Exploring Ethical Requirements Elicitation for Applications in the Context of AI

José Antonio Siqueira de Cerqueira[1], **Edna Dias Canedo (Advisor)**[1]

[1]Department of Computer Science, University of Brasília (UnB)
P.O. Box 4466, Brasília, DF, Zip Code 70910-900, Brazil

jose.siqueira.cerqueira@gmail.com, ednacanedo@unb.br

Abstract. *Ethical concerns arises from the proliferation of Artificial Intelligence (AI) based systems in use. AI ethics has been approached mainly in guidelines and principles, not providing enough practical guidance for developers. Hence, we aim to present RE4AI Ethical Guide and its evaluation. We used the Design Science Research methodology to understand the problem, present the guide and evaluate it through a focus group. The Guide is composed of 26 cards across 11 principles. We evaluated it with 5 AI professionals and our preliminary results reveal that it has the potential to facilitate the elicitation of ethical requirements. Thus, we contribute to bridge the gap between principles and practice by assisting developers to elicit ethical requirements and operationalise ethics in AI.*

1. Introduction

Artificial Intelligence is with no doubt one of today's hottest topics in the academic, social and industrial fields [Mayer et al. 2021]. Artificial Intelligence offers several opportunities for individuals and society on a broad scale to dramatically improve and enhance their capabilities in performing complex tasks, being able to reinvent society by introducing the technologies, which, because they have such extraordinary and disruptive potential, also introduce proportional risks [Floridi et al. 2018], beyond the business area, but also as global climate, legal system, medicine, global surveillance, transportation, wars, among others, from which arises the need to investigate AI ethics pointed out by several authors, including [Benjamins et al. 2019] [Krafft et al. 2020] [Vakkuri et al. 2020b] [Jobin et al. 2019] [Guizzardi et al. 2020] and European institutions [Commission 2019], and is currently a highly debated topic [Ryan and Stahl 2020].

Overall, AI ethics has been approached, in the literature, in its theoretical field, through presentation of ethical guidelines [Mittelstadt 2019]. Whereas it is essential to discuss guidelines and principles, actual practical directions for developers have been generally overlooked by researchers [Mittelstadt 2019]. Ethical principles in AI are not automatically translated to practice [Mittelstadt 2019]. These principles are often broad and abstract, and the higher the level of abstraction of ethical values, the more interpretation is required to perform the translation to practice [Smit et al. 2020]. Moreover, developers lack proper training, both in real projects and in academic studies.

Requirements elicitation phase provide a fruitful setting for discussions on ethical issues by enabling increased interaction between different actors involved in software development and its use [Kostova et al. 2020], also, when considering ethical issues in the early stages of software development, rather than as an afterthought, a reduction in additional work occurs [Vakkuri et al. 2020a].

Traditional techniques used in Requirements Engineering can also be used to develop AI-based systems in compliance with ethical principles and guidelines [Guizzardi et al. 2020]. Vakkuri et al. [Vakkuri et al. 2020a] presented a method to implement ethics in AI-based systems, called ECCOLA, based on Planning Poker. This method consists of a set of 21 cards, divided into 8 themes, with questions to be answered by the Product Owners and developers. The ethical principles laid out in the AI HLEG [Commission 2019] and IEEE EADv1 [IEEE 2019] guidelines served as a basis for the cards classification into themes and in designing the questions shown in the cards. This set of cards is modular, the relevant cards for the iteration (sprint) in question can be used, i.e., it is suitable for agile development. It is stated that the ECCOLA method only helps to increase the ethical awareness of the development team, providing no means of measuring the impact of the use of the tool, nor did they include examples or assessments of the use of the method in practice.

With the purpose of mitigating AI-based systems ethical issues, the aim of this work is to provide a Guide for Artificial Intelligence Ethical Requirements Elicitation (RE4AI Ethical Guide), its development, evaluation and final remarks. The Guide will help software development teams to elicit ethical requirements for AI, in the first phase of Software Development Life Cycle (SDLC) with a focus in agile software development. The Guide consists of a deck of 26 cards across 11 ethical principles. The users will answer the questions presented in the cards and the answer will then be user stories – ethical requirements – to be included in the sprint backlog without limiting the context where the AI-based system is used.

For its creation, the Design Science Research methodology [Vaishnavi and Kuechler 2015] was adopted in order to understand the problem, develop a prototype and evaluate it through a focus group with AI professionals. From the evaluation, the proposed Guide is both useful and practical and can help in the elicitation of ethical requirements in the context of agile development. Thus, our preliminary results reveal that the Guide contributes to bridging the gap between high-level and abstract principles and practice by assisting developers and Product Owners, especially in agile development projects, to elicit ethical requirements and implement ethics in AI using user stories.

Potential benefit from the use of the proposed Guide is to mitigate negative impacts of AI-based systems operating in various sectors of society by providing practical means to operationalise AI ethical principles from the initial stage of the SDLC.

The goal of this study is to help to operationalise AI ethics in practice by developers and multiple stakeholders, presenting two contributions:

1. To provide a Guide for Artificial Intelligence Ethical Requirements Elicitation (RE4AI Ethical Guide);
2. To evaluate the Guide through a focus group with AI practitioners.

2. Related works and Theme Overview

The proposed guide is intended to provide a way to develop a system based on ethical requirements before it is deployed, however, it can also be used to evaluate a system in use. Furthermore, AI-based systems in use in society need to be demonstrated ethical to the general public. This consideration is addressed in one of the challenges of Cyber-Physical

Emergency Response Systems, as reported in [Araujo et al. 2017], since AI-based systems can also be Cyber-Physical systems, such as driverless cars [Guizzardi et al. 2020]. More specifically one of the main challenges we face is related to how evaluate such systems since they are heterogeneous and involving different types of actors in frequently changing scenarios.

2.1. Guidelines and Principles

In the literature, most studies focus on the conceptual part of AI ethics, and one of them is the compilation, presentation and evaluation of ethical guidelines and their principles. Several authors have used different methodologies to explore sets of documents and extract the most recurrent principles and their definitions, usually concluding that they are too general, have high level of abstraction and degree of difficulty in applying them in real contexts, besides there is overlap between the principles [Hagendorff 2020, Jobin et al. 2019, Fjeld et al. 2020, Zeng et al. 2019, Smit et al. 2020, Floridi and Cowls 2019, de Ágreda 2020, Ryan and Stahl 2020, Rothenberger et al. 2019]. At least 84 of public and private initiatives have published reports describing ethical principles for the development and deployment of AI [Mittelstadt 2019]. We selected for the purpose of our study, the 11 ethical principles and issues listed in [Ryan and Stahl 2020], that is: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, sustainability, dignity and solidarity. For the best of our knowledge, this is the study that approached the most comprehensive set of guidelines, 94.

2.2. AI ethics in practice

In 2019 a greater focus has begun, among academia, on how to carry out this translation of ethical principles into practice, that is, the translation from "what" to "how" [Newman 2020]. Despite difficulties to operationalise abstract principles into practice, several tools and methods to implement ethics in AI exist, and surprisingly, the vast amount of available tools by itself is already a drawback [Schiff et al. 2020b].

In Newman's study [Newman 2020] 6 tools and 12 best practice frameworks were evidenced, and in Morley et al. [Morley et al. 2019] 106 tools were found, presenting a wide list of tools and methods and flag some significant challenges, such as: (a) the tools included are relatively immature, have little documentation and are not necessarily ready for use, resulting in little usability by developers and additional work to be put into practice; (b) difficulty in assessing their scope of use; (c) difficult to encourage their adoption by the practical mind of AI and ML developers; d) the vast amount of available tools makes it difficult to evaluate and choose, being difficult to compare one with another, despite there being articles or public repositories on GitHub, if the tool is not supported by a community, with active users – both developers and scholars – public availability of the source code and ample documentation, there will not be an adherence nor usefulness of it [Schiff et al. 2020b].

We searched for tools that assist implementing AI ethics in a previous study [Cerqueira et al. 2021b] and we found 21 tools, such tools will be filtered, and matched with respective principles, and then will be part of a card in our Guide, as seen in our other previous study [Cerqueira et al. 2022].

2.3. Requirements Engineering for AI

The process of Requirements Engineering for AI-based systems (**RE4AI**) is different from traditional systems [Kostova et al. 2020], and there is an additional complexity to the development of AI-based systems [Nguyen-Duc et al. 2020], because there is a dependency between the large amount of data and algorithms [Belani et al. 2019], being observed in some cases the use and extension of already well-established approaches, principles and tools in Software Engineering for the development of AI-based systems [Schleier-Smith 2015]. Some authors have explored the challenges of Requirements Engineering, as well as Software Engineering, for AI-based systems (e.g., [Belani et al. 2019] [Nguyen-Duc et al. 2020] [Schleier-Smith 2015] [Lwakatare et al. 2019]). Thus, in the context of AI-based systems there is a difficulty in tracing the output of a model back to system requirements, as they may not be explicitly documented, and possible issues arise only when the system is deployed [Raji et al. 2020]. Furthermore, "there is a lack of a formalization of a standard development or practice model, or process guidelines for when and in what context it is appropriate to implement certain recommendations" [Raji et al. 2020].

Schiff et al. [Schiff et al. 2020b] suggest that interdisciplinary teams in both higher education institutions and organizations, with humanities and social science students partnering with engineering and computer science students, and developers partnering with non-technical teams (e.g., social scientists, lawyers, ethicists), respectively, should aim to learn each other's languages and work together. Then, they listed 6 criteria for a framework for responsible AI development (broad, operationalizable, flexible, iterative, guided and participatory) in order to propose a framework, based on the IEEE 7010 standard. We take into account this list of criteria during the creation of our guide, as seen in [Cerqueira et al. 2022]. Although the authors' proposal is designated framework, they end up introducing an impact assessment list, that is, this tool would also have its greatest practical utility in the final phases of the Software Development Life Cycle. Moreover, the authors do not present a very detailed definition of the framework nor how to follow it, and the use case presented includes few technical details on how to actually implement the framework, presenting broad and generic recommendations.

We based the creation of our guide on the ECCOLA method [Vakkuri et al. 2020a] mentioned in the Introduction Section, because it provides means to elicit ethical requirements for AI – being used in the first phase of the Software Development Life Cycle, while not limiting stakeholders participation and by motivating users to answer questions and think about ethical issues that they are not used to consider in their everyday work.

3. Methodology

In order to target our goals, we used the Design Science Research methodology [Vaishnavi and Kuechler 2015] to understand the problem, develop a prototype and evaluate it through a focus group with AI practitioners.

For the first stage a Systematic Literature Review was conducted in order to: 1) identify the existing techniques, methodologies, methods, frameworks, processes and tools in the literature to support the operationalisation of ethical requirements in AI; 2) identify the works that investigate ethics in the elicitation of requirements for applications

in the context of AI and Machine Learning; 3) identify the existing ethical principles and guidelines in the literature and industry in the context of AI.

Of the 33 primary studies selected, few explicitly address the use or present new proposals for practical means to implement ethics in AI. The proposals found present a low level of maturity, few practical examples and lack of documentation. This demonstrates how practical AI ethics is still in its early stages, above all regarding practical guidelines, ethical requirements, and tools.

After the analysis of techniques, tools, methods, frameworks and processes found in the literature, we identified the ECCOLA method [Vakkuri et al. 2020a] as the most suitable for our context, consisting of a deck of cards, based on Planning Poker, for the elicitation of ethical requirements in AI, made available in a static way. We also found the need for the inclusion of traditional Software Engineering practices, such as requirements elicitation [Guizzardi et al. 2020], for the context of AI, besides the characteristics of a Guide to implement ethics in AI: 1) Broad; 2) Operationalisable; 3) Flexible; 4) Iterative; 5) Guided; 6) Participative [Schiff et al. 2020a].

4. Guide for Artificial Intelligence Ethical Requirements Elicitation – RE4AI Ethical Guide

The Guide for Artificial Intelligence Ethical Requirements Elicitation was implemented as a web-based system and is divided into: Introduction, presenting a brief introduction and how to use it; Guide, presenting the set of cards; Principles, presenting all the principles present in the guide; Tools, presenting which tools are present in the guide related to the principles; Trade-offs, presenting which trade-offs may occur when developing AI-based systems that take ethical issues into consideration; and About, briefly presenting the authors, information about the guide and references used. In Figure 1 we present the initial screen of the guide, where its subdivisions are present.

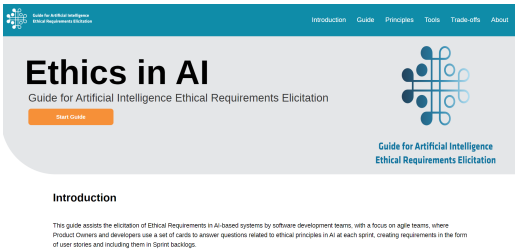

Figure 1. Home page of the guide. Own source

By clicking Start Guide, the user will be presented by default with all the cards, and the options to filter or compare cards, as illustrated in Figure 2.

In order to display only the cards related to a specific principle, the user must select the desired principle from the Filters menu (on the left). At the top of the card there is the card number, its ethical issues, and the ethical principle, besides that, the principles are related to different colours. The user can click on the tool provided in the Tool Suggestion field, where a new tab will open in the browser, displaying the source code repository on GitHub of the respective tool. If 2 or more cards are selected, the user can click Compare cards (on the right), where only those cards will be displayed. The user can then click

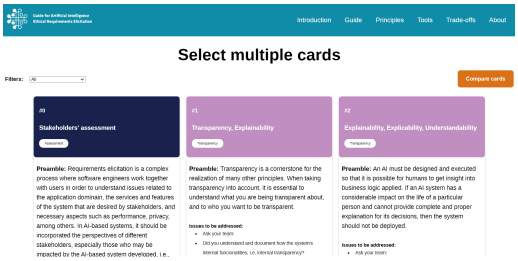

Figure 2. Card selection page. Own source

Start again to return to the previous screen where all cards are displayed. In Figure 3 we illustrate the scenario where 4 cards of different principles are selected and compared.

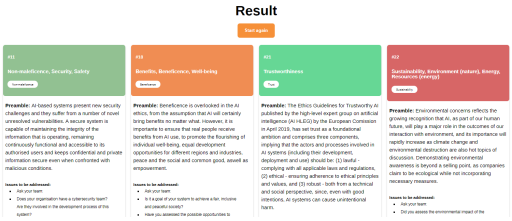

Figure 3. Different cards compared. Own source

It is available in the footnotes the address of the source code of the system, in addition to the license Creative Commons 4.0 International (CC BY 4.0), in order to allow the sharing and adaptation of the guide preserving the attribution of the credit to the authors [Méndez et al. 2020]. In addition, a free font is used throughout the system – UnB Office, allowing greater portability and maintainability, while users (i.e., developers, ethicists, public organizations, academics) of the system should not have to worry about patents or copyright licenses for fonts, or any other aspect of the system.

The source code of the guide is available at https://github.com/josesiqueira/RE4AIEthicalGuide and the system at https://josesiqueira.github.io/RE4AIEthicalGuide/. In the guide 24 cards are provided, distributed along the 11 principles adopted for the elaboration of the guide, plus 2 additional cards: Stakeholders' assessment, and Overall ethical evaluation, both under the topic of Assessment. Thus, in total there are 26 cards. In the sprint backlog meeting, the actors must choose the cards that will be used in that sprint, read aloud the content of the card, then the development team will elicit the ethical requirements in the form of user stories, also writing down the reasoning that led them to those user stories. Validation should be done by development teams together with customers and multiple stakeholders, who may request changes.

The prototype designed – the Guide to Eliciting Ethical Requirements for AI –, and the ECCOLA method developed by Vakkuri et al. [Vakkuri et al. 2020a], differ in many aspects. While the latter is presented only as a deck of cards in Portable Document Format, our guide is developed as a web-based system (using HTML, CSS and JS), allowing interactivity in card selection through filters and comparisons between multiple cards, as well as extensive supporting material (how to use, principles, tools, trade-offs) and the addition of tool suggestion in the content of the cards. We also assigned free licenses and made the source code available, in order to allow the study of the tool and

future adaptations to new contexts. Moreover, we contemplated in our guide all the 11 principles listed by Ryan and Stahl [Ryan and Stahl 2020] and presented 26 cards, while in the ECCOLA method are contemplated only 7 principles with a total of 21 cards. In Table 1 we point out the main differences between ECCOLA and the Guide for Artificial Intelligence Ethical Requirements Elicitation.

Table 1. Main differences between ECCOLA and the Guide for Artificial Intelligence Ethical Requirements Elicitation

ECCOLA	RE4AI Ethical Guide
Static	Interactive
7 principles	11 principles
21 cards	26 cards
Does not suggest any tools	Suggested tools available
Copyrighted	Open license with source code available
No support material available	Support material available

4.1. Focus group

Through a focus group, the moderator presented the RE4AI Ethical Guide and its contextualization to the participants – experienced AI professionals. The moderator conducted the participants in the use of the Guide, allowed interaction between the participants when using the Guide and eliciting requirements, and after this phase addressed a script of questions. Focus group is a quick and low-cost method used to obtain information related to the experiences of professionals and users of some technology/product/service, providing qualitative information [Kontio et al. 2004]. Focus groups are "carefully planned discussions designed to elicit the perceptions of group members about a defined area of interest" [Kontio et al. 2004].

We used adaptations of Kotio's guideline [Kontio et al. 2004] to conduct the focus group. The steps for evaluating the Guide through the focus group are presented in Figure 4, adapted for our case, and arranged into three phases: planning, execution and analysis.

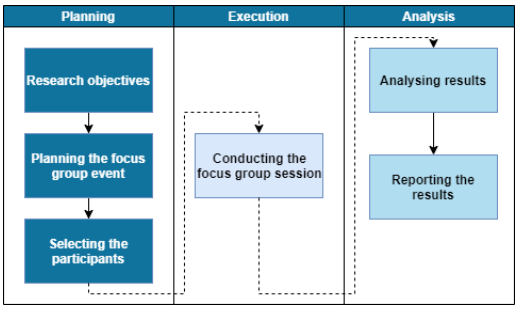

Figure 4. Steps for the evaluation of the Guide through the Focus Group. Adapted from [Kontio et al. 2004] e [Kitchenham and Pfleeger 2002]

We began in the planning phase with the steps of: defining the objectives and research questions of our evaluation – creating and defining the objectives of the feedback obtained from the use of our Guide, and the script of questions to be answered by the participants; planning the focus group event – setting a pre-determined structure and sufficient time for participants to understand the issues and meaningful discussions to occur [Kontio et al. 2004]; selecting participants – choosing representative, experienced, and

motivated participants. Then, in the execution phase: conducting the focus group session – presenting an introduction of the objectives and fostering discussion and interaction, and video recording of the session was made. In the analysis phase: analysing and reporting the results, and finally, the final discussion.

The overall aim of the evaluation of the Guide is to provide evidence that its use may have an impact on the governability of AI-based systems. For this purpose, participants working in software development teams that implement AI-based systems were selected. They were invited to use the Guide and to answer a question script composed of the same questions present in the survey, with the addition of one question: Do you think the Guide helped the software development team to identify and elicit the ethical requirements?

The focus group was composed of 5 professionals who develop AI-based systems. The working area of these professionals are: major Brazilian banks (private and public), private companies that provide services to public organizations, and major multinational IT companies. The session began with an overview of the study objectives and a discussion of how participants should act during the session [Kontio et al. 2004]. Participants were encouraged to collaborate with each other, and the organiser ensured the confidentiality and anonymity of the discussion. We chose the Product and System User Testing technique [Langford and McDonagh 2002] for the focus group session. This is because of the technique's main advantages: feedback is based on participants' experience of performing real tasks; providing stimulus for discussion; useful when evaluating systems or prototypes under development. Observations (comments made throughout the session) and a question script were used to provide additional information. The moderator made available the link of the Guide, and shared the screen, in which the participants experienced the assisted use of the guide, exploring the ethical requirements elicitation, and collaborated with each other through this task. Next, the questions in the questionnaire were read aloud, and participants were able to answer the questions, allowing for a discussion. Finally, participants were asked to present their final considerations.

5. Results

5.1. Focus Group Analysis

Regarding the use of the Guide, the participants created a hypothetical scenario of an AI-based essay correction system, to be employed by a selection board for admission to an institution. The system aims to select candidates by assigning a score to their essays. Throughout the session, they addressed three cards, and answered questions in the Issues to be addressed section. As detailed in the Guide information, they should start with card 0, on stakeholder assessment. After that, they chose among themselves the next cards to be addressed – cards 7 and 6 – where various ethical requirements were elaborated by the participants.

5.1.1. Card #0: Stakeholder's Assessment. Analysis made by the group:

The different stakeholders identified are: the selection board, the candidates, and the institution. These have contractual interests in selecting the best candidate, given the available criteria. "The system is the one that decides the selection, it can select the candidates

automatically, but if it has any bias, it does so in a wrong way, negatively impacting the stakeholders." Collaboration among multiple stakeholders was not identified.

5.1.2. Card #7: Interpretability, Showing - Transparency. Ethical requirements elicited:

1. The algorithm should be known to the candidates.
2. The dataset should be disclosed after the application of the test. This can help students to prepare for the next exams, since the theme of the essay is modified in each selection process.
3. It should be communicated widely that the correction is being done automatically by a system, in the public notice, in the form of a contract.
4. Regarding compliance with the General Law on Data Protection (LGPD): There should be no personal data within the dataset, and users' data should be anonymised.

5.1.3. Card #6: Explainability, Explicability, Disclosure - Transparency. Ethical requirements elicited:

1. The system should not retain personal data.
2. The system should ensure that a candidate receives their own score, not someone else's.
3. Data used for training must be anonymous.
4. Reproducibility of the system should be allowed by the Ministry of Education.
5. Documented system code should be disclosed to external auditing bodies.
6. The system should be periodically monitored, with a portion of the training data, and made publicly available – create these monitors and publicly disclose the results.

In Card #6: Questions related to testing. Ethical requirements elicited:

1. Professors on the board should be able to perform curation by evaluating an essay and observing the results of the AI-based system, thus enhancing the system – Enable a curation process.
2. Record how and who performed the curation process, make reports of that process (making logs of the curation).
3. It should be possible to explain and document the code and metrics involved.
4. Make available non-technical documentation of the system part – make metrics public.
5. Tests that fail (e.g. essays that receive unduly low scores) should be exposed to the public.
6. Appropriate metrics should be used in order to publicly demonstrate the percentage of successes, as well as which essays were wrongly graded or scored.
7. It should be checked in which cases the metric was not satisfied.
8. The cases where the AI-based system is not able to repair should be identified, and then a curation process should be undertaken – a follow up by a certified essay proofreader.

Regarding the tools presented, the participants explored the tool **InterpretML** on GitHub. It was identified that the tool helps to explain how the model works, as well as why the system acts in a certain way. The advantages found are: "finding out if the model being developed is doing what we want, if the dataset is sufficient, or if counter example data is needed", "Helps to find out how the model works and how to improve it, as well as finding out the reasons, of the model's classification features, even when using a black box model".

Regarding the question script addressed, they assessed that the principles present are not so basic: "propose questions that you would never ask yourself nor raise the possibility of the problem, and each card brings discussions of hours, it's very complex, and it all becomes easy to visualize, in items". They have never had contact with the suggested tools, but believe that the tools have application to ethics in AI. 3 participants considered that the principle of Transparency is easier to apply (in particular, card #5). While Dignity is the most difficult, because "not all systems reach that point". Only 2 participants mentioned that no principle is easily implementable, because "all of them are very complex, one should be careful not to see them as simple". Concerning the questions presented in the cards, they mentioned in relation to their practicality: "very practical, direct and objective". However, they also mentioned that they can be very broad: "the questions are clear, but the answers are not so clear." They pointed out the help of the questions for the elicitation of ethical requirements: "When provoked, we remember curating process, public notice, among others."

The participants, acting as a software development team, reported that the Guide helped to identify and elicit ethical requirements: "The guide helped us a lot to elicit ethical requirements, because the Guide is well structured, well divided and in a simple way." Also, a Product Owner present would help to have more questions to be raised. Overall, the Guide has improved the ethical awareness and learning of the participants: "By reading all the cards, we see principles that we were not aware of." 4 participants considered the requirements analysis phase as the most suitable phase of the software development process for using the Guide, while only 1 considered the implementation and maintenance phase. All participants stated that they would use the RE4AI Ethical Guide in requirements elicitation in their projects.

Several suggestions for improvement were offered. The Guide was considered to be too lengthy, and they suggested reducing the scope of the Guide and providing a "reduced version". Due to the high coverage and complexity of the various questions, they suggested more guidance on the use of the Guide: "guidance on which cards and tools to use for specific problems". This is also due to the lack of availability of a Product Owner, who usually works in areas outside IT, and: "would not have enough time to use this guide". In addition, they suggested a separation of the Guide into categories: "possible phase divisions, such as documentation, testing, coding and maintenance". Finally, they considered the scope of the Guide of great interest, where "just the content already provides considerable knowledge" and "Fundamental to bring the debate of this problem, which is already a current problem that impacts everyone". They pointed out that there are other ways of benefiting from the content: "not just as Planning Poker, but the reading itself already raises the debate". They mentioned the possible application of the Guide acting as a checklist: "to assess and grade an already implemented system."

We identified 6 perceived positive points, such as: a) the support information presented is adequate for understanding and use; b) the questions contained in the cards are easy to understand – objective and clear; c) the use of the Guide helps the creation of user stories through the questions in the cards; d) there is an increase in ethical awareness through the use of the Guide; e) applicability of the Guide in the requirements elicitation phase; f) there is an interest from the participants in using the guide in the requirements elicitation phase in their future projects. In addition, from the focus group we noted the usefulness and practicality of the Guide in eliciting requirements in development teams, since the participants were able to elicit 18 requirements for a hypothetical scenario, during the session.

Our findings suggest that the RE4AI Ethical Guide is perceived to be of great interest by participants, receiving an overall positive evaluation. The Guide, by operationalising ethical principles, can help mitigate challenges present in the literature, such as: lack of tools to implement AI ethics at the project level [Morley et al. 2019], [Vakkuri et al. 2020b]; lack of tools that assist software development teams as a whole [Vakkuri et al. 2020a]; with practicality and usability offering help to be used in practice [Morley et al. 2019]; as well as the lack of tools that do not focus mostly on explicability [Morley et al. 2019].

We observed 5 negative points and suggestions for improvement offered by the participants, such as: a) the suggested tools are not known by the participants; b) very extensive and broad guide, suggesting a reduced version (reduction of the scope) with cards and tools oriented to a particular context/problem; c) to divide the Guide in categories for the phases of documentation, tests, codification and maintenance; d) make the Guide available in other languages; e) offer an order of importance of the principles (prioritization of the principles). These problems will be mitigated in future works and in further versions of the Guide.

Concerning the generalization of the focus group results, like other qualitative studies, they have usual limitations on this topic [Langford and McDonagh 2002]. We had available only 5 professionals who develop AI-based systems in different contexts, which influenced the generalization of the final results, increased by the fact that their answers are affected by the organization's goals and previous experiences on the subject. Also, the results presented are preliminary and the work has not been used in real cases. However, despite the fact that generalization is not possible, these data are valid and complementary with other studies, e.g., increasing the number of participants and using the Guide in real scenarios.

6. Academic and Social Relevance

When we consider the academic relevance of our work, the impact of this dissertation starts with the three published papers [Cerqueira et al. 2021a, Cerqueira et al. 2021b, Cerqueira et al. 2022], the first is focused on finding suitable ethical principles to work with, the second is focused on exploring literature through a meta-analytic approach and finding tools, methods and techniques that implement AI ethics, and the third is focused on presenting the proposed Guide and its development process. A fourth paper is in production that aims to present a systematic literature review on this topic in a journal. Our study advances the ECCOLA method, adding cards, principles, possible tools, and

making it as a web-based system, not only static. Furthermore, we evaluated our Guide through a focus group, obtaining great feedback. Negative ethical impacts of AI is happening incessantly, and to operationalise AI ethics is an on going effort. Our Guide has great social relevance due to its potential in mitigating various negative ethical impacts of AI in our society, through the elicitation of ethical requirements, enabling ethical principles: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, sustainability, dignity and solidarity.

References

Araujo, R., Maciel, R., and Boscarioli, C. (2017). I grandsi-br: Grandes desafios de pesquisa em sistemas de informaçao no brasil (2016-2026). *Relatório Técnico. Comissão Especial de Sistemas de Informação (CE-SI) da Sociedade Brasileira de Computação (SBC). 67p*.

Belani, H., Vukovic, M., and Car, Z. (2019). Requirements engineering challenges in building ai-based complex systems. In *27th IEEE International Requirements Engineering Conference Workshops, RE 2019 Workshops, Jeju Island, Korea (South), September 23-27, 2019*, pages 252–255, 10.1109/REW.2019.00051. IEEE.

Benjamins, R., Barbado, A., and Sierra, D. (2019). Responsible ai by design in practice. *arXiv*, 1:1–10.

Cerqueira, J. A. S., Acco Tives, H., and Canedo, E. D. (2021a). Ethical guidelines and principles in the context of artificial intelligence. In *XVII Brazilian Symposium on Information Systems*, pages 1–8.

Cerqueira, J. A. S., Dos Santos Althoff, L., Santos De Almeida, P., and Canedo, E. D. (2021b). Ethical perspectives in ai: A two-folded exploratory study from literature and active development projects. In *Proceedings of the 54th Hawaii International Conference on System Sciences*, page 5240.

Cerqueira, J. A. S., Pinheiro De Azevedo, A., Acco Tives, H., and Canedo, E. D. (2022). Guide for artificial intelligence ethical requirements elicitation - re4ai ethical guide. *Proceedings of the Annual Hawaii International Conference on System Sciences*.

Commission, E. (2019). *Ethics Guidelines for Trustworthy AI High-Level Expert Group on artificial intelligence*.

de Ágreda, Á. G. (2020). Ethics of autonomous weapons systems and its applicability to any ai systems. *Telecommunications Policy*, 44:101953.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., and Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *SSRN Electronic Journal*, 1(1):1–39.

Floridi, L. and Cowls, J. (2019). A unified framework of five principles for AI in society. *Issue 1*, 1.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., and Vayena, E. (2018). Ai4people - an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach.*, 28(4):689–707.

Guizzardi, R. S. S., Amaral, G. C. M., Guizzardi, G., and Mylopoulos, J. (2020). Ethical requirements for AI systems. In *Canadian Conference on AI*, volume 12109 of *Lecture Notes in Computer Science*, pages 251–256, https://doi.org/10.1007/978-3-030-47358-7_24. Springer.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds Mach.*, 30:99–120.

IEEE (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems (EAD FirstEdition)*.

Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, 1(9):389–399.

Kitchenham, B. A. and Pfleeger, S. L. (2002). Principles of survey research part 4: questionnaire evaluation. *ACM SIGSOFT Softw. Eng. Notes*, 27(3):20–23.

Kontio, J., Lehtola, L., and Bragge, J. (2004). Using the focus group method in software engineering: Obtaining practitioner and user experiences. In *2004 International Symposium on Empirical Software Engineering (ISESE 2004), 19-20 August 2004, Redondo Beach, CA, USA*, pages 271–280. IEEE Computer Society.

Kostova, B., Gurses, S., and Wegmann, A. (2020). On the interplay between requirements, engineering, and artificial intelligence. In *Proceedings of REFSQ-2020 Workshops, Pisa, Italy, 2020*, pages 1–5.

Krafft, T., Hauer, M., Fetic, L., Kaminski, A., Puntschuh, M., Otto, P., Hubig, C., Fleischer, T., Grünke, P., Hillerbrand, R., Hustedt, C., and Hallensleben, S. (2020). From principles to practice - an interdisciplinary framework to operationalise ai ethics.

Langford, J. and McDonagh, D. (2002). Focus groups: Supporting effective product development.

Lwakatare, L. E., Raj, A., Bosch, J., Olsson, H. H., and Crnkovic, I. (2019). A taxonomy of software engineering challenges for machine learning systems: An empirical investigation. In Kruchten, P., Fraser, S., and Coallier, F., editors, *Agile Processes in Software Engineering and Extreme Programming - 20th International Conference, XP 2019, Montréal, QC, Canada, May 21-25, 2019, Proceedings*, volume 355 of *Lecture Notes in Business Information Processing*, pages 227–243. Springer.

Mayer, A.-S., Haimerl, A., Strich, F., and Fiedler, M. (2021). How corporations encourage the implementation of ai ethics. In *ECIS*.

Méndez, D., Graziotin, D., Wagner, S., and Seibold, H. (2020). Open science in software engineering. In Felderer, M. and Travassos, G. H., editors, *Contemporary Empirical Methods in Software Engineering*, pages 477–501. Springer.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11):501–507.

Morley, J., Floridi, L., Kinsey, L., and Elhalal, A. (2019). From what to how: An initial review of publicly available ai ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, pages 1–28.

Newman, J. (2020). *Decision Points in AI Governance: Three case studies explore efforts to operationalize AI principles*. Center for Long-Term Cybersecurity - UC Berkeley.

Nguyen-Duc, A., Sundbø, I., Nascimento, E., Conte, T., Ahmed, I., and Abrahamsson, P. (2020). A multiple case study of artificial intelligent system development in industry. In Li, J., Jaccheri, L., Dingsøyr, T., and Chitchyan, R., editors, *EASE '20: Evaluation and Assessment in Software Engineering, Trondheim, Norway, April 15-17, 2020*, pages 1–10. ACM.

Raji, I. D., Smart, A., White, R., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., and Barnes, P. (2020). Closing the ai accountability gap: defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*.

Rothenberger, L., Fabian, B., and Arunov, E. (2019). Relevance of ethical guidelines for artificial intelligence - a survey and evaluation. In vom Brocke, J., Gregor, S., and Müller, O., editors, *27th European Conference on Information Systems - Information Systems for a Sharing Society, ECIS 2019, Stockholm and Uppsala, Sweden, June 8-14, 2019*, volume 27, pages 1–12.

Ryan, M. and Stahl, B. C. (2020). Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *Journal of Information, Communication and Ethics in Society*.

Schiff, D., Ayesh, A., Musikanski, L., and Havens, J. C. (2020a). Ieee 7010: A new standard for assessing the well-being implications of artificial intelligence. *ArXiv*, abs/2005.06620.

Schiff, D., Rakova, B., Ayesh, A., Fanti, A., and Lennon, M. (2020b). Principles to practices for responsible ai: Closing the gap. *arXiv preprint arXiv:2006.04707*.

Schleier-Smith, J. (2015). An architecture for agile machine learning in real-time applications. In Cao, L., Zhang, C., Joachims, T., Webb, G. I., Margineantu, D. D., and Williams, G., editors, *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, August 10-13, 2015*, pages 2059–2068. ACM.

Smit, K., Zoet, M., and van Meerten, J. (2020). A review of AI principles in practice. In Vogel, D., Shen, K. N., Ling, P. S., Hsu, C., Thong, J. Y. L., Marco, M. D., Limayem, M., and Xu, S. X., editors, *24th Pacific Asia Conference on Information Systems, PACIS 2020, Dubai, UAE, June 22-24, 2020*, page 198.

Vaishnavi, V. K. and Kuechler, W. (2015). *Design science research methods and patterns: innovating information and communication technology*. Crc Press.

Vakkuri, V., Kemell, K., and Abrahamsson, P. (2020a). ECCOLA - a method for implementing ethically aligned AI systems. *CoRR*, abs/2004.08377.

Vakkuri, V., Kemell, K.-K., Kultanen, J., and Abrahamsson, P. (2020b). The current state of industrial practice in artificial intelligence ethics. *IEEE Software*.

Zeng, Y., Lu, E., and Huangfu, C. (2019). Linking artificial intelligence principles. *ArXiv*, abs/1812.04814.