and gave feedback about them.

The fifth limitation of the empirical evaluation is that the recruitment game did not represent a typical project of Solita because the system was developed by students. However, we see that recruitment is an important domain and the selection of the recruitment game was a good decision as the use of AI in hiring portrayed as a sensitive and complex topic [30].

8. Conclusions

The goal of this study was to investigate what ethical guidelines organizations have defined for the development of transparent and explainable AI systems and evaluate how explainability requirements can be defined in practice. Our study shows that explainability is tightly coupled to transparency and trustworthiness of AI systems. This leads to the conclusion that the systematic definition of explainability requirements is a crucial step in the development of transparent and trustworthy AI systems.

In this paper, we propose a model of explainability components that can facilitate the definition of explainability requirements of AI systems. The purpose of the model is to assist practitioners with identifying explainability needs by answering the following important questions: 1) to whom to explain, 2) what to explain, 3) in what kind of situation to explain, and 4) who explains. This paper also introduces a template for representing explainability requirements in a structured and consistent way. In addition, we also recommend a set of good practices for defining explainability requirements of AI systems.

The results of this study indicate that the clarity of the purpose of an AI system affects the definition of explainability requirements. Furthermore, we have made the conclusion that the analysis of potential negative consequences and the usage of multidisciplinary teams bring important perspectives to the definition of explainability requirements.

One important direction in our future research is to perform case studies to understand how transparency and explainability requirements are currently defined in AI projects. We also aim to evaluate the usability and usefulness of the model of explainability components and the template for defining explainability requirements with practitioners in real context and study their experiences. Our long-term plan is to investigate how explainability requirements can be used in the testing of AI systems.

CRedit authorship contribution statement

Nagadivya Balasubramaniam: Conceptualization, Methodology, Investigation, Validation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Marjo Kauppinen:** Conceptualization, Methodology, Investigation, Validation, Writing – original draft, Writing – review & editing, Project administration, Supervision. **Antti Rannisto:** Investigation, Validation, Writing – review & editing, Project administration. **Kari Hiekkänen:** Conceptualization, Investigation, Writing – review & editing, Supervision. **Sari Kujala:** Conceptualization, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

We want to thank all the participants from Solita who participated in the interviews and workshop in Phase 2 of this study.

References

- [1] B. Abdollahi, O. Nasraoui, Transparency in fair machine learning: the case of explainable recommender systems. *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent*, Springer, 2018, pp. 21–35.

- [2] K. Ahmad, M. Bano, M. Abdelrazek, C. Arora, J. Grundy, What's up with requirements engineering for artificial intelligent systems?, in: International Requirements Engineering Conference, 2021, pp. 1–12.
- [3] N. Balasubramaniam, M. Kauppinen, S. Kujala, K. Hiekkänen, Ethical guidelines for solving ethical issues and developing AI systems. Product-Focused Software Process Improvement, 2020, pp. 331–346.
- [4] K. Charmaz, Constructing Grounded Theory, 2nd edition, SAGE Publications Inc, 2014.
- [5] L. Chazette, Mitigating challenges in the elicitation and analysis of transparency requirements, in: International Requirements Engineering Conference, 2019, pp. 470–475.
- [6] L. Chazette, W. Brunotte, T. Speith, Exploring explainability: a definition, a model, and a knowledge catalogue, in: International Requirements Engineering Conference, 2021, pp. 197–208.
- [7] L. Chazette, O. Karras, K. Schneider, “Do end-users want explanations? Analyzing the role of explainability as an emerging aspect of non-functional requirements, in: International Requirements Engineering Conference, 2019, pp. 223–233.
- [8] L. Chazette, K. Schneider, Explainability as a non-functional requirement: challenges and recommendations, Requirements Eng. 25 (4) (2020) 493–514.
- [9] M. Cohn, Agile Estimating and Planning, Prentice Hall, 2006.
- [10] J. Corbin and A. Strauss, “Basics of Qualitative Research,” SAGE, 4th edition, 2015.
- [11] L.M. Cysneiros, Using *i** to elicit and model transparency in the presence of other non-functional requirements: a position paper. Istar, Citeseer, 2013, pp. 19–24.
- [12] L.M. Cysneiros, J.C.S. do Prado Leite, Non-functional requirements orienting the development of socially responsible software, in: Enterprise, Business-Process and Information Systems Modeling, 2020, pp. 335–342.
- [13] L.M. Cysneiros, M. Raffi, J.C. Sampaio do Prado Leite, Software transparency as a key requirement for self-driving cars, in: International Requirements Engineering Conference, 2018, pp. 382–387.
- [14] J.C.S. do Prado Leite, C. Cappelli, Software transparency, Bus. Inf. Syst. Eng. 2 (3) (2010) 127–139.
- [15] K. Drobotowicz, M. Kauppinen, S. Kujala, Trustworthy AI services in the public sector: what are citizens saying about it? Requirements Eng.: Foundation for Software Quality (2021) 99–115.
- [16] European commission: ethics guidelines for trustworthy AI. <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines>. Accessed 24 Oct 2021.
- [17] R. Guizzardi, G. Amaral, G. Guizzardi, J. Mylopoulos, John, Ethical requirements for AI systems, in: 33rd Canadian Conference on Artificial Intelligence, 2020.
- [18] K.M. Habibullah, J. Horkoff, Non-functional requirements for machine learning: understanding current use and challenges in industry, in: International Requirements Engineering Conference, 2021, pp. 13–23.
- [19] J. Horkoff, Non-functional requirements for machine learning: challenges and new directions, in: International Requirements Engineering Conference, 2019, pp. 386–391.
- [20] IEEE: ethically aligned design, first edition <https://ethicsinaction.ieee.org/>. Accessed 24 Oct 2021.
- [21] D. Kwan, L.M. Cysneiros, and J.C.S. do P. Leite, “Towards achieving trust through transparency and ethics (Pre-Print),” 2021, Accessed: Aug. 30, 2021. [Online]. Available: <http://arxiv.org/abs/2107.02959>.
- [22] M.A. Köhl, K. Baum, M. Langer, D. Oster, T. Speith, D. Bohlender, Explainability as a non-functional requirement, in: International Requirements Engineering Conference, 2019, pp. 363–368.
- [23] B. Lepri, N. Oliver, E. Letouzé, A. Pentland, P. Vinck, Fair, transparent, and accountable algorithmic decision-making processes, Philos. Technol. 31 (4) (2018) 611–627.
- [24] B. Paech, K. Schneider, How do users talk about software? Searching for common ground, in: Workshop on Ethics in Requirements Engineering Research and Practice, 2020, pp. 11–14.
- [25] SIIA (Software and information industry association): ethical principles for artificial intelligence and data analytics, pp. 1–25, 2017.
- [26] B. Zieni, R. Heckel, TEM: a transparency engineering methodology enabling users' trust judgement, in: International Requirements Engineering Conference, 2021, pp. 94–105.
- [27] V. Stirbu, M. Raatikainen, J. Rontynen, V. Sokolov, T. Lehtonen, and T. Mikkonen, “Towards multi-concern software development with everything-as-code,” *IEEE Software*, 2021.
- [28] Artificial intelligence governance and auditing (AIGA), [Online]. Available: <https://ai-governance.eu/>, Accessed 17 Jul 2022.
- [29] N. Balasubramaniam, M. Kauppinen, S. Kujala, K. Hiekkänen, Transparency and explainability of AI systems: ethical guidelines in practice, in: Requirements Engineering: Foundation for Software Quality, 2022, pp. 3–18.
- [30] B. Dattner, T. Chamorro-Premuzic, R. Buchband, L. Schettler, The legal and ethical implications of using AI in hiring, Harv Bus Rev (2019) [Online]. Available: <https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring>. Retrieved 22 Jul 2022.
- [31] B. Attard-Frost, A. De los Ríos, D.R. Walters, The ethics of AI business practices: a review of 47 AI ethics guidelines, AI Ethics (2022), <https://doi.org/10.1007/s43681-022-00156-6>.
- [32] M.Lenca Jobin, E. Vayena, The global landscape of AI ethics guidelines, Nat. Mach. Intell. 1 (2019) 389–399.
- [33] Sommerville, P. Sawyer, Requirements Engineering: A Good Practice Guide, Wiley, 2006.
- [34] T. Gorschek, P. Garre, S. Larsson, C. Wohlin, A model for technology transfer in practice, IEEE Softw 23 (6) (Nov. 2006) 88–95, <https://doi.org/10.1109/MS.2006.147>.
- [35] T. Gorschek, K. Wnuk, Third generation industrial co-production in software engineering, in: M. Felderer, G.H. Travassos (Eds.), Contemporary Empirical Methods in Software Engineering, Springer International Publishing, Cham, 2020, pp. 503–525, https://doi.org/10.1007/978-3-030-32489-6_18.
- [36] C. Wohlin, P. Runeson, Guiding the selection of research methodology in industry–academia collaboration in software engineering, Inf. Softw. Technol. 140 (Dec. 2021), 106678, <https://doi.org/10.1016/j.infsof.2021.106678>.