Preprint notes
Title of the article:
ECCOLA - a Method for Implementing Ethically Aligned AI Systems

Authors:
Ville Vakkuri, Kai-Kristian Kemell, Pekka Abrahamsson

Notes:
- This is the author's version of the work
- Paper has been submitted for a peer review in an IEEE Conference
- ECCOLA method’s Internet resources are accessible: https://doi.org/10.6084/m9.figshare.12136308

Copyright notice:
© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
ECCOLA - a Method for Implementing Ethically Aligned AI Systems

Ville Vakkuri[0000-0002-1550-1110], Kai-Kristian Kemell[0000-0002-0225-4560], Pekka Abrahamsson[0000-0002-4360-2226]
Faculty of Information Technology, University of Jyväskylä
Jyväskylä, Finland
ville.vakkuri@jyu.fi, kai-kristian.o.kemell@jyu.fi, pekka.abrahamsson@jyu.fi

Abstract— Various recent Artificial Intelligence (AI) system failures, some of which have made the global headlines, have highlighted issues in these systems. These failures have resulted in calls for more ethical AI systems that better take into account their effects on various stakeholders. However, implementing AI ethics into practice is still an on-going challenge. High-level guidelines for doing so exist, devised by governments and private organizations alike, but lack practicality for developers. To address this issue, in this paper, we present a method for implementing AI ethics. The method, ECCOLA, has been iteratively developed using a cyclical action design research approach. The method makes the high-level AI ethics principles more practical, making it possible for developers to more easily implement them in practice.

Keywords—Artificial Intelligence, AI ethics, Ethics, implementing, method

1. INTRODUCTION

As we make increasing progress on Artificial Intelligence (AI), the systems become increasingly widespread and exert a growing impact on society. This has also resulted in us witnessing various AI system failures, which have served to highlight various ethical issues associated with these systems. Many of these failures have made the global headlines and resulted in public backlash. Especially privacy issues related to facial recognition technology have become a prominent topic among the general public, as well as for policymakers\(^1\).

The systems we develop, despite us having had some collective learning experiences from past system failures, are still far from being problem-free. Ethical issues persist, and more arise as the technologies become more sophisticated. Aside from the obvious physical damage potential of systems such as autonomous vehicles, data handling alone is ripe with ethical issues without universal answers.

The discussion on the field of AI ethics has soared in activity in the past decade following this technological progress, resulting in the birth of some key principles that are now widely acknowledged as central issues in AI ethics. One such issue is the demand for AI systems that are explainable [1]. The problem thus far has been transferring this discussion into practice, i.e., how to actually influence the development of these systems?

For the time being, this has mostly been carried out either via guidelines or laws and regulations. Guidelines have been devised by companies [2], governments [3] and standardization organizations [4]. Yet, these guidelines have been lacking in actionability. Developers struggle to implement abstract ethical guidelines into the development process [5,6].

Methods and practices in the area remain highly technical, focusing on specific issues in e.g. machine learning [7]. While certainly useful in their specific contexts, these types of tools do not help companies in the design and development process as a whole. Thus, development methods are still required to bridge this gap between research and practice in the area.

In this paper, we present our work on an AI ethics method: ECCOLA. It has been developed iteratively over the past two years through empirical use and data resulting from it, with each iteration improving the method. ECCOLA is intended to help organizations implement AI ethics in practice, in an actionable manner.

The rest of this paper is structured as follows. The second section discusses the theoretical background of the paper: AI ethics, methods in AI ethics, as well as the Essence Theory of Software Engineering used in devising the method in question. The third section presents the method, ECCOLA. In the fourth section we discuss how ECCOLA was iteratively developed and what kind of data were used in doing so. In the fifth and final section we discuss the method in relation to extant literature and conclude the paper.

2. THEORETICAL BACKGROUND

This section is split into three subchapters. In the first one, we provide an overview of the current state of AI ethics in research. In the second one, we focus on the state of the practical implementation of AI ethics, discussing the methods

\(^1\) https://www.bbc.com/news/technology-48276660
and other tools that currently exist to help practitioners implement it. In the third and final one, we discuss the Essence Theory of Software Engineering, and specifically the idea of essentializing software engineering practices, as this an approach we have utilized in devising ECCOLA.

A. AI Ethics

AI ethics is a long-standing area of research. In the past, much of the debate focused on hypothetical future scenarios that would result from technological progress. However, as these hypothetical future scenarios start to become reality following said progress, which to many has been faster than anticipated, the field has become increasingly active.

Much of the research in the area has focused on theory, and specifically to define AI ethics by highlighting key ethical issues in AI systems. This discussion has focused on principles. Many have been proposed and discussed, and, by now, some have become largely agreed-upon [8]. Based on an analysis of the numerous AI ethics guidelines that now exist, Jobin et al. [8] listed the key principles that could be considered central based on how often they appear in these guidelines: “transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability, and solidarity.”

To provide an example of the type of research that has been conducted on these principles, we can look at transparency. Transparency [9] is widely considered one of the central AI ethical principles. Transparency is about understanding AI systems, how they work, and how they were developed [9,10]. It has been argued to be the very foundation of AI ethics: if we cannot understand how the systems work, we cannot make them ethical either [11]. The discussion on transparency has, aside from defining what it is, focused on how to achieve it. For example, Ananny & Crawford [10] discussed the limitations of the idea of transparency in relation to the complexity brought on by machine learning. Is being able to see inside the system really enough or even helpful? Transparency is featured as a key principle in the high-profile guidelines of EU [3] and IEEE [4], for example.

Though principles are one way of categorizing the discussion in the area, it is ultimately about bringing attention to potential ethical issues in AI, with or without pinning them under a specific principle. Privacy issues, for example, have been one prominent topic of discussion both in academia and the media following various practical examples of (ethical) AI system failures. Privacy issues have been discussed in relation to data handling, technology such as facial recognition, as well as racial bias, which falls under the principle of fairness.

Indeed, guidelines have, thus far, been the main way of bridging the gap between research and practice in the area. The purpose of these guidelines has been to distill the discussion in the area into a tool. However, past research has shown that guidelines are rarely effective in software engineering. McNama et al. [6] studied the impact the ACM Code of Ethics2 had had on practice in the area, finding little to none. This seems to also be the case in AI ethics: in a recent paper [5], we studied the current state of practice in AI ethics and found that the principles present in literature are not actively tackled out on the field.

This state of affairs underlines a need for more actionable tools for implementing AI ethics in practice. In the context of software engineering, we thus turn to methods, i.e. ways of working that direct how work is carried out [12]. As software engineering in any mature organization is carried out using some method, out-of-the-box ones or in-house ones, incorporating AI ethics as a part of these methods would be a goal to strive for.

B. Methods in AI Ethics

There are already various methods and tools for implementing AI ethics, as highlighted by Morley et al. [7] in their systematic review. These are largely tools for the technical side of AI system development, such as tools for machine learning. On the other hand, we are not currently aware of any method focusing on the higher-level design and development decisions surrounding AI systems. Guidelines have been devised for this purpose but seem to remain impractical given their seeming lack of adoption out on the field [5].

Aside from AI ethics methods and tools, some ethical tools from other fields do exist that could potentially be used to design ethical AI systems. One example of such a tool is the RESOLVED method from the field of business ethics [13]. We have, in a past study [14], studied the suitability of this particular method for the AI ethics context, with our results suggesting that dedicated methods would be more beneficial. Such methods, however, are currently lacking.

Aside from ECCOLA, there is currently some other activity in method development for the area as well, e.g., Leikas et al. [15] recently presented an “ethical framework for designing autonomous intelligent systems”. In devising ECCOLA, our method, we have turned to the Essence Theory of Software Engineering for method engineering. Specifically, we have utilized the theory’s philosophy of essentializing software engineering practices in devising a method. We will discuss this in the following subsection.

C. Essentializing to Create Methods from Practices

A The Essence Theory of Software Engineering (Jacobson et al. [12]) is a method engineering tool. It comprises a method core, which the authors refer to as a kernel, as well as a language. The kernel, they argue [12], contains all the core elements present in any software engineering project.

To this end, the kernel contains three types of items: alphas (i.e. things to work with), activities (things to do), and competencies (skills required to carry out the tasks). There

---

2 https://www.acm.org/code-of-ethics
are seven alphas, which form the core of the kernel: opportunity, stakeholders, requirements, software system, work, team and way-of-working. The kernel provides a basis for constructing methods using the Essence language to describe them. I.e., the theory consists of basic building blocks which can be utilized by using the language to extend the base to build a method. On its own, the kernel could be used as a generic software engineering method, but the point of Essence is to construct new methods using the language, while utilizing the kernel as an extensible starting point for doing so.

Software engineering methods consist of practices. A practice is a more atomic unit of work, such as pair programming. In creating ECCOLA, we have utilized the idea of essentializing software engineering practices. In short, this refers to describing them using the Essence language. This offers one way of breaking down practices into different elements in order to describe them, making them easier to understand. This also serves to make practices more modular, as describing them in the same notational language makes it easier to combine them into methods.

Essentializing practices is described as a process by Jacobson [16] as follows:

- “Identifying the elements – this is primarily identifying a list of elements that make up a practice. The output is essentially a diagram [...]”
- “Drafting the relationships between the elements and the outline of each element – At this point, the cards are created.”
- “… Providing further details – Usually, the cards will be supplemented with additional guidelines, hints and tips, examples, and references to other resources, such as articles and books”

As can be observed in the above quote, Essence utilizes cards to describe methods. This is also an approach we have utilized in ECCOLA. The ECCOLA method is utilized via a physical (or digital) set of cards.

Essence was also chosen due to its method-agnostic approach and modular philosophy on methods. From the get-go, ECCOLA was never intended to be a stand-alone method, but rather, a modular extension to existing software development methods that would bring in AI ethics into the process.

Originally, we planned on using the Essence language to describe ECCOLA. For example, principles such as transparency could have been alphas (i.e. things to work with) in the method. However, as the development of the method progressed and we began to test its early versions in practice, Essence turned out to make the method confusing to its users. This ultimately resulted in Essence taking less of a role in the later iterations of ECCOLA, as we discuss further in the study design section.

3. ECCOLA - A METHOD FOR DESIGNING ETHICALLY ALIGNED AI SYSTEMS

As we have discussed in section 2, AI ethics is currently an area with a prominent gap between research and practice. Much of the research has been theoretical and conceptual, focusing on defining key principles for AI ethics and how to tackle them. The numerous guidelines for AI ethics that currently exist [8] have tried to bridge this gap to bring these principles to the developers but seem to not have had much success. Indeed, ethical guidelines tend to not have much impact in the context of SE [6]. To bridge this gap with another approach, we propose a method for implementing AI ethics: ECCOLA.

ECCOLA4 (figure 1) is intended to provide developers an actionable tool for implementing AI ethics. To utilize the various AI ethics guidelines in practice, the organization seeking to do so has to somehow make them practical first. ECCOLA, on the other hand, is intended to be practical as is, and ready to be incorporated into any existing method. ECCOLA does not provide any direct answers to ethical problems, as arguably correct answers are a rare breed in ethics in general, but rather asks questions in order to make the organization consider the various ethical issues present in AI systems. Though ultimately how these questions are then tackled is up to the organization in question, ECCOLA does encourage taking into account the potential ethical issues it highlights.

ECCOLA is built on AI ethics research. It utilizes both existing theoretical and conceptual research, as well as AI ethics guidelines that have been devised based on existing research as well. In terms of guidelines, the cards are based primarily on the IEEE Ethically Aligned Design guidelines [4] and the EU Trustworthy AI guidelines [4]. As these guidelines have already distilled much of the existing research on the topic under various principles, these principles have been utilized in ECCOLA as well. AI ethics research, then, has been used to further expand the way these principles are covered in ECCOLA.

In practice, ECCOLA takes on a form of a deck of cards. This approach was based on the Essence Theory of Software Engineering [12], which was used to describe the first versions of the method. Methods described using the Essence language are utilized through cards. However, using cards in the context of software engineering methods is not a novel idea, nor one proposed by Essence, e.g., Planning Poker in Agile uses cards and the idea of Kanban is founded around using cards in the form of sticky notes.

There are 21 cards in total in ECCOLA. These cards are split into 8 themes, with each theme consisting of 1 to 6 cards. These themes are AI ethics ones found in various ethical guidelines, such as transparency or data. Each individual card deals with a more atomic aspect of that theme, such as, in the case of data, data privacy and data

---

3 http://semat.org/alpha-definitions-overview/competency-cards

4 https://figshare.com/articles/Internet_resource_for_ECCOLA_-_A_Method_for_Implementing_Ethically_Aligned_AI_Systems/12136308

This is the author’s version of the work
This is the author’s version of the work. ©2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

quality. Aside from the main set of cards, ECCOLA also features an A5-sized game sheet that describes how the method is used.

Each card in ECCOLA is split into three parts: (1) motivation (i.e. why this is important), (2) what to do (to tackle this issue), and (3) a practical example of the topic (to make the issues more tangible). Each card also comes with a note-making space. As the cards are generally utilized as physical cards, the card is split into two with the left half of each card containing the textual contents and the right half containing white space for making notes. This note-making space has been included to make using the cards more convenient in practice.

ECCOLA supports iterative development. During each iteration, the team is to choose which cards, or themes, are relevant for that particular iteration. ECCOLA is also method-agnostic, making it possible to utilize it with any existing or in-house SE method.

In the next section, we discuss how ECCOLA has been developed. The method has gone through multiple iterations and has been improved based on empirical data in each iteration.

4. ECCOLA DEVELOPMENT PHASES AND DATA

ECCOLA has been developed iteratively through multiple phases. For this purpose, we have utilized the Cyclical Action Research method described by Susman and Evered [17] in developing it. Thus far, ECCOLA has gone through 6 iterations. In each phase, we have collected empirical data, based on which the method has then been iteratively improved.

The six subsections of this section each cover one iteration. In each subsection, we discuss what ECCOLA looked like at the time, how it was tested, and how it was changed based on the data. This process is also summarized in the table below (Table 1). The summary of the changes made to ECCOLA in each iteration can be found as a list at the end of each sub-section.

![Image](https://example.com/image.png)

**Figure 1 Cyclical Action Research process on ECCOLA. Including Cycle of Action, Observation, Reflection on each iteration.**

| Phase | Version | Background Theories | Study Setting | Timing | Study Participants |
|-------|---------|---------------------|---------------|--------|--------------------|
| 1     | N/A     | RESOLVEDD, EAD, Essence | Class          | Q1-Q2 2018 | 5 teams of 4-5 students |
| 2     | 1       | RESOLVEDD, EAD, Essence | Class          | Q2 2018 - Q2 2019 | 27 teams of 3-5 students |
| 3     | 2       | RESOLVEDD, EAD, Essence | Class          | Q2 2018 - Q2 2019 | 27 teams of 3-5 students |
| 4     | 3       | RESOLVEDD, EAD, Essence | Class          | Q2 2018 - Q2 2019 | 27 teams of 3-5 students |
| 5     | 4       | EU AI HLEG, EAD      | Blockchain Project | Q2-Q3 2019 | 2 sw development team members |
| 6     | 5       | EU AI HLEG, EAD      | Conference Workshop | Q4 2019 | 8 researchers |

**Table 1 Cyclical Action Research Phases**

A. Phase 1 (Q1-Q2 2018)

In early 2018, prior to starting our work on ECCOLA, we searched for existing methods for AI ethics, ultimately finding none. Thus, we expanded our horizons and looked at ethical tools from other fields instead, to see if anything would seem applicable in the context of AI ethics as well. This led us to eventually test an existing ethical tool from the field of business ethics, the RESOLVEDD strategy [13], in the context of AI ethics. Our aim was to see if existing ethical tools, even if they were not specifically created for AI ethics, could be suitable for that context.

We conducted a scientific study on RESOLVEDD in the context of AI ethics. These findings have been published in-depth elsewhere (see Vakkuri & Kemell [14]). In short, we discovered that forcing developers to utilize RESOLVEDD did have some positive effects. Namely, it produced transparency in the development process, and the presence of an ethical tool made the developers aware of the potential importance of ethics, resulting in ethics-related discussions within the teams. However, the tool itself was not considered well-suited for the context by the respondents. Moreover, when forcing developers to utilize such a tool, the commitment towards it quickly vanished when the tool was no longer compulsory.

**Phase 1 actions:**

- The development of ECCOLA was initiated
Figure 2 ECCOLA - a Method for Implementing Ethically Aligned AI Systems
This is the author's version of the work. ©2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

B. Phase 2 (Q2 2018 - Q2 2019)

I. Creating Version 1 (Q2 2018 - Q1 2019)

Based on the results of this study, we began to develop a method of our own, ECCOLA, during the latter half of 2018. This initial version of the method was based on three primary theories: (1) RESOLVEDD strategy, (2) The Essence Theory of Software Engineering, and (3) The IEEE Ethically Aligned Design guidelines.

We utilized some of the general ideas of RESOLVEDD, which were deemed useful based on the data we collected. Namely, we took to RESOLVEDD for ideas on how to make the tool support iterative development. Additionally, we included some of the aspects of RESOLVEDD which were shown to support transparency of system development (e.g. the idea of producing formal text documents while using the method).

We began to describe the method using the Essence language (see section 2.3). Methods described using Essence are visualized through cards, and thus, ECCOLA took on the form of a card deck as well. This also meant that we included the various elements of Essence into the cards. For example, we made some of the key AI ethics principles, namely transparency, accountability, and responsibility, into alphas in the context of Essence (i.e., measurable things to work on). The cards also included various activities that were to be performed in order to progress on these alphas, as well as patterns and other Essence elements.

The AI ethics contents of the method, at this stage, were based primarily on the IEEE Ethically Aligned Design guidelines [4]. We included key principles from the guidelines such as transparency and accountability, which have been prominent topics of discussion in AI ethics. Additionally, we utilized various research articles. For example, to expand on transparency, we utilized the studies of Dignum [9] and Ananny & Crawford [10], among others.

Much like how while using RESOLVEDD one produces text answering some questions posed by the tool, we incorporated the same idea of producing text while using ECCOLA into the initial version of the method. The theoretical background of this early version was based primarily on the IEEE EAD guidelines and the idea of the ART principles of AI Ethics [9].

II. Testing Versions 1 (Q1 2019)

This first version of ECCOLA was tested in a large-scale project-based course on systems development at the University of Jyväskylä in the first quarter of 2019. In the course, 27 student teams of 4-5 students worked on a real-world case related to autonomous maritime traffic. Each team was tasked with coming up with an innovation that would help make autonomous maritime traffic possible. The teams were not required to actually develop these innovations into functional products, given the time and capability constraints in a course setting, but rather, to hone the ideas as far as they could in the context of the course. Some teams ultimately did produce technical demos, but this was not required. The results of these projects have been published in an educational book.

As any such innovation would involve AI directly or indirectly, given the autonomous maritime traffic context, we chose to test ECCOLA by having these teams utilize it to reflect on the ethical issues their ideas might pose. The teams were introduced to ECCOLA during a course lecture and were handed a physical card deck. Each team was then told to utilize the card deck in whatever way they saw fit, while writing down notes on the cards as if they used them. Additionally, unstructured interview data was collected from the teams through their weekly meetings with their assigned mentor and this feedback was taken into account in developing the method.

Prior to the course, the students had been tasked with reading a book on Essence, Software Engineering Essentialized [16], which explains the tool. Though the educational goal of this was elsewhere, this also served to make sure the students would not be overly confused with this version of ECCOLA being described using the Essence language.

After the students had utilized the cards for a week, they were collected and the written notes on them analyzed. Based on this data, and the discussions the teams had had with their mentors in the weekly meetings, ECCOLA was improved as follows. This first iteration of ECCOLA.

Actions based on iteration 1 of phase 2, for version 2:

- Alpha states were added to the alphas in order to make tracking progress on them easier
- Practical examples were added to the cards to make it easier to understand the practical implications of the ethical issues in the cards
- Reduced the amount of academic jargon on the cards, focusing on practice over theory
- Removed list of academic references from each card.

III. Testing Version 2, (Q1 2019)

This iteration took place during the same systems development course described in the preceding subsection. This iteration was carried out in the same manner as the previous one. The same student teams were tasked with utilizing the new version of ECCOLA again while writing down notes on them as they did. Additional data was again collected in the weekly mentor meetings. Overall, this was, in terms of time elapsed, a brief iteration carried out during the course.

After another week, ECCOLA was once more improved based on the data collected. The following changes were made to the method.

5 https://jyx.jyu.fi/handle/123456789/63051

This is the author's version of the work
Actions based on iteration 2 of phase 2, for version 3:

- Added a game sheet describing how the cards (and the method) should be used. We realized that the method, in this version, required teaching to be understood
- Added numbering to the cards
- Further reduced the amount of academic jargon on the cards.

IV. Testing Version 3 (Q1 2019)

As was the case with the previous two iterations in this phase, the third version of ECCOLA was tested in the systems development course in a similar manner. However, as this was towards the end of the course, there were no further iterations to be tested in the same setting. Thus, we took our time to analyze the feedback from all three versions, reflect on it, and study new publications in the area to improve the method.

This resulted in a lengthier creation process for the subsequent version. Based on the data and our reflection we made larger changes to the method. We discuss these in the following subsection.

V. Creating Version 4 (Q2 2019)

Data from phase 2 indicated that the method, though cumbersome to use, did help the teams implement AI ethics. Analyzing the notes, they had made on the cards showed that they had conducted ethical analyses successfully and changed their ideas based on their analyses. The AI ethics portion of the method thus worked. However, the method was not easy to use.

After the course had concluded, we had time to make larger improvements to the method based on the data collected. We opted to lessen the role of Essence in the method, forgoing the idea of using the Essence language to describe it. It seemed that Essence had made ECCOLA more confusing than it otherwise would have been, as in addition to learning the method, its users would have to learn the Essence notation and Essence in general. We stopped using the Essence elements in the cards and instead split the cards into different AI ethics themes. However, the general approach of using cards for the method seemed to work and thus this approach was kept.

Additionally, based on the data, the method seemed to be too heavy to use. ECCOLA was initially designed to be a linear process that was iteratively repeated. Its users, however, would be free to modify the process based on their development context and based on their use experience. Nonetheless, this approach was considered too rigid, and the respondents felt it was just another process tacked onto their other work processes. We thus changed the approach, making the cards more stand-alone, so that the users of the method could choose which cards to utilize based on which ones they felt were relevant for their current situation.

During this time period, before the next empirical test, we also expanded the theoretical basis of the method. The initial version of the EU Guidelines for Trustworthy AI was published in early 2019, some aspects of which we chose to incorporate into ECCOLA. Other novel literature was also included to expand on theoretical basis of the method.

Actions based on phase 2:

- The use of Essence to describe the method was discontinued
- Contents of the cards reformatted and reformulated
- Method made modular rather than one linear, iterative process
- Expanded the AI ethics theoretical basis of the method.

C. Phase 3 (Q2-Q3 2019)

As the primary concern with the versions 1 to 3 had been the way ECCOLA was used as a method in practice rather than its AI ethical contents, we chose to focus on making a method that it would be easier and more practical to use. For this purpose, we made a spin-off of ECCOLA for the context of blockchain ethics. Many of the AI ethical themes such as transparency and data issues could be translated into this context, even if the contents of the cards had to be modified to be better suited for it. Additional blockchain specific issues were also added into these cards.

In this phase, ECCOLA was utilized in a real-world blockchain project by two of the project team members. Data was collected through observation and various unstructured interviews. The team was free to utilize the cards as they wished and was encouraged to reflect on how the method would best suit their SE development method of choice. However, the team could also receive consultation from one of the researchers where needed on how to use the cards, as well for clarification on their contents, if needed. As a result, we gained a better understanding of how the method was utilized in practice (e.g., how many cards were used per iteration on average, which was 6) in a real-world SE context.

Based on the data gathered from the blockchain project, the main ECCOLA card deck was iteratively improved. The lessons learned from studying the use of the blockchain ethics version of ECCOLA were incorporated into ECCOLA.

The following changes were made:

- A note-making space was added to each card
- Added new cards
- Split the cards into themes, such as transparency or data.

Added more contextual content into each card, as opposed to focusing largely on instructions on what to do. This resulted in revamping the “motivation” and “practical example” section of many of the cards.
In this paper, we have presented a method for implementing AI ethics: ECCOLA. It is intended to help organizations develop more ethical AI systems by providing them with means of implementing AI ethics in a practical manner. ECCOLA has been developed iteratively using the Cyclical Action Research approach [17]. Though development on the method continues, we have reached a state of maturity where we want to share the method with the scientific community.

The purpose of ECCOLA is to help us bridge the gap between research and practice in the area of AI ethics. Despite the increasing activity in the area, the academic discussion on AI ethics has not reached the industry [5]. Through ECCOLA, we have attempted to make some of the contents of the IEEE EAD guidelines [4] and the EU Trustworthy AI guidelines [3] actionable, alongside other research in the area.

In developing ECCOLA, we have had three main goals for the method:

- To help create awareness of AI ethics and its importance
- To make a modular method suitable for a wide variety of SE contexts, and
- To make ECCOLA suitable for agile development, and to also make ethics a part of agile development in general.

In relation to the first goal, there is currently no way of benchmarking what is, so to say, sufficiently ethical in the context of AI ethics. This is arguably a limitation for any such method in the context currently. Benchmarking ethics is difficult and thus it is equally difficult for a method to have a proven effect in a quantitative manner. Moreover, ethical issues are often context-specific and require situational reflection. This has been why we have instead chosen to focus on raising awareness and highlighting issues rather than trying to provide direct answers for them. Raising awareness has also been a goal of the IEEE EAD initiative [4]. Raising awareness is important as the area of AI ethics is new for the industry.

ECCOLA provides a starting point for implementing ethics in AI. Based on our lessons learned thus far, we argue that ECCOLA facilitates the implementation of AI ethics in two confirmable ways. First, ECCOLA raises awareness of AI ethics. It makes its users aware of various ethical issues and facilitates ethical discussion within the team. Secondly, ECCOLA produces transparency of systems development. In utilizing the method, a project team produces documentation of their ethical decision-making by means of e.g. making notes on the note-making space in the cards and non-functional requirements in product backlog. Transparency is one key issue in AI systems, both in terms of systems and in terms of systems development [9]. These documents, as we have done while testing the method, can also be analyzed to understand how the method was used.

---

6 https://icsob2019.wordpress.com/workshops/
This is the author’s version of the work. ©2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

aside from seeking to understand the reasoning behind the ethical decisions made during development.

The second goal has been based on the method-agnostic philosophy of the Essence Theory of Software Engineering [12]. Industry organizations use a wide variety of methods, from out-of-the-box ones to, more commonly, tailored in-house ones[18]. ECCOLA is not intended to replace any of these. Rather, ECCOLA is intended as a modular tool that can be added to existing methods and used in conjunction with them.

This, in turn, leads us to the third goal. As agile development is currently the trend, ECCOLA has been designed to be an iterative process from the get-go. However, during its iterative development, we noticed that a strict process was not a suitable approach due to being too heavy to use. The users of the method opted out of adhering to the process and used the cards in a modular fashion despite the instructions. Now, ECCOLA is a modular tool by design. Being a card deck, this means that its users are able to select the cards they feel are relevant for each of their iterations, as opposed to having to go through the same process every time. Moreover, ECCOLA is intended to become a part of the agile development process in general. Ethics should not be merely an afterthought, but rather, a requirement, as well as a part of the user stories.

ECCOLA is a tool for developers and product owners. Ethics cannot be outsourced, nor can ethics be implemented by hiring an ethics expert [5]. AI ethics should be in the requirements, formulated in a manner also understood by the developers working on the system.

As governments and policymakers have already begun to regulate AI systems in various ways (e.g. bans on facial recognition for surveillance purposes7), this trend is likely to only accelerate. With more and more regulations imposed on AI systems, organizations will need to tackle various AI ethics issues while developing their systems. This will consequently result in an increasing demand for methods in the area. While this will also inevitably result in the birth of various new methods, developed by companies, scholars, and standardization organizations alike in the future, for the time being ECCOLA can serve as one initial option where there currently are none.

REFERENCES

[1] Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1, 206–215 (2019)
[2] Pichai, S. AI at Google: our principles. Blog (2018). https://www.blog.google/technology/ai/ai-principles/ 2019/5/13.
[3] AI HLEG (High-Level Expert Group on Artificial Intelligence): Ethics guidelines for trustworthy Al. (2019) https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai, last accessed 2019/6/14.
[4] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition. IEEE. (2019). https://standards.ieee.org/content/ieee-standards/en/industry-connections/ethics-standards.html,
[5] Vakkuri, V., Kemell, K. K., Kultanen, J., & Abrahamsson, P. The Current State of Industrial Practice in Artificial Intelligence Ethics. (2020). IEEE Software.
[6] McNamara, A., Smith, J., Murphy-Hill, E.: Does ACM’s code of ethics change ethical decision making in software development? Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering ESEC/FSE pp. 729-733, 2018.
[7] Morley, J., Floridi, L., Kinsey, L., & Ehitlal, A.: From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. Preprint arXiv:1905.06876 (2019)
[8] Jobin, A, Marcello I, and Effy V. "The global landscape of AI ethics guidelines." Nature Machine Intelligence 1.9 (2019): 389-399.Nature vanha
[9] Dignum, V.: Responsible Autonomy. (2017). Preprint arXiv:1706.02513.
[10] Ananny, M., Crawford, K.: Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. In: New Media & Society vol. 20(3), pp. 973–89. (2018).
[11] Turilli, M., Floridi, L.: The ethics of information transparency. Ethics and Information Technology, vol. 11(2), 105-112. (2009).
[12] Jacobson, I, et al. "The essence of software engineering: the SEMAT kernel." Communications of the ACM 55.12 (2012): 42-49.
[13] Pfeiffer, R.S., Forsberg, R.P.: Ethics on the Job: Cases and Strategies. Wadsworth Publishing Company, California (1993)
[14] Vakkuri, V., and Kemell KK, “Implementing AI Ethics in Practice: An Empirical Evaluation of the RESOLVEDD Strategy.” International Conference on Software Business. Springer, Cham, 2019.
[15] Leikas, J., Koivistio, J. and Gotcheva N. “Ethical framework for designing autonomous intelligent systems.” Journal of Open Innovation: Technology, Market, and Complexity 5:1 (2019): 18.
[16] Jacobson, I., et al. “The Essentials of Modern Software Engineering.” ACM, New York (2019).
[17] Susman, G., and Evered. R., "An assessment of the scientific merits of action research." Administrative science quarterly (1978): 582-603.
[18] Ghanbari, H., “Investigating the causal mechanisms underlying the customization of software development methods.” Jyväskylä studies in computing 258 (2017).

7 https://www.bbc.com/news/technology-51148501

This is the author’s version of the work