

## Guide for Artificial Intelligence Ethical Requirements Elicitation – RE4AI Ethical Guide

<p>José Antonio Siqueira de Cerqueira University of Brasília (UnB) <a href="mailto:jose.siqueira.cerqueira@gmail.com">jose.siqueira.cerqueira@gmail.com</a></p>	<p>Anayran Pinheiro de Azevedo University of Brasília (UnB) <a href="mailto:anayran.azevedo@gmail.com">anayran.azevedo@gmail.com</a></p>	<p>Heloise Acco Tives Federal Institute of Paraná (IFPR) <a href="mailto:heloise.acco@ifpr.edu.br">heloise.acco@ifpr.edu.br</a></p>	<p>Edna Dias Canedo University of Brasília (UnB) <a href="mailto:ednacanedo@unb.br">ednacanedo@unb.br</a></p>
---	--	---	---

### Abstract

*Development and use of Artificial Intelligence (AI) based systems are growing at a fast pace in our society, simultaneously ethical concerns are arising from them. Addressing AI ethics is a continual issue and has provoked much debate among researchers. The aim of this work is to provide a Guide for Artificial Intelligence Ethical Requirements Elicitation (RE4AI Ethical Guide). The Design Science Research methodology was adopted in order to understand the problem, develop a prototype and evaluate it through a survey. The proposed Guide, composed of 26 cards along 11 ethical principles, is both useful and practical and can help in the elicitation of ethical requirements for AI in the context of agile development. 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 to elicit ethical requirements and implement ethics in AI.*

### 1. Introduction

The evolution of the emergence of software that makes use of Artificial Intelligence (AI) techniques, mostly Machine Learning (ML), amplifies the manifestations of accidents and the awareness of the associated ethical issues [1]. As an example, a ML algorithm with extensive use in hospitals in the United States had a racial bias when referring black people to better health programmes [2]. AI-based systems developed without proper ethical cautions are prone to negatively impact society at large, e.g., in criminal justice, education, healthcare [2].

In general, ethics in AI has been addressed, in the literature, in its theoretical field, through ethical guidelines [3]. While the existence of guidelines and principles is necessary, little practical direction exists for developers – those responsible for implementing ethics

in AI-based systems – to apply in real contexts, even more with the market delivery demands [3], where often the ethical considerations involved is a quality to be considered in the software only after its deployment [4]. Furthermore, developers do not receive adequate training within development projects, nor during their academic studies. There are no legal consequences for not implementing AI ethics, as the guidelines present in the literature, and proposed by organisations, are often non-binding laws (soft law). Thus, there is neither motivation nor punishment for developers in the area of AI ethics.

During the requirements elicitation phase there is a greater interaction between different actors involved in software development and its use, providing a fertile environment for debate on ethical issues [5], and there is a reduction in additional work by considering ethical issues in the early stages of software development, rather than as an afterthought [4].

Guizzardi et al. [6] presented a discussion on ethical requirements, pointing out that well-established techniques used in Requirements Engineering can also be used to develop AI-based systems in compliance with ethical principles and guidelines. Although they argue that traditional Requirements Engineering techniques can be used for such task, they do not detail which technique is used, neither define its steps, and the context of autonomous cars is the only AI-based system addressed.

Vakkuri et al. [4] presented a method to implement ethics in AI-based systems, called ECCOLA. 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 [7] and IEEE EADv1 [8] 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.

To mitigate 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.

This Guide differ from the ECCOLA method by Vakkuri et al. [4] 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. Furthermore, the addition of tools suggestion in the content of the cards, as well as extensive supporting material (how to use, principles, tools, trade-offs). Moreover, we considered in our guide all the 11 principles listed by Ryan and Stahl [9] and presented 26 cards, while in the ECCOLA method only 7 principles are observed with a total of 21 cards. We also found the need for the inclusion of traditional software engineering practices, such as requirements elicitation, for the context of Artificial Intelligence, in addition to the characteristics of a Guide to implement ethics in AI [10]: broad, operationalisable, flexible, iterative, guided and participatory.

For its creation, the Design Science Research methodology [11] was adopted in order to understand the problem, develop a prototype and evaluate it through a survey with undergraduate and graduate students. 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.

Additional non-exhaustive concepts, background and literature review relevant to understand the context of this work are provided in our previous study [12], considering that this paper is a continuation of our preceding study.

## 2. Initial Guide criteria

We took into consideration the approach devised by Schiff et al. [10] – broad, operationalisable, flexible, iterative, guided, participatory – to bring ethical principles in AI closer to practice, and made some adjustments to our study. The proposed Guide will be **broad**, by considering different ethical principles, besides indicating possible tools of restricted scope with usability for a specific principle, such as XAI tools for the Transparency principle (e.g., InterpretML [13] and TransparentAI [14]).

The Guide will be **operationalisable** by allowing users to elicit requirements and include them in their Sprint backlogs in the form of user stories, being part of a larger context in which are included the functional and non-functional requirements that the system must fulfill, in addition to the ethical ones. This criterion is aligned with the purpose of our guide, to help the creation of ethical user stories to serve as items of the product backlog in an agile software development context. In other words, the system requirements are the user stories, which are present in the product backlog, i.e., a list of requirements, that will be worked on by the development team in iterations, called sprints, lasting from 1 to 4 weeks [15].

The guide will be **flexible**, since the cards have open questions and there are no single answers or only one context for applying the guide. The Guide will be **iterative**, as users are free to decide the best moment to use the cards, and they can be reused, with the inclusion of different stakeholders in the process. The Guide will be **guided** as there will be user documentation, providing the user with a prompt familiarization with the system in a simple and intuitive way. The guide will be **participatory**, i.e., different stakeholders of the organization can be part of the ethical requirements elicitation activity, including users participation in software development meetings. Our artifact, therefore, will be a guide that will serve as an assistant to support the elicitation of ethical requirements in the context of AI-based systems.

### 3. Development of the initial prototype of the Guide

In this phase, we will follow a set of steps for the development of the RE4AI Ethical Guide. First, we will delimit the set of ethical principles to be used, then we will define the set of possible tools that will serve as a suggestion to developers, finally, we will present the cards that will compose the deck, where each card will comprise a tool and a principle, besides other pertinent information, with the goal of creating a prototype and presenting it. Figure 1 presents an overview of the steps that will guide the process of developing the Guide content.

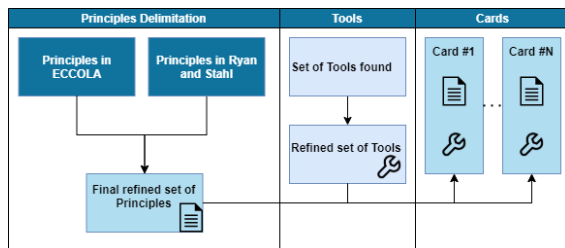


Figure 1. Exploration diagram of the development of the Guide for Elicitation of Ethical Requirements in AI. Own source

#### 3.1. Ethical Principles delimitation

There are several guidelines containing ethical principles serving as normative guides for AI ethics, and as of November 2019, at least 84 organisations – public, private, government, the academia and civil society – have been publishing reports describing ethical principles, values or other abstract high-level requirements for the development and deployment of AI [3]. Therefore, there is an initial challenge in choosing which ethical principles will be used in the development of the guide proposed in this work.

Several authors have proposed compilations of various ethical guidelines into few principles, e.g., [16], [17], [18], [19], [20], [21]. To the best of our knowledge, Ryan and Stahl [9] did the most comprehensive compilation, with 91 guidelines, aimed at developers and users. Thus, in the development of the Guide, it will be taken as central axis the principles listed by Ryan and Stahl [9]. We standardized the principles in Vakkuri et al. [4] with the principles in Ryan and Stahl [9] through mappings presented in tables. However, it is a challenge to include all the principles and ethical issues identified by [9] – for example: how to elicit requirements related to the principles of Solidarity, or Dignity? Moreover, the ethical principles present in the

ECCOLA method cards are not part of the same set of principles identified in [9], and are often named in different ways, but with similar content.

Initially, there are seven principles present in the ECCOLA method – based primarily on the AI HLEG [7] and the IEEE EADv1 [8]. Both set of principles and their mapping are presented in Tables 1, 2 and 3 of our supplementary material <https://zenodo.org/record/5520174>. In the ECCOLA method, for each principle, there are a number of cards that may contain an ethical issue distinct from another in its same set. In other words, within a set of cards of the same principle, there is more than one ethical issue associated. For this reason, we related each ECCOLA card to a principle and an ethical issue in Ryan and Stahl. Standardizing these principles allows us to reuse the ECCOLA method cards, as well as being an initial starting point for our Guide.

#### 3.2. Tools that assist the implementation of AI ethics

In this next step, we identified the tools that will be presented in the content of the cards. We explore the tools identified in our previous work [12], where only 21 tools that assist the implementation of AI ethics were highlighted. In this work, the tools were mapped with principles and ethical issues presented by Ryan and Stahl – the resultant relation to selected principles are presented in Table 4 of our supplementary material <https://zenodo.org/record/5520174>.

Repositories were identified that have no usefulness to our goal (T16, T20, T21) – **Discarded** – because they contain supporting code, curated list (e.g., a list on a given topic that has been carefully compiled, usually by a survey) of tools outside our scope, and address ethical dilemmas in AI. In addition, some tools will be present in the supporting material of the Guide, as an extra reference, however, they will not be in the body of any card (T2, T3, T5, T6, T9, T10, T11) – **Apart** –, because they do not present a practical tool, but include information pertinent to developers about implementing ethical principles in AI in some way (e.g., educational tool, privacy implementation in a project). The tools included in the content cards of our guide (T1, T4, T7, T8, T12, T13, T14, T15, T17, T18, T19) – that actually assist in implementing ethical principles in AI-based systems – total in 11 tools, composing our refined set of tools, visible in the Tools lane of Figure 1.

From the initial set of tools (21), we filtered 11, and based on their relationships with the set of principles, we note that there is a trend in tools operationalising the principles of Transparency and Justice and fairness.

Some tools operationalise more than one principle:

1. Transparency - T1, T4, T7, T8, T17, T19 - Total: 6
2. Justice and fairness - T7, T12, T13, T14, T15, T17, T18. Total: 7
3. Non-maleficence - T7 - Total: 1
4. Responsibility - T7 - Total: 1
5. Sustainability - T7 - Total: 1

#### 4. Cards Contents

The Guide for Artificial Intelligence Ethical Requirements Elicitation will consist of a deck of cards. The cards will be separated by principles and ethical issues. Each principle may have more than one ethical issue, i.e., more than one card will be available for each principle. Each card is composed of four parts:

1. **Preamble** – why this is important;
2. **Issues to be addressed** – to tackle this issue;
3. **Illustration** of this topic – to further exemplify the issue;
4. **Tool Suggestion** – tools available on GitHub that support the implementation of the ethical issue.

Items 1, 2 and 3 are adapted from Vakkuri et al. [4] and are available in our prototype. In **Preamble**, there is an observation of why it is important to address this issue, as something positive to be achieved that reflects on the user in the end, or provides an overview of that topic. It is noticeable that, in **Issues to be addressed**, there is no single, direct and objective indication of what developers should do, but there are questions, which users of the guide need to discuss, in order to operationalise ethics in AI. In this way, ethical awareness among the development team is increased. In **Illustration** we further illustrate the issue by offering a case where ethical requirements were not considered and led to incidents, or the illustration of the topic in a specific context. In **Tool Suggestion** we offer the options of available tools in the refined set of Tools, however, it was seen that this set of Tools does not cover all the principles in the Guide, i.e., this field is not mandatory and will not appear in all the cards, as there are no tools available for all ethical issues.

Two more cards compose the guide, related to stakeholder analysis and assessment. Important to note that, in the ECCOLA method originally conceived there is an initial card, which must be addressed before the

user explores other cards, called Stakeholder Analysis. This card motivates the developer to answer questions related to stakeholder assessment, who they are, how they are affected and how they are related. Stakeholder analysis, as described by Vakkuri et al. [4] converges with the concept of Runtime Stakeholders introduced by Guizzardi et al. [6]. The latter work, states that a key concept for obtaining/eliciting ethical requirements is that of Runtime Stakeholders: “These include those stakeholders that are using, affected by, or influencing the outcomes of a system as it is operating”. An example of Runtime Stakeholders in the context of AI-based systems for healthcare are: patients and their families, the doctors, nurses, x-ray operators and other healthcare professionals. Thus, this card shows to be crucial to allow the other cards to be applied.

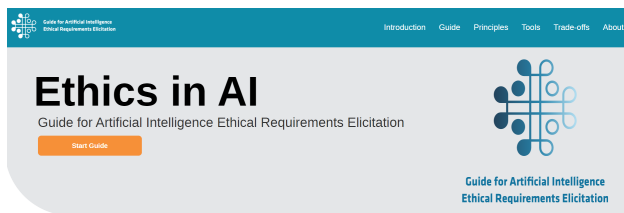
The second additional card, deals exclusively with the evaluation of the AI-based system being developed. Ethical evaluation must become an integral part of the operation of a system, or there is no guarantee that tools – such as this guide – will have any positive impact on the ethical implications of AI systems [1]. There is a caution to insert on some cards throughout the deck, under “Issues to be addressed”, whether the system allows for evaluation (e.g., internal, or external, and to what extent), however, it is observed the urgency of inserting a card exclusively for this purpose, in that oversight, at the evaluation stage, is “concerned with whether the algorithmic system is continuing to operate in the right way once deployed, needs to be revised, or can be improved” [1]. In other words, even after the system is deployed, the development team should define a time interval between one evaluation and another. The need for the inclusion of this card also emerges from the distribution of responsibility between the components defined by Morley et al. [1] on AI ethics governance: an independent multi-disciplinary ethics board; and the AI professionals themselves. According to Morley et al. [1], positive ethical qualities are susceptible to progressive increase, that is, “an algorithm can be increasingly fair, and fairer than another algorithm or a previous version, but makes no sense to say that it is fair or unfair in absolute terms”.

In sum, the Stakeholders’ assessment card will be card # 0, as in Vakkuri et al. [4], to be addressed at the beginning (before the other cards), and we have added a card dealing with system assessment, as a last card, which should be periodically revisited.

#### 5. Guide Overview

The Guide for Artificial Intelligence Ethical Requirements Elicitation is available at

<https://josesiqueira.github.io/RE4AIEthicalGuide/index.html> and its source code at <https://github.com/josesiqueira/RE4AIEthicalGuide>. RE4AI Ethical Guide 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 2 we present the initial screen of the guide, where its subdivisions are present.



### Introduction

This guide assists the elicitation of Ethical Requirements in AI-based systems by software development teams, with a focus on agile teams, where Product Owners and developers use a set of cards to answer questions related to ethical principles in AI at each sprint, creating requirements in the form of user stories and including them in Sprint backlogs.

Figure 2. 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 3.

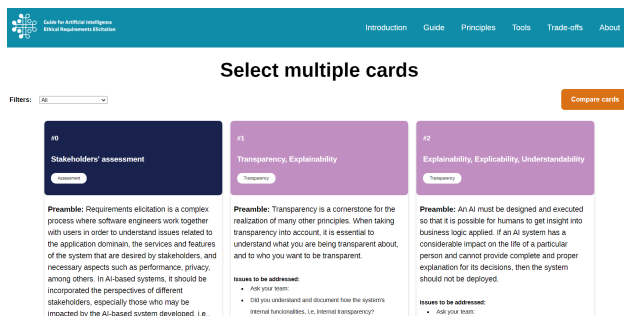


Figure 3. Card selection page. Own source

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

Initially 24 cards are provided, distributed along the 11 principles adopted for the elaboration of the guide, plus 2 additional cards: Stakeholders' assessment, and

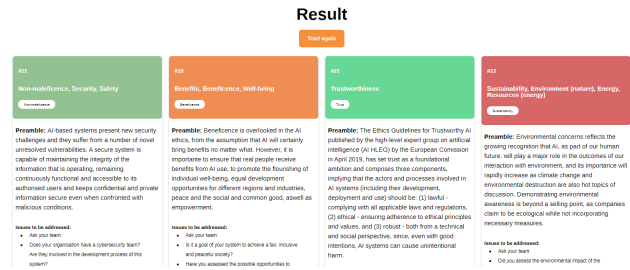


Figure 4. Different cards compared. Own source

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.

## 6. Evaluation

The objective of the evaluation of the Guide through a survey is to verify the viability of the guide, as well as the perceptions of users about the content provided. According to Morley et al. [1], there is little evidence that the use of tools that operationalize AI ethics impacts the governability of a system. Thus, 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 the planning of the survey, its execution and analysis of the responses, the guideline proposed by Pfleeger and Kitchenham [22] and the phases proposed by Moller et al. [23] were used. In preparing the questionnaire, the guidelines presented by Kitchenham and Pfleeger [24] were used.

We started in the planning phase with the following steps: defining the objectives and questions of our evaluation – we created and defined the objectives for obtaining feedback from the use of our Guide, and the questions to be answered by the participants; choosing participants – undergraduate and graduate students. Then, in the execution phase: we designed a questionnaire – we created questions of different types with Likert-scale and open-ended questions; and administered the questionnaires. Finally, in the analysis phase: we analyzed and reported the results – presenting the qualitative analysis through the Krippendorff [25] content based analysis technique, and the quantitative analysis.

The participants are undergraduate and graduate

students of the disciplines of Data Science, Requirements Analysis, Artificial Intelligence and Cognitive Computing, and Artificial Intelligence of the semester of 2021.1, that were invited to use the guide and answer a questionnaire, remotely. The disciplines were lectured at the University Center of Brasilia (UniCeub), in Brasilia-DF, Brazil. These disciplines are part of the curriculum of the courses of Computer Science, Computer Engineering and Systems Analysis. There was an average of 20 students per class. A total of 40 complete and 49 incomplete responses were received, amounting to 89 responses. Only completed questionnaires were considered, therefore 40 completed questionnaires were analysed.

**Q1: Regarding the supporting content present in the guide, was the information sufficient for its understanding and use?** Regarding the participants' opinion on the support material provided in the Guide, i.e., the introduction, principles, tools, and trade-offs sections, in relation to their understanding and usefulness, the majority of the participants, 24 (60%), agreed that the information presented in the Guide was sufficient for its understanding and use; 9 (22.5%) strongly agreed; and 7 (17.5%) neither agreed nor disagreed, as presented in Figure 5.

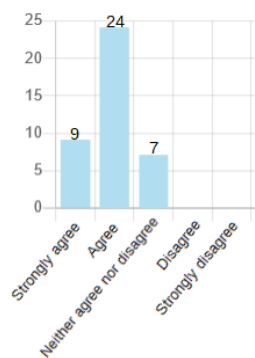


Figure 5. Feedback on the supporting content of the Guide

**Q2: Did you already know any of the tools suggested by the Guide?** Regarding the previous knowledge of the tools suggested by the Guide, most of them are unknown to the participants, 36 (90%) said they did not know any of the tools presented in the guide.

**Q3: Regarding the suggested tools, do you believe they have utility in implementing AI ethics?** Regarding the applicability of the tools, 11 participants (27.5%) strongly agreed about the applicability of the tools suggested by the Guide in the implementation of ethics in AI, 19 (47.5%) agreed, 8 (20%) neither agreed nor disagreed. Only 2 (5%) disagreed with its applicability.

**Q4: Which Principles do you consider most easily implementable?** Multiple answer. With regard to the perception about the applicability of the ethical principles in AI provided by the guide, 30 (75%) participants chose the principle of Transparency. This choice reinforces the idea that this is the principle that enables the other principles [26]. In second place, participants chose the principle of Responsibility (with 23) and in third place the principle of Privacy (with 21), as presented in Figure 6.

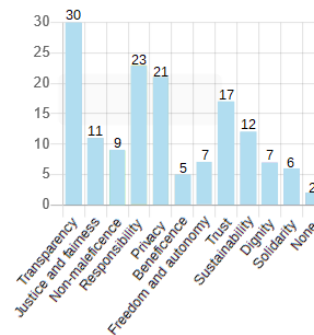


Figure 6. Applicability of ethical principles in AI

**Q5: Did you find the questions in the Guide cards easy to understand?** For the success in the elicitation of ethical requirements, it is imperative to understand the questions that must be answered by the development team. Regarding the ease of understanding of the questions available in the letters, 12 (30%) participants strongly agreed, 19 (47.5%) agreed, while 9 (22.5%) were neutral, as presented in Figure 7.

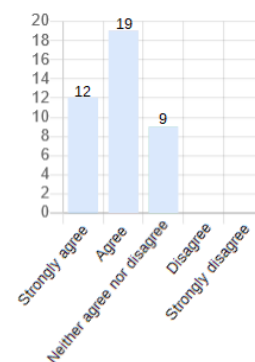
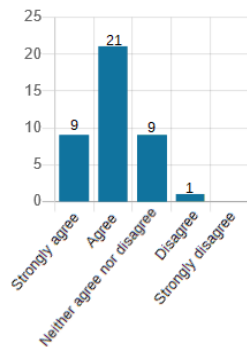


Figure 7. Comprehension of the questions available on the cards

**Q6: In relation to the questions present in the Guide cards, can the questions answered by the use of the cards help to elicit ethical requirements?** 9 participants (22.5%) strongly agreed regarding the usefulness of the answers obtained through the questions on the cards in creating user stories, 21 (52.5%) agreed,

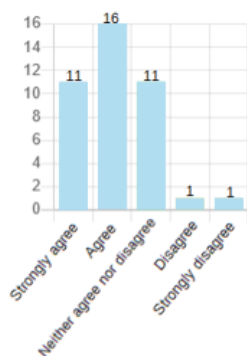
9 (22.25%) were neutral, while only 1 (2.5%) disagreed, as presented in Figure 8.



**Figure 8. Feasibility of eliciting requirements through the answers obtained**

The final objective of the proposed Guide is to assist in the creation of user stories. Some participants commented positively on this assistance: “The guide cards help a lot in the elaboration of clear user stories, placing emphasis mainly on what the software should do, for example making it explicit that there should be no discrimination of software users.”

**Q7: Has the Guide improved your ethical awareness and learning?** In addition to producing ethical requirements, the Guide is intended to assist in increasing the ethical awareness of its users. 11 (27.5%) of the participants strongly agreed that the Guide can increase ethical awareness, 16 (40%) said they agreed, 11 (27.5%) remained neutral, 1 (2.5%) disagreed, and 1 (2.5%) strongly disagreed, as presented in Figure 9.



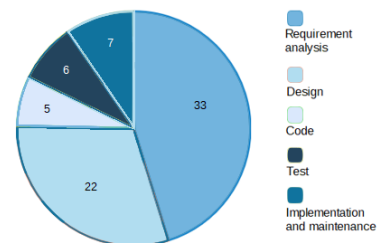
**Figure 9. Ethical awareness acquired through the use of the Guide**

Several participants stated that the Guide was helpful for learning and noted a positive experience regarding learning about the topic of AI ethics: “In the same way that it helps the team to keep the project on track, the guide is also helpful for learning with its consistent presentation of information and context”. Some

participants who had some superficial or no contact with issues related to AI ethics stated: “The guide helped me understand the importance of ethics in AI software projects and I had no idea about the principles that need to be taken into consideration when building software in this context”; “The guide opened my mind about AI ethics. This subject needs to be increasingly thought and discussed by the software development community, due to the evolution of systems that use some AI component and its direct and indirect impacts on end users social well-being”.

On the other hand, some participants stated that the proposed Guide is too extensive and commented on the practicality: “I learned a lot of new concepts from reading the guide. Although I miss something briefer to help in practice, like a checklist for the day-to-day, or a template for documentation related to ethical issues.”

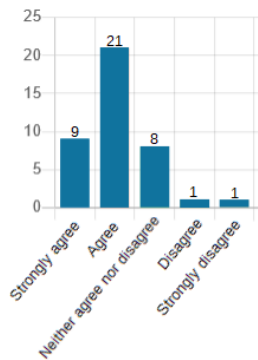
**Q8: At which stage of the software development process do you consider it most feasible to use the Guide?** Multiple answer. Regarding the applicability of the Guide in relation to the software development process phases, 33 (82.5%) of the participants consider it to be applicable in the Requirements Analysis phase, 22 (55%) in the Design phase, 5 (12.5%) in the Coding phase, 6 (15.5%) during the Testing phase, finally, 7 (17.5%) stated that it is applicable in the Implementation and Maintenance phase, as presented in Figure 10.



**Figure 10. Applicability of the Guide in relation to the software development phases**

**Q9: Would you use the RE4AI Ethical Guide in requirements elicitation?** About the future use of the Guide during the requirements elicitation phase of an AI-based system, 9 (22.5%) of the participants strongly agreed to use it, 21 (52.5%) agreed, 8 (20%) were neutral, 1 (2.5%) disagreed, and 1 (2.5%) strongly disagreed, as presented in Figure 11.

Overall, the Guide was well accepted by the respondents, in relation to practicality. Some comments were: “I would use it because of the practicality of addressing specific requirements elicitation contexts that can be improved”; “I would use the guide because of the practicality and the usefulness of the cards. I believe that these resources would help a lot in requirements



**Figure 11. Future use of the RE4AI Ethical Guide by participants**

elicitation”. Regarding commonly overlooked aspects of requirements: “The guide is very useful and is a good way, especially in the requirements phase, to not overlook commonly neglected aspects of building AI systems.”

**Q10: Do you have any suggestions for improving the Guide?** Some participants made suggestions for improvement:

- Regarding the extension of the content, one finds the need for the presentation of the Guide’s content in a reduced form and adjustments in the interface: “I missed a slightly more summarized version of the guide. A version that could be used for smaller and simpler AI projects, like a project that does not necessarily have a team of developers, but for example 1 or 2 people just working on an AI model that is going to be used or consumed by some other process.”
- In relation to how to use the guide, impacting its use process: “It would be interesting to have a prioritization of the principles, something that would give an order of importance. What should be treated or resolved with more priority, for cases where the software development company does not have enough resources or time to evaluate all the ethical principles proposed by the guide.”
- We also observed the need to make the Guide available in other languages, in order to facilitate understanding by users who are not proficient in English, or are not native speakers: “It would be interesting to make the content of the guide also available in Portuguese to facilitate understanding.”; “I couldn’t find a translation on the site and the google translation often leaves something to be desired, it would be nice to implement in other languages.”

## 7. Main evaluation discussions

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.

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 [26], [27]; lack of tools that assist software development teams as a whole [4]; with practicality and usability offering help to be used in practice [26]; as well as the lack of tools that do not focus mostly on explicability [26].

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).

## 8. Limitations and threats to validity

There is an impossibility to generalize the result of the survey, in part because it was conducted only in Brazil, with undergraduate and graduate students, who may have their perceptions impacted, both by their previous experiences and by the quality of the educational institution. Also, there was a difficulty in obtaining a significant number of participants. However, these limitations indicate opportunities to replicate this study in different countries and contexts. In order to mitigate this problem, a focus group will be conducted with AI professionals in a future work.

## 9. Final remarks

In this work a Guide for Artificial Intelligence Ethical Requirements Elicitation was developed, also referred to as RE4AI Ethical Guide, to assist the



implementation of ethics in AI by software development teams. For its creation, the Design Science Research methodology was adopted in order to understand the problem, develop a prototype and evaluate it. We identified the ECCOLA method [4] 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, for the context of AI, in addition to the characteristics of a Guide to implement ethics in AI [10]: broad, operationalizable, flexible, iterative, guided and participatory.

RE4AI Ethical Guide was developed, composed of 26 cards, containing the 11 principles in [9], as a web-based system allowing interactivity in the selection of the cards through filters and comparisons among multiple cards, as well as a support material. It was evaluated through a survey with 40 undergraduate and graduate students who evaluated the Guide through an online questionnaire.

It is seen in the literature that developers of AI-based systems may choose the ethical tools according to their needs, i.e., that are more convenient for them, that adhere to the limits of subjectivity of the text interpretation of ethical principles and the needs of the market in which the organization is inserted and its goals, instead of choosing those that are more aligned with the ethical understanding of the society as a whole (i.e., those affected by the running AI-based system) [28]. In order to mitigate this real and current problem, the proposed Guide takes into consideration the creation of documentation of the decision-making process of the translation of principles into ethical requirements by the development teams, through reports performed at each sprint, enabling auditability by different external agents, enabling the distribution of responsibility, and public judgment about possible manipulations such as ethics shopping – choosing ethical principles in order to adapt some pre-existing behaviors – and ethics washing – making misleading claims in order to appear more ethical than it is [29].

Furthermore, according to Morley et al. [1]: “if the responsibility for the entire process remains solely with the AI practitioners themselves, there remains the risk that the operationalization process itself will become subject to manipulation and may be used only for ethics washing purposes.” In order to address this problem, cards were included that encompassed different stakeholders in the considerations of the development team using the Guide, sometimes including them in the decision-making process.

The implementation of ethics in AI is an ongoing challenge, which should not be seen as a final goal that can be objectively achieved, as a checklist to be completed, but as a development process, i.e., a set of repeatable procedures, and re-evaluated on a recurring basis.

The operationalization of ethical principles and guidelines in AI is subject to the subjectivity of those involved in the elicitation process, and more importantly, in the developers. Among other features, we have tried to develop the guide so that it is perceived by users as a reflective development process, which helps AI practitioners to understand their own subjectivity and biases within a given set of circumstances [1].

There is a need for more work that focuses on teaching AI ethics in the training of future professionals as part of the curriculum adopted in courses related to the development of AI-based systems, in order to increase ethical awareness among students in computing courses, as well as the training of IT professionals by organizations.

We hope to contribute in the development of future research in the context of AI ethics, both in academia and industry, and in choosing tools and processes that support the implementation of ethics in AI-based systems, as well as raising awareness about the various ethical issues involved in the use of AI-based systems and their challenges in the development process.

Moreover, We identified the need for the implementation of a client side in the developed web-based system, in which software development teams can create and access an account, modify and insert cards through a graphical interface, and that users can store the elicited requirements in a user story format related to a particular card, i.e., to the ethical principle. These user stories can serve as examples to new users of the guide. In addition, other ethical requirements may arise. In this way, we suggest the use of Natural Language Processing for the creation of a dataset, where it would be possible to train these data and generate a machine learning model so that new user stories can be automatically validated.

For future works, further evaluations and examples of use of the guide are needed to identify the perceptions of a diverse set of AI practitioners in the use of the guide in different contexts and propose improvements. The traceability of ethical requirements in the implemented code is also an attractive field of research where attention is needed, as it requires the evaluation and understanding of what has been accomplished by the developers. It is interesting to provide ways to perform this task, as well as examples of these mappings –

between ethical requirements and code. Furthermore, to evaluate the application of the proposed Guide when it is desired to evaluate AI-based systems already developed and in use by users. Finally, we propose the elaboration of a catalogue or database of ethical requirements in AI devised by the use of our Guide. Overall, our work is an important cornerstone for enabling and steering such future research through the presentation and use of the proposed Guide, among other aspects.

## References

- [1] J. Morley, A. Elhalal, F. Garcia, L. Kinsey, J. Mokander, and L. Floridi, "Ethics as a service: a pragmatic operationalisation of AI ethics," *CoRR*, vol. abs/2102.09364, 2021.
- [2] H. Ledford, "Millions of black people affected by racial bias in health-care algorithms," *Nature*, vol. 574, pp. 608–609, Oct. 2019.
- [3] B. Mittelstadt, "Principles alone cannot guarantee ethical AI," *Nature Machine Intelligence*, vol. 1, pp. 501–507, Nov. 2019.
- [4] V. Vakkuri, K. Kemell, and P. Abrahamsson, "ECCOLA - a method for implementing ethically aligned AI systems," *CoRR*, vol. abs/2004.08377, 2020.
- [5] B. Kostova, S. Gurses, and A. Wegmann, "On the interplay between requirements, engineering, and artificial intelligence," in *Proceedings of REFSQ-2020 Workshops, Pisa, Italy, 2020*, pp. 1–5, 2020.
- [6] R. S. S. Guizzardi, G. C. M. Amaral, G. Guizzardi, and J. Mylopoulos, "Ethical requirements for AI systems," in *Canadian Conference on AI*, vol. 12109 of *Lecture Notes in Computer Science*, pp. 251–256, Springer, 2020.
- [7] European Commission, *Ethics Guidelines for Trustworthy AI High-Level Expert Group on artificial intelligence*, Apr. 2019.
- [8] IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems (EAD First Edition)*, 2019.
- [9] M. Ryan and B. C. Stahl, "Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications," *Journal of Information, Communication and Ethics in Society*, 2020.
- [10] D. Schiff, B. Rakova, A. Ayesh, A. Fanti, and M. Lennon, "Principles to practices for responsible AI: closing the gap," *CoRR*, vol. abs/2006.04707, 2020.
- [11] V. K. Vaishnavi and W. Kuechler, *Design science research methods and patterns: innovating information and communication technology*. Crc Press, 2015.
- [12] J. A. Siqueira De Cerqueira, L. Dos Santos Althoff, P. Santos De Almeida, and E. Dias Canedo, "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*, p. 5240, 2021.
- [13] H. Nori, S. Jenkins, P. Koch, and R. Caruana, *InterpretML - Alpha Release*, Oct. 2019. <https://github.com/interpretml/interpret>.
- [14] Nathanlauga, *InterpretML - Alpha Release*, Oct. 2019. <https://github.com/Nathanlauga/transparentai>.
- [15] V. T. Heikkilä, D. E. Damian, C. Lassenius, and M. Paasivaara, "A mapping study on requirements engineering in agile software development," in *41st Euromicro Conference on Software Engineering and Advanced Applications*, pp. 199–207, IEEE Computer Society, 2015.
- [16] T. Hagendorff, "The ethics of AI ethics: An evaluation of guidelines," *Minds Mach.*, vol. 30, pp. 99–120, 2020.
- [17] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of ai ethics guidelines," *Nature Machine Intelligence*, vol. 1, no. 9, pp. 389–399, 2019.
- [18] J. Fjeld, N. Achten, H. Hilligoss, A. Nagy, and M. Srikumar, "Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI," *SSRN Electronic Journal*, vol. 1, no. 1, pp. 1–39, 2020.
- [19] Y. Zeng, E. Lu, and C. Huangfu, "Linking artificial intelligence principles," *ArXiv*, vol. abs/1812.04814, 2019.
- [20] K. Smit, M. Zoet, and J. van Meerten, "A review of AI principles in practice," in *24th Pacific Asia Conference on Information Systems*, p. 198, 2020.
- [21] L. Floridi and J. Cows, "A unified framework of five principles for ai in society," *Available at SSRN 3831321*, 2019.
- [22] S. L. Pfleeger and B. A. Kitchenham, "Principles of survey research: part 1: turning lemons into lemonade," *ACM SIGSOFT Softw. Eng. Notes*, vol. 26, no. 6, pp. 16–18, 2001.
- [23] J. S. Molléri, K. Petersen, and E. Mendes, "Survey guidelines in software engineering: An annotated review," in *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, pp. 58:1–58:6, ACM, 2016.
- [24] B. A. Kitchenham and S. L. Pfleeger, "Principles of survey research part 4: questionnaire evaluation," *ACM SIGSOFT Softw. Eng. Notes*, vol. 27, no. 3, pp. 20–23, 2002.
- [25] K. Krippendorff, *Content analysis: An introduction to its methodology*. Sage publications, 2018.
- [26] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, "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*, pp. 1–28, 2019.
- [27] V. Vakkuri, K.-K. Kemell, J. Kultanen, and P. Abrahamsson, "The current state of industrial practice in artificial intelligence ethics," *IEEE Software*, 2020.
- [28] M. Krishnan, "Against interpretability: a critical examination of the interpretability problem in machine learning," *Philosophy & Technology*, pp. 1–16, 2020.
- [29] L. Floridi, "Translating principles into practices of digital ethics: Five risks of being unethical," *Philosophy & Technology*, vol. 32, no. 2, pp. 185–193, 2019.