JYX
JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Vakkuri, Ville; Kemell, Kai-Kristian; Jantunen, Marianna; Halme, Erika; Abrahamsson, Pekka

Title: ECCOLA : a method for implementing ethically aligned AI systems

Year: 2021

Version: Published version

Copyright: © 2021 The Author(s). Published by Elsevier Inc.

Rights: CC BY 4.0

Rights url: https://creativecommons.org/licenses/by/4.0/

Please cite the original version:

Vakkuri, V., Kemell, K.-K., Jantunen, M., Halme, E., & Abrahamsson, P. (2021). ECCOLA : a method for implementing ethically aligned AI systems. Journal of Systems and Software, 182, Article 111067. https://doi.org/10.1016/j.jss.2021.111067

The Journal of Systems & Software 182 (2021) 111067

Contents lists available at ScienceDirect

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

ECCOLA — A method for implementing ethically aligned AI systems ☆

Ville Vakkuri *, Kai-Kristian Kemell, Marianna Jantunen, Erika Halme, Pekka Abrahamsson

University of Jyväskylä, PO Box 35, FI 40014, Jyväskylä, Finland

ARTICLE INFO

Article history:
Received 10 January 2021
Received in revised form 4 May 2021
Accepted 17 August 2021
Available online 2 September 2021

Keywords:
Artificial intelligence
AI ethics
Ethics
Implementing
Method

ABSTRACT

Artificial Intelligence (AI) systems are becoming increasingly widespread and exert a growing influence on society at large. The growing impact of these systems has also highlighted potential issues that may arise from their utilization, such as data privacy issues, resulting in calls for ethical AI systems. Yet, *how* to develop ethical AI systems remains an important question in the area. How should the principles and values be converted into requirements for these systems, and what should developers and the organizations developing these systems *do*? To further bridge this gap in the area, in this paper, we present a method for implementing AI ethics: ECCOLA. Following a cyclical action research approach, ECCOLA has been iteratively developed over the course of multiple years, in collaboration with both researchers and practitioners.

© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

As Artificial Intelligence (AI) technology is developed with speeding progress, these systems become increasingly widespread and exert a growing impact on society. This has led to us witnessing a number of AI system failures, many of which have made global headlines and resulted in public backlash. Occasionally, these failures have served to highlight some of the various potential ethical issues associated with AI systems, in cases where these systems are found to, for example, exercise unfair bias or act in socially unacceptable ways. Some such famous incidents occurred when AI-based systems have endorsed or exercised unethical behavior such as gender discrimination[1] or racism.[2] Especially issues related to privacy, in cases like facial recognition technology, have become a prominent topic among the general public, as well as for policymakers.[3]

Though these incidents have resulted in collective learning experiences, the systems we developed are still far from being problem-free. Ethical issues persist, and more arise as the level of sophistication of AI-related technologies rises. Aside from the obvious physical damage potential of systems such as autonomous vehicles, many areas of AI systems and their development are ripe with ethical issues without universal answers, starting from well-known topics such as data handling and extending to complex societal impacts of future systems (advanced general AI, etc.) currently still unattainable without further progress in the area.

The discussion on the field of AI ethics has soared in activity in the past decade following AI-related technological progress, resulting in the birth of some key principles that are now widely acknowledged as central issues in AI ethics. These principles cover a wide range of subjects, such as a demand for AI systems to be explainable (Rudin, 2019) and aligned with human rights and well-being (IEEE Global Initiative, 2019). The problem thus far has been transferring this discussion into practice, i.e., how to actually influence the development of these systems.

So far, this has mostly been carried out either via guidelines or laws and regulations. Guidelines have been devised by various parties, such as companies (e.g., Google (Pichai, 2018)), governments (e.g., EU (HLEG, 2019)) and standardization organizations (e.g., IEEE (IEEE Global Initiative, 2019)). Despite their ubiquity, guidelines alone have been lacking in actionability. Developers struggle to implement abstract ethical guidelines into the development process (Vakkuri et al., 2020; McNamara et al., 2018). There may be no consequences for deviating from codes of ethics or using them mainly as a marketing strategy, and there is no guarantee that ethics guidelines will affect the actual decision-making of developers (Hagendorff, 2020).

Methods and practices in the area remain highly technical, focusing on, e.g., specific machine learning issues (Morley et al.,

☆ Editor: Raffaela Mirandola.

* Corresponding author.
E-mail addresses: ville.vakkuri@jyu.fi (V. Vakkuri), kai-kristian.o.kemell@jyu.fi (K.-K. Kemell), marianna.s.p.jantunen@jyu.fi (M. Jantunen), erika.a.halme@jyu.fi (E. Halme), pekka.abrahamsson@jyu.fi (P. Abrahamsson).

1 https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html.

2 https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

3 https://www.bbc.com/news/technology-48276660.

https://doi.org/10.1016/j.jss.2021.111067
0164-1212/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

2019). While certainly useful in their specific contexts, these types of tools do not help companies in the design and development process as a whole. For example, tools for machine learning, though key in AI systems, do not help companies make decisions regarding the system and its future usage context in the big picture. Thus, other approaches such as development methods for ethical AI 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.ECCOLA is a sprint-by-sprint process designed to facilitate ethical thinking in AI and autonomous systems development, and designed to be used together with existing methods. It takes on the form of a deck of 21 cards, split into 8 AI ethics themes (e.g. transparency). While designing ECCOLA, we had three goals for it: (1) to help create awareness of AI ethics and its importance, (2) to make a modular method suitable for a wide variety of SE contexts, and (3) to make ECCOLA suitable for agile development, while also helping make ethics a part of agile development in general. Overall, ECCOLA is intended to help organizations implement AI ethics in practice, in an actionable manner.

ECCOLA has been developed iteratively over the past three years through empirical use and data resulting from it, with each iteration improving the method. In doing so, we have followed a Cyclical Action Research approach (based on Susman and Evered (1978) and Davison et al. (2004)). So far, there have been 6 stages in this process. ECCOLA has been used and evaluated in student, industry, and academic contexts (e.g. conference workshops), with the evaluation and usage shifting towards the industry over time. This article extends an existing paper presenting an earlier version of ECCOLA published in the proceedings of DSD/SEAA 2020 (Vakkuri et al., 2020). Since then, we have focused on seeing how companies utilize ECCOLA in practice while continuing to develop ECCOLA in collaboration with other researchers.

The rest of this paper is structured as follows. The second section discusses the theoretical background of ECCOLA. The third section presents the ECCOLA method itself. In the fourth section we introduce our research approach. In the fifth section we discuss how ECCOLA was iteratively developed. In the sixth section we discuss the implications of ECCOLA. In the seventh section we discuss threats to validity. The eighth and final conclusions section concludes the paper.

2. Theoretical background

This section is split into four subsections. 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 and other tools that currently exist to help practitioners implement it. In the third we discuss Value Sensitive Design to further position this method using existing literature. In the fourth and final one, we discuss the Essence Theory of Software Engineering, and specifically the idea of essentializing software engineering practices, as this is an approach we have utilized in devising ECCOLA.

2.1. AI ethics

AI ethics is a long-standing area of research. In the past, much of the debate has 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 on defining 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 (Jobin et al., 2019). Based on an analysis of the numerous AI ethics guidelines that now exist, Jobin et al. (Morley et al., 2019) 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 (Dignum, 2017) is widely considered one of the central AI ethical principles. Transparency is about understanding AI systems, how they work, and how they were developed (Dignum, 2017; Ananny and Crawford, 2018). 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 (Turilli and Floridi, 2009). The discussion on transparency has, aside from defining what it is, focused on how to achieve it. For example, Ananny and Crawford (2018) 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? For example, transparency is featured as a key principle in the high-profile guidelines of EU (HLEG, 2019) and IEEE (IEEE Global Initiative, 2019).

Principles are but one way of categorizing the discussion in the area. The discussion in the area 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. For example, privacy issues have been discussed in relation to data handling, and technologies such as facial recognition. Privacy issues are hardly a topic of discussion unique to the field of AI ethics either. Data issues such as bad data have also been discussed in relation to racial bias, which falls under the principle of fairness.

Guidelines have been utilized as a way of bridging the gap between research and practice, with the purpose to distill the discussion in the area into tools in the form of guidelines. However, past research has shown that guidelines are rarely effective in software engineering. McNamara et al. (2018) studied the impact the ACM Code of Ethics[4] 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 (Vakkuri et al., 2020), 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. Moreover, we found that AI development endeavors did not differ from generic development endeavors in this regard, with companies developing AI no more focused on tackling them differently than any other software company. This gap, and the issues with guidelines, are also acknowledged by Johnson & Smith in their gap analysis (Johnson and Smith, 2021).

The state of affairs as presented here, underlines a need for more actionable tools for implementing AI ethics in practice. In the context of software engineering, we therefore turn to methods; ways of taking action that direct how work is carried out (Jacobson et al., 2012). 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. In this next subsection, we look at methods in the area.

[4] https://www.acm.org/code-of-ethics.

2.2. Methods in AI ethics

There are already various methods and tools for implementing AI ethics, as highlighted by Morley et al. (2019) in their systematic review of the field. The study consists of largely tools for the technical side of AI system development, such as tools for machine learning. The study by Morley et al. reviews a collection of tools or methods that are utilized by various companies and organizations for implementing ethics in AI development, and a typology based on ethical principles is used to analyze the results.

The review by Morley et al. brought certain challenges to light regarding AI ethics tools; the study showed that some of the researched tools are immature, and there is an "uneven distribution of effort across the 'Applied AI Ethics' typology" (Morley et al., 2019). Morley et al. believe that creating ethical machine learning technologies is realistically possible, but efforts have so far been focused on the "what", and not the "how" of AI ethics (Morley et al., 2019). The debate has been focusing on the topic on ethical principles, instead of applying them in practice. They suggest that turning ethical principles into design protocols will require increased coordination, and patience to tolerate a slow progression of turning theory into practice, with mistakes along the way (Morley et al., 2019).

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 (Vakkuri et al., 2020). The field remains active, for example, Leikas et al. (2019) recently proposed an "Ethical Framework for Designing Autonomous Intelligent Systems" and an AI ethics MOOC at the Helsinki University has devoted a chapter to AI ethics in practice (Rusanen et al., 2021).

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 RESOLVEDD method from the field of business ethics. We have studied the suitability of this particular method for the AI ethics context in the past, with our results suggesting that dedicated methods specifically devised for implementing AI ethics would be more beneficial (Vakkuri and Kemell, 2019). Additionally, we feel that Value Sensitive Design (VSD) is another approach worth mentioning in this context, even though it is not specific to AI ethics. Due to its prominence in existing research (specifically in Information Systems (IS)), we discuss it separately in the following subsection.

2.3. Values in value sensitive design

In addition to looking at the field of AI ethics from the point of view of SE, we feel that a brief look at ethics and value consideration discussion from IS is in order as well to better position ECCOLA. In particular, Value Sensitive Design (VSD) is a prominent approach that has been utilized out on the field. However, as VSD is not specific to AI ethics, we have separated it from the preceding subsection.

VSD can be traced back to the 1990s when the HCI (Human–Computer Interaction) community took a stand on value-oriented design in IS research (Shilton, 2018). The context-specific nature of ethical issues has been acknowledged in VSD as well, with Friedman remarking that different individuals and people have different ideas of ethics and values (Friedman et al., 2013). In the context of Information Systems Design (ISD), Friedman et al. (2008) proposed 13 values: Human Welfare, Ownership and Property, Privacy, Freedom from Bias, Universal Usability, Trust, Autonomy, Informed Consent, Accountability, Courtesy, Identity, Calmness, and Environmental Sustainability. Looking at this list of values, there is a reasonable amount of overlap with the common AI ethics principles summarized by Jobin et al. (2019) that we discussed in Section 2.1 above.

Even outside the context of AI ethics, integrating ethical considerations into practice in software engineering (SE) is a recurring challenge. For example, the ACM/IEEE Software Engineering Code of Ethics and Professional Practice, while in many ways useful according to Biffl et al. (2006), has also been difficult to integrate into traditional SE. Indeed, a more recent study (McNamara et al., 2018) has also argued that the ACM Ethical Guidelines (Gotterbarn et al., 2018) have not changed the way developers work.

Value Sensitive Design (VSD) is a methodology meant to encourage designers to consider ethics and values in the design process, and is "primarily concerned with values that center on human well-being, human dignity, justice, welfare, and human rights". VSD Lab (2021). VSD is at the cross-section of four fields closely related to HCI, namely Computer Ethics, Social Informatics, Participatory Design, and Computer-Supported Cooperative Work. Friedman and Kahn set up a seven principle composite that the VSD is based on, and one of the main principles is that VSD is a proactive methodology (Friedman et al., 2002). VSD encompasses 14 methods for incorporating value consideration into the design process (Davis and Nathan, 2015).

VSD has seen some success out on the field as well, with multinationals such as Intel and Microsoft utilizing it in some projects (Manders-Huits, 2011). Overall, its use has been documented in a wide variety of projects. Perhaps the most notable VSD method in terms of industry utilization has been the Tripartite Method, which is used to involve value consideration into the design process (Winkler and Spiekermann, 2018). Envisioning Cards[5] can be utilized in deploying the method. Physical tools are commonly used to deploy methods in practice, be it cards or other approaches. We have also chosen to focus on a physical presentation for ECCOLA by making it a card deck.

VSD has, however, also been argued to have its shortcomings. In particular, it has been criticized for lacking in pragmatism and methodological guidance (van der Duin, 2019; Winkler and Spiekermann, 2018). Nonetheless, it has seen some success out on the field, which has been a recurring challenge for any method or tool involving ethics. We have also looked at VSD for some inspiration while designing ECCOLA, as we discuss further in the discussion section.

2.4. Essentializing to create methods from practices

In this final subsection of this section, we discuss a background theory that was utilized especially early on in the development of ECCOLA. The Essence Theory of Software Engineering (Jacobson et al. (2012)) is a method engineering tool. It comprises of two parts: (1) what its authors refer to as a kernel, and (2) a language. In short, the kernel offers premade building blocks for constructing methods using the language, and the language itself is used to model practices and methods.

More specifically, the kernel contains, as its authors argue (Jacobson et al., 2012), all the essential elements found in any SE project. The theory posits that every SE project, at bare minimum, has these elements in it, in addition to any additional project-specific elements. These elements are split into three types of items: alphas (i.e., things to work with), activities (i.e., things to do), and competencies (i.e., the skills required to carry out the project). Moreover, these elements are split into three areas of concern (i.e., categories): customer, solution, and endeavor.

The heart of the kernel consists of the aforementioned alphas, of which there are seven. In the customer area of concern, there

[5] https://www.envisioningcards.com/.

are two alphas: (1) opportunity, and (2) stakeholders. There are also two alphas in the solution area: (3) requirements, and (4) software system. Finally, the endeavor area of concern contains the three final alphas: (5) work, (6) team and (7) way-of-working. Aside from helping the users of the tool structure methods, alphas are used to track progress on a project. Each alphas has alpha *states* that denote progress on that part of the project (e.g. requirements).

Originally, we intended to use the Essence language to describe the ECCOLA method. Essence was 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. Our plan was to devise alphas for AI ethics and to use the language to portray practices used to progress on them.

However, as we discuss in detail the following sections, we ultimately ended up giving up on the idea of using Essence to describe ECCOLA. Briefly put, utilizing Essence to describe ECCOLA made the method too heavy. Not only would the users of ECCOLA have to learn to use ECCOLA itself, they would also have to learn to use, or at least understand, Essence.

On the other hand, though ECCOLA is no longer described using the Essence language, we utilized the idea of *essentializing* practices in ECCOLA. Essentializing practices is described as a process by Jacobson (Jacobson et al., 2019) 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 the above quote highlights, 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. The cards are also created in a similar manner, although with some extra steps as ECCOLA cards have more (and different) content than traditional Essence practice cards. Although Essence is no longer used to describe the method itself, we still utilize the idea of essentializing practices to draft the cards for ECCOLA.

3. ECCOLA - A method for Implementing 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 (Morley et al., 2019) 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 (McNamara et al., 2018). To bridge this gap with another approach, we propose a method for implementing AI ethics: ECCOLA.

ECCOLA (Fig. 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 how these questions are ultimately tackled is up to the users of ECCOLA, ECCOLA does encourage them to take into account the potential ethical issues it highlights.

Table 1
ECCOLA card themes.

Card themes (8)	Card number	Card amount (total 21)
Analyze	#0	1
Transparency	#1–6	6
Safety & Security	#7–9	3
Fairness	#10–11	2
Data	#12–13	2
Agency & Oversight	#14–15	2
Wellbeing	#16–17	2
Accountability	#18–20	3

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

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

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 (IEEE Global Initiative, 2019) and the EU Trustworthy AI guidelines (HLEG, 2019). 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. Existing AI ethics research has then been utilized to expand the way these principles are covered in ECCOLA.

In practice, ECCOLA takes on the form of a deck of cards. This approach was based on the Essence Theory of Software Engineering (Jacobson et al., 2012), 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 originally proposed by Essence. E.g., Planning Poker in Agile uses cards. Moreover, various SE methods encourage the use of physical tools in general while using the method. The idea of Kanban, for example, is founded around using sticky notes on a signboard.

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 data privacy and data quality in the case of data. Aside from the main set of cards, ECCOLA also features an A5-sized game sheet that describes how the method is used (see Table 1).

Each card (see Fig. 2) 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.

Fig. 1. ECCOLA - a method for implementing ethically aligned AI systems.

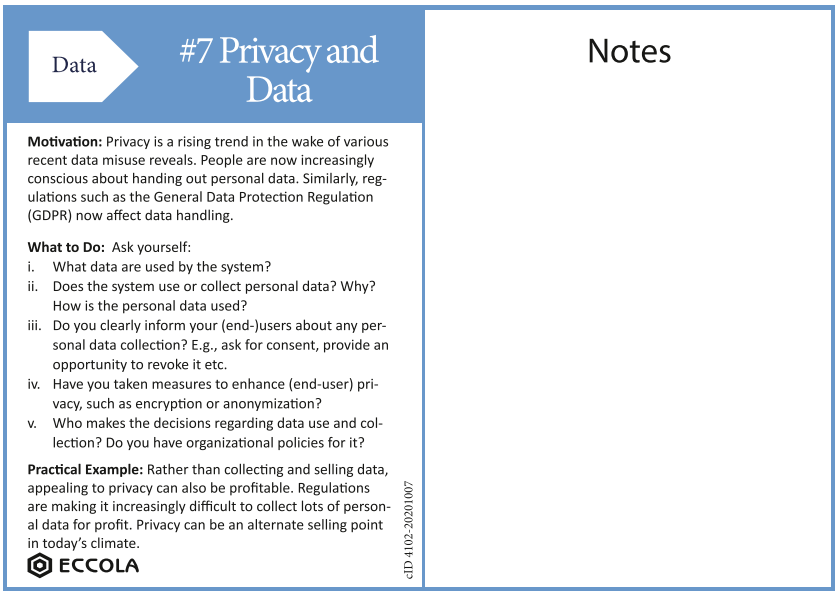

Fig. 2. Card example from ECCOLA, Card #7 privacy and data.

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 following subsection, we discuss how to use ECCOLA in practice.

3.1. How to use ECCOLA in practice?

Expanding on what we already discussed in this section, i.e. what ECCOLA is, this section describes how to implement the ECCOLA method in practice. It includes descriptions of how ECCOLA has been used for different purposes, and our recommendations on how to proceed with using the ECCOLA cards in software development projects.

ECCOLA is a modular, sprint-by-sprint process that has been designed to facilitate ethical thinking in AI/S (Artificial Intelligence/Autonomous System) development. While using ECCOLA, you choose the cards you feel are relevant for your work currently and then evaluate the situation again after each sprint. Using ECCOLA results in a paper trail of choices and trade-offs that documents the ethical consideration conducted during development. This documentation provides a way of evaluating the trustworthiness of the system.

ECCOLA is intended to be used during the entire design and development process in a three step process that is repeated in every iteration. (1) Prepare: Choose the relevant cards for the current sprint. (2) Review: Keep the selected cards on hand during work tasks. Write down on the cards the actions you have taken and (ethical) discussions you have had. (3) Evaluate: Review to ensure that all the planned actions were taken. Revise the card deck as needed, and repeat the process. Remember to do a retrospective afterwards.

Everyone involved with using the cards should read the cards thoroughly at least once before the sorting process in order to familiarize themselves with the topics of the cards as well as their contents. This is recommended not only to make the decision process easier, but also to save time when selecting cards for each sprint.

ECCOLA cards are designed to offer a variety of viewpoints to prompt thoughts during the development process, and the idea is to utilize different cards in different stages of development - and to not necessarily use all cards in every project either. Each software development endeavor is unique, e.g. in relation to the requirements and the scope of the project. ECCOLA cards should therefore also be selected based on the project and tasks at hand. Cards irrelevant to the current situation can be discarded during the sorting process. The sorting should preferably be conducted before the development process starts, so that the prompts presented by the cards can be utilized from the beginning. The sorting process should include everyone who will be using the cards, and possibly other members of the project who are involved with the product's development.

Before starting to use the cards in a development process, we recommend sorting the cards into piles based on which stage of the development they will be used in. Cards that are deemed irrelevant for the project can simply not be used during that project. This selection process should be documented by briefly explaining why some cards were selected and why some were considered irrelevant in each iteration, to support transparency in the context of systems development. Documenting ethical choices in general is encouraged while using the method. Our recommendation for sorting the ECCOLA cards is to create three piles of cards.

Pile 1 for the early stages and planning stages in a project. Pile 2 for any other parts of the project, throughout development. These should be adjusted on a sprint-by-sprint basis as well. The chosen cards, or specific parts of each card, can then be considered in relation to the activities in that sprint. Finally, Pile 3, if needed, towards the end of the project if there is a need to evaluate a decisions, or if there have been any unexpected occurrences.

When introducing ECCOLA to new organizations and people interested in using it, we have typically held an introductory workshop, which we discuss in the subsection below.

3.1.1. Getting acquainted with the cards/tutorial sessions

To introduce new users to ECCOLA, we have held tutorial sessions in the form of workshops. Similar sessions could also be held in organizations looking to start using ECCOLA. Below is a brief outline of these sessions.

The following outline has been used for ECCOLA tutorials:

1. A presentation on ECCOLA (and AI ethics if necessary).

2. Introducing the hypothetical product and planning its features and requirements.
3. Sprints 1, 2 and 3 where new features or requirements are introduced for each sprint. Each sprint lasts e.g., 15–20 min.
4. Discussion and feedback.

The introduction should familiarize the participants with the method, and can contain a brief introduction to AI ethics as well, focusing on why it is important and what it is, with a focus on practical issues. After the introductory presentation, the participants are given a task to work on. For example, during the COVID-19 pandemic, we had workshop participants design an AI-based mobile application for tracking and limiting its spread. The participants then split into groups (e.g., 5 per group) and design such a system according to the given requirements while using the ECCOLA cards.

This work is carried out in three sprints of e.g., 15–25 min. Each sprint can contain pre-selected cards, or the participants can be instructed to choose the cards themselves for each sprint. If the participants are to select their own cards, the sprints should also be longer in duration. Between sprints you can have a brief discussion session, or you can go through the sprints in quick succession and have a longer one afterwards.

4. Research method

In this section, we discuss the Cyclical Action Research approach we have utilized to develop ECCOLA. Our approach was based on that discussed by Susman and Evered (1978) and, in further detail, by Davison et al. (2004). We chose this approach as we wanted to iteratively develop the method over time, testing it in different contexts in the process. Moreover, Action Research (AR) is well-suited for using different data collection methods in different contexts (Susman and Evered, 1978).

Thus far, we have completed 7 Action Research (AR) cycles and are currently conducting an eighth one. These have been split into 6 research *stages*, with most research stages featuring one cycle, aside from stage 2 that consisted of three cycles. These are shown in Fig. 4 and Table 2, and each stage is further discussed in the following data analysis section. In this current section, we discuss the cyclical research approach of this study more generally from a methodological point of view.

Past the very first AR cycle that focused on testing an existing tool, each cycle has proceeded in the same general manner. In each cycle, we have tested a version of ECCOLA in practice in some context, collected data from its use, and then used the data to improve the method. After this, we have started a new cycle. In the diagnosis phase of each cycle, we have looked at literature on the topic to determine whether ECCOLA should be further modified based on literature before a new test in a different context.

The initial cycles (Stages 1–2) focused on student testing. We used student projects early on as we wished to make the method more mature before industry testing. In Stage 3, we started to also include industry testing in the form of a small-scale blockchain project. In addition to this, in Stage 3, we began to host academic workshops at conferences, as well as privately organized academic workshops, to collect feedback from the scientific community (using the Tutorial Session outline in Section 3.1.1). Finally, we shifted our focus further towards industry testing in Stages 5 and 6, and we are currently cooperating with multiple companies using ECCOLA. The way we have progressed from student testing to industry testing in this fashion is also inspired by the continuous co-experimentation approach described by Mikkonen et al. (2018).

In our industry testing, we have utilized an approach has been referred to as industry-as-a-lab by Potts (1993). This approach focuses on "what people actually do or can do in practice". As many of the current problems in the area resulting in the gap between research and practice seem to stem from a lack of practical tools, we have focused on making ECCOLA practical. To achieve this, we have focused on receiving continuous feedback primarily through formal data collection and throughout the process improving the method based on the feedback before then testing it again. A more recent example of this approach is the study of Mikkonen et al. (2018).

Finally, perhaps worth noting is that the research team behind this endeavor has past experience in developing methods as well. Namely, one of the authors proposed the Mobile-D approach for developing mobile applications in an Agile manner when Agile was still emerging (Abrahamsson et al., 2004).

In the subsections below, we discuss each phase of the Cyclical Action Research model discussed by Susman and Evered (1978) (and Davison et al. (2004)). Susman and Evered (1978) highlight five phases (Fig. 3) in this cyclical process that they posit are all necessary. We describe our process according to these phases in the subsections of this section.

4.1. Diagnosis

In the initial cycle, diagnosing the problem was focused on understanding the gap in AI ethics in general. We have published papers about this in the past, with Vakkuri et al. (2020) looking at this gap quantitatively and e.g. Vakkuri et al. (2020) looking at it qualitatively. While collecting data for these papers, we began to see that there is indeed a gap between research and practice in the area, and started to also look for ways to bridge the gap.

In Stages 2 and up, when we were already developing ECCOLA, the diagnosis phases focused on better understanding *what* is AI ethics and, to this end, what exactly is the problem ECCOLA should help solve. In addition to improving ECCOLA based on our data from each preceding cycle, in the diagnosis phase of each cycle, we looked at motivation behind ECCOLA. Whereas Action Research traditionally focuses on solving problems an organization has, in this case, it was largely up to us to define the problem and then convince organizations that it was a real problem. However, towards the latest stage, we have noticed that AI ethics has become much more topical out on the field to the point where we have had organizations volunteering to work with us on developing ECCOLA.

The main question in the diagnosis phase of each cycle was always whether our idea of AI ethics was still up-to-date. Was ECCOLA still in line with the current discussion on AI ethics? For example, the EU guidelines on AI ethics (HLEG, 2019) were published after Stage 2 (Fig. 4), and in our minds presented a major contribution to the field, which we felt should also influence ECCOLA.

4.2. Action planning

In the first stage (Section 5.1) where we ultimately tested the RESOLVEDD strategy, we considered alternative courses of action. Having identified a gap in the area, we looked at different alternatives for solving the problem. Using the existing AI ethics guidelines to bridge the gap was one option. However, existing papers argued that ethical guidelines alone were unlikely to work in AI ethics (Mittelstadt, 2019) or SE engineering in general (McNamara et al., 2018).

We therefore turned to methods that could help us tackle it. First, we looked at existing methods for implementing ethics. As a result, in Stage 1 of our study (Section 5.1), we studied

Table 2
Cyclical action research stages.

Stage	Version in action	Primary background theories	Study setting	Timing	Participant
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
2	2	RESOLVEDD, EAD, Essence	Class	Q2 2018 - Q2 2019	27 teams of 3–5 students
2	3	RESOLVEDD, EAD, Essence	Class	Q2 2018 - Q2 2019	27 teams of 3–5 students
3	4	EU AI HLEG, EAD	Blockchain Project	Q2–Q3 2019	2 sw development team members
4	5	EU AI HLEG, EAD	Conference Workshop	Q4 2019	8 researchers
5	6	EU AI HLEG, EAD	Industrial & Conference Workshops	Q1–Q3 2020	2 Company cases & 10+ ICT researchers
6	7	EU AI HLEG, EAD	Industrial	Ongoing	

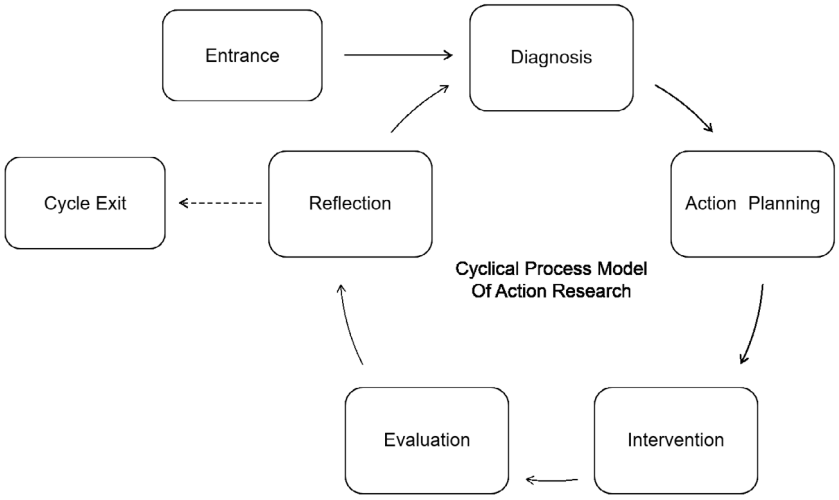

Fig. 3. Based on Davison et al. (2004) and Susman and Evered (1978).

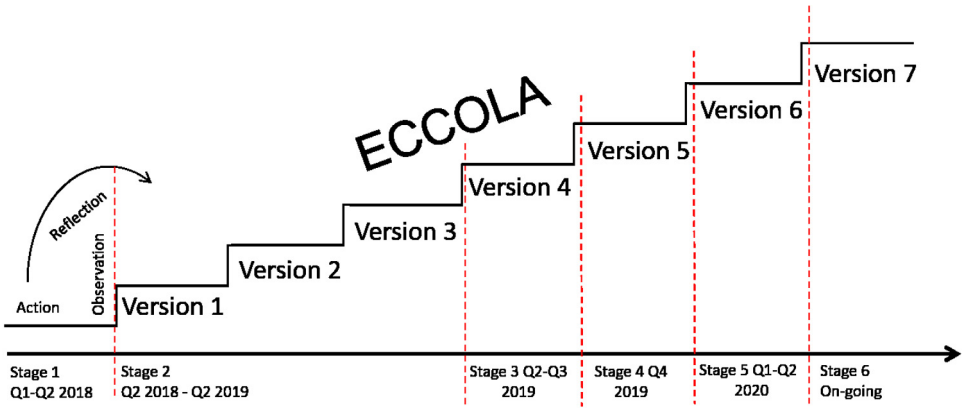

Fig. 4. Cyclical action research process on ECCOLA. Including cycle of action, observation, reflection on each iteration.

an existing ethical tool from the field of business ethics, the RESOLVEDD strategy, in the context of AI ethics, and argued based on our findings that methods and tools specific to AI ethics are required (Vakkuri and Kemell, 2019). As a result, in the absence of existing AI ethics methods, we began to work on ECCOLA.

In the stages past Stage 1, Action Planning was focused on determining how to test each version of ECCOLA. This included deciding on what type of data to collect and how. As we had already committed to developing ECCOLA, we no longer actively considered other ways of tackling the gap.

4.3. Intervention (or action taking)

The main intervention in all the stages of this study past the first one has been the introduction of ECCOLA. In the student and industry contexts, the project would have existed and been carried out with or without ECCOLA. ECCOLA was simply introduced as a framework for conceptualizing a problem (i.e. various ethical issues). This can be likened to the way Susman (1976) describe surprise in interventions: "the element of surprise evoked by an intervention results when the change agent offers members of the target organization a new way to conceptualize an old problem and offers it in a language or framework that differs from that by which members of the organization define their present situation". On the other hand, the academic workshops were created for the sole purpose of having the participants use ECCOLA, even though the mini-projects of the workshops could have been carried out without ECCOLA as a framework.

The introduction of ECCOLA has been accompanied by other actions taken to facilitate its adoption and use. These have varied between the research stages, but each stage has generally included 1) an introductory lecture or a workshop on ECCOLA, and 2) various check-ups to discuss the use of ECCOLA and any problems faced while using it. These have been used for data collection purposes as well, with especially the check-ups serving as a way of generating important data in the form of feedback for the evaluation phase of each Action Research (AR) cycle.

In student contexts, the use of ECCOLA continued for a set amount of weeks during a course project. In academic contexts, i.e. workshops, the use of ECCOLA lasted some hours. In industry contexts, the use of ECCOLA lasted for a duration of a project (Stage 3) or is still on-going (Stage 6).

4.4. Evaluation

Evaluation was conducted both during and after the use of ECCOLA in each stage. The focus of the evaluation was to understand what effect ECCOLA had had on the way its users worked, i.e. how it had changed existing practices and whether it had added new work practices. In doing so, we wished to also understand how the users of ECCOLA had felt about ECCOLA while using it.

We collected different types of data in different stages of the study (Fig. 4, Table 2). Across these stages, we have used work products (sheets, notes, text etc.), ECCOLA cards with notes on them, observation, unstructured interviews, and informal discussions as sources of data. In the next section (Section 5), we discuss what types of data were used in each stage in the respective subsections. The data collected in each stage is also summarized in Table 3.

4.5. Reflection (or specifying learning)

As we have developed ECCOLA iteratively in this process, the reflection phases have primarily focused on improving ECCOLA based on the data collected in each research stage. Indeed, the evaluation of ECCOLA has also been the focus of the data collection. In each reflection phase, we looked at ECCOLA from two points of view.

First, we looked at how ECCOLA had worked as a method in that stage. Had the method itself been clear to its users? Had the users managed to follow the process suggested by ECCOLA? To determine this, we looked at the notes on the ECCOLA cards and other work products to see how (or if) the cards had been utilized, or discussed their use with the subjects for example.

Secondly, we looked at the theory behind ECCOLA, i.e. AI ethics. Were we presenting the principles in an understandable way and were the users of ECCOLA grasping the concepts? Was something missing based on the data, or did something need to be further emphasized? For example, sometimes we would receive direct feedback regarding the wording on some of the cards.

Additionally, we critically evaluated our research process and choices regarding it. We looked at shortcomings in our data collection methods and how we introduced ECCOLA into the research context in each cycle. For example, the introductory session we have hosted at workshops and for companies (see Section 3.1.1) has been improved over time as well.

5. ECCOLA development stages and data

ECCOLA has been developed iteratively through multiple stages. In each stage, we have collected empirical data, which has then been used to iteratively improve the method. The current version of ECCOLA is its seventh version. The subsections of this section each present one development stage in the iterative development process of ECCOLA. At the end of each section is a brief summary of what changes were made in each stage. This process is also summarized in Table 2 below, as well as in Fig. 4.

5.1. Stage 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 (Jacobson et al., 2012), 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 and Kemell (2019)). 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 developers, and they felt that using the tool was detached from the rest of the processes. Moreover, when forcing developers to utilize such a tool, the commitment towards it quickly vanished when the tool was no longer compulsive.

Stage 1 actions: The development of ECCOLA was initiated

5.2. Stage 2 (Q2 2018 - Q2 2019)

5.2.1. 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 (Pfeiffer and Forsberg, 1993), (2) The Essence Theory of Software Engineering (Jacobson et al., 2012), and (3) The IEEE Ethically Aligned Design guidelines (IEEE Global Initiative, 2019).

We utilized some of the general ideas of RESOLVEDD, which were deemed useful based on the data we collected. Namely, we (1) looked at RESOLVEDD for ideas on how to make the tool function in conjunction with iterative SE methods, and (2) for ideas on how to conduct comprehensive stakeholder analyses as the basis of the ethical analysis. We also included some of the aspects of RESOLVEDD which were shown (Vakkuri and Kemell, 2019) to support transparency of systems 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 2.4). 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 (IEEE Global Initiative, 2019). The field in general was still less formulated than it currently is, and thus the main AI ethics principles were still under more discussion than they currently are (e.g., Jobin et al. (2019) show that the field has since reached some consensus). 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

Table 3
Research stage and data collection.

Research stage	Data collection tools
1	Semi structured interviews for users
2	Note taking, mentor meetings, work-product (course), ECCOLA cards (user notes)
3	User interview, note taking, ECCOLA cards (user notes)
4	Note taking during workshop, unstructured participant interview, workshops recording
5	Note taking during workshop, unstructured participant interview, workshops recording, ECCOLA cards (user notes)
6	Note taking during tutorial, works, recurring project meetings, workshops recording, unstructured developer interview, ECCOLA cards (user notes), project documentation

utilized the studies of Dignum (2017) and Ananny and Crawford (2018), 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 academic articles discussing some individual principles.

5.2.2. Testing Version 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 (Q1 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 refine the ideas as far as they could in the context of the course. The results of these projects have been published in an educational book[6]

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 – or if – they used them. After the students had utilized the cards for a week, they were collected and the written notes on them analyzed. 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 (Jacobson et al., 2019), which explains the tool. Though the educational goal of this was elsewhere, this also served to make sure the students would not be overtly confused with this version of ECCOLA being described using the Essence language.

Based on the data collected, the language on the cards was considered difficult to understand and overall they were considered too academic by the teams. The cards were considered impractical, with the teams having difficulties applying their contents into practice. The students were also confused by the Essence notation.

Actions based on Iteration 1 of Stage 2, for Version 2: (1) Alpha states were added to the alphas in order to make tracking progress on them easier. (2) Practical examples were added to the cards to make it easier to understand the practical implications of the ethical issues in the cards. (3) Reduced the amount of academic jargon on the cards, focusing on practice over theory. (4) Removed list of academic references from each card.

5.2.3. Testing Version 2 (Q1 2019)

This iteration took place during the course described above and was carried out in the same manner as the previous one. The same student teams utilized this newer version of ECCOLA again while writing down notes on the cards 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 evaluated using the data we collected. The teams still found the method confusing. In particular, they found it difficult to understand how the cards tied together, and how they should be utilized. Even if the individual cards were made more practical, the language was still considered difficult to understand. Thus, the following changes were made to the method based on the data.

Actions based on Iteration 2 of Stage 2, for Version 3: (1) 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. (2) Added numbering to the cards. (3) Further reduced the amount of academic jargon on the cards.

5.2.4. Testing Version 3 (Q1 2019)

The third version of ECCOLA was also tested in the same course as the previous two. 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.

In analyzing the data from the teams, we focused on evaluating the level of utilization. This was done by analyzing the notes the teams made on the cards. The notes were evaluated on a scale of 0 = no notes or markings, 1 = single words or markings, 2 = sentences or more.

Also, we evaluated the cards independently based on the notes. The cards that were utilized the most and affected the projects the most were either cards with practical themes (e.g. data handling), or cards focusing on the big picture of the project at hand (e.g. cards focusing on 'what' and 'how' questions). On the other hand, the cards that were utilized the least, were the ones focused on accountability and other AI ethics specific issues. It seemed that many of the AI ethics principles, even with practical examples, were considered difficult (or irrelevant) by the teams. The cards describing AI ethics principles were utilized by 53% of the teams, whereas the other cards had a utilization level of 75% on average.

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

5.2.5. Creating Version 4 (Q2 2019)

The earlier versions of ECCOLA were cumbersome to use based on initial tests (see above). Utilizing these versions did result in ethical analyses and had an impact on the projects. However, the method was difficult to understand and especially the AI ethics

[6] http://urn.fi/URN:ISBN:978-951-39-7689-7.

principles in particular were difficult to grasp for the teams utilizing the tools. After the course in which the first three versions of the method were tested, we made larger improvements based on the data.

First, we changed the way the method was described. We opted to lessen the role of Essence in ECCOLA. The Essence language used to describe the method seemed to make the method even more difficult to learn, as its users had to learn to use the method *and* to learn to understand the Essence language (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 making the method a card deck seemed to work and thus this approach was kept.

Secondly, the method seemed to be too heavy to use. ECCOLA was initially designed to be a linear process that was iteratively repeated. The idea was that its users could modify this process based on the context at hand to adjust the method to their projects. Nonetheless, this approach was considered too rigid, and the respondents felt, that it was just another process tacked onto their other work processes. Thus, we made the method modular, with the cards being more stand-alone on average, though some cards were still linked together in some ways. The users of ECCOLA could, following this approach, choose which cards to utilize in each situation (e.g., sprint) based on the context. The intent behind this was to make ECCOLA more suited for use with Agile methods.

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 (HLEG, 2019) were 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.

Changes made based on Stage 2 overall: (1) The use of Essence to describe the method was discontinued. (2) Contents of the cards reformatted and reformulated. (3) Method made modular rather than one linear, iterative process. (4) Expanded the AI ethics theoretical basis of the method.

5.3. Stage 3 (Q2-Q3 2019)

As the primary concern with the versions 1–3 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, which is 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 stage, 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 5th version of ECCOLA.

Changes made based on Stage 3: (1) A note-making space was added to each card. (2) Added new cards. (3) Split the cards into themes, such as transparency or data. (4) 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. (5) Added new content focusing on stakeholder analysis and requirements, in order to help the users of the method gain an understanding of the big picture at hand.

5.4. Stage 4 (Q4 2019)

After improving ECCOLA based on the lessons learned from the blockchain project, ECCOLA was presented in a workshop in a scientific conference (ICSOB2019). In this workshop the participants utilized ECCOLA to discover potential ethical issues in a hypothetical AI development scenario. The participants of the workshop were split into two groups for the task.

The first group was tasked with developing an idea for an AI-based drone that would help farmers improve their harvests. The second group was tasked with developing an AI-based system that would filter and evaluate immigration applications. During the workshop, the groups worked on the ideas in timed iterations. Each group had a customer stakeholder that progressively presented them with more requirements at the end of each iteration. For every iteration, the groups would select the ECCOLA cards they felt were the most relevant for the requirements of that iteration.

At the end of the workshop, verbal feedback from the participants was collected. This was done in the form of a discussion where the participants talked about their experiences with each other and between the two groups. These group interviews were recorded and later transcribed for analysis. The feedback was then utilized to develop the 6th version of ECCOLA.

Changes made based on Stage 4: (1)The themes in the cards were color coded for clarity. (2)The practical examples in the cards were improved.

5.5. Stage 5 (Q1-Q3 2020)

A paper presenting the early 2020 (i.e., 6th) version of ECCOLA was published at DSD/SEAA2020 (Vakkuri et al., 2020). This paper extends said DSD/SEAA paper.

In the first half of 2020, ECCOLA was presented at the XP2020 conference in a workshop. The workshop was organized in a similar manner as the one at ICSOB2019 described in the previous subsection, with some modifications. The participants were split into three groups and tasked with working on a hypothetical AI/S project where they were to design a system for COVID-19 spread monitoring, while using ECCOLA to dwell on the potential ethical issues. This time, as the conference was held remotely, the participants communicated online, utilized a digital version of ECCOLA, and produced work products online. The work products (written documents) produced by the teams were collected for later analysis of the use of ECCOLA.

Additionally, we have held three privately organized ECCOLA workshops not associated with any scientific conference. These have been workshops for researchers active in the field, for the purposes of various research projects. These have been organized in a similar manner to the conference workshops, with the participants utilizing ECCOLA to work on a hypothetical project after a brief introduction to the method.

During 2020, ECCOLA was also adopted by three companies. One of these companies began using ECCOLA as early as late Q1 2020. In preparation for further company adoption, we utilized the workshop data, preliminary feedback from this one case (unstructured), and the other data collected in earlier stages, to create the current (7th) version of ECCOLA.

Changes made based on Stage 5 (resulting in the current version of ECCOLA): (1) Improved card layout based on company feedback (numbered card contents for easier referencing). (2) Improved individual card readability and textual content based on early company feedback with a focus on reducing the chance of any of the content being misunderstood. (3) Made changes based on current academic discussion. (4) Improved some of the practical examples on the cards with a focus on making them less tied to any current real events. (5) Fine-tuned the visual appearance of the cards.

5.6. Stage 6 (on-going)

Currently, we are cooperating with three companies to collect industry use data on ECCOLA. These companies are detailed in Table 4. With each company, we have held a workshop similar to the ones we have held at conferences to introduce them to the method. After this, we have kept in touch with the companies regarding the utilization of the method through recurring meetings. While we have collected data from these meetings as notes and discussed their experiences using ECCOLA during the meetings, these cases are still pending formal data collection.

So far, in our discussions with the participants, the companies have indicated that they have successfully utilized ECCOLA in conjunction with their existing methods. They feel that ECCOLA has successfully been modular. To this end, ECCOLA also seems to work in conjunction with agile methods, as all the companies consider themselves agile. However, we have not yet collected any work products or ECCOLA cards with notes from the companies. The projects are also still on-going, and thus we have not yet been able to conduct formal interviews discussing their ECCOLA use experiences in more detail. As a result, this stage is still on-going as well.

Additionally, ECCOLA has been accepted for presentation in another scientific conference workshop at ICSE2021. This workshop will be held in a similar manner in hopes of further improving the method where needed. Though the developmment of ECCOLA continues, we feel that we have reached a stage where we wish to share ECCOLA with the scientific community and the industry at large. Given the current lack of methods for AI ethics, with the industry largely reliant on guidelines to implement AI ethics, ECCOLA can serve as a starting point in the area, as we discuss next.

6. Discussion

The ECCOLA method was created 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 (Vakkuri et al., 2020). Through ECCOLA, we have attempted to make some of the contents of the IEEE EAD guidelines (IEEE Global Initiative, 2019) and the EU Trustworthy AI guidelines (HLEG, 2019) actionable, alongside other research in the area.

We use the three goals we had for ECCOLA, which we discussed in the Introduction and Section 3, to structure the discussion in this section. These goals were (1) to help create awareness of AI ethics and its importance, (2) to make a modular method suitable for a wide variety of SE contexts, and (3) to make ECCOLA suitable for agile development, while also helping 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 also been why we have, for now, chosen to focus on raising awareness and highlighting (potential) issues rather than trying to provide direct solutions for ethical questions. Raising awareness has also been a goal of the IEEE EAD initiative (IEEE Global Initiative, 2019). In general, raising awareness is important as AI ethics is a new topic for the industry.

On the other hand, it would be possible to select a specific set of AI ethics guidelines, such as the EU ones (HLEG, 2019), and study whether a tool or a method would help organizations implement those. While ECCOLA is not based on any *one* set of guidelines, the EU guidelines have heavily influenced it, and this is something future studies on ECCOLA should tackle. So far, as ECCOLA is still being iteratively developed further, we have not yet conducted such a study, focusing instead on improving the method before looking to further confirm its usefulness past what we have presented here.

Currently, 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: (1) ECCOLA raises awareness of AI ethics. It makes its users aware of various ethical issues and facilitates ethical discussion within the team. This could be seen on the notes made on the cards we collected from the users of ECCOLA during the different stages of its development, as well as in the discussions and interviews we had with its users. (2) 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 the product backlog. This could be seen in the notes made on the ECCOLA cards we analyzed while developing ECCOLA.

Transparency is one key issue in AI systems, both in terms of systems and in terms of systems development (Dignum, 2017). These documents, as we have done while testing the method, can also be analyzed to understand how the method was used, aside from seeking to understand the reasoning behind the ethical decisions made during development. Using ECCOLA produces a paper trail of decisions and choices as notes on the cards, alongside other types of written documents such as meeting notes.

So far, we have not utilized control groups while developing ECCOLA, focusing instead on improving the method before aiming to further quantify its effectiveness. We cannot thus argue, based on our data on ECCOLA so far, that ECCOLA would have increased ethical consideration over a baseline of no ethical tool being utilized. On the other hand, we did study the use of the RESOLVEDD strategy in a past paper, which we also briefly discussed here due to its relevance in motivating the development of ECCOLA, and argued that the presence of an ethical tool in general seems to increase ethical consideration (in a student setting). Moreover, out on the field, the baseline largely seems to be that ethical aspects are currently ignored (Vakkuri et al., 2020; Vakkuri et al., 2020). With these studies in mind, we consider it likely that ECCOLA does increase ethical consideration over a baseline of no tool being utilized. However, the effects of ECCOLA on ethical consideration should be further looked into in future studies. This could be done by e.g. studying whether ECCOLA helps fulfill the requirements of one particular set of guidelines, as we have discussed above.

Compared to a baseline where no ethical methods are used, ECCOLA can thus already be argued to increase ethical consideration during development based on this data. This was also the case when we studied student teams using the RESOLVEDD strategy in an existing paper: it increased ethical consideration over the baseline of no ethical tool being used (Vakkuri and Kemell,

Table 4
Participant companies.

Company	Stage	Company description	ECCOLA users
Company A	5&6	small (<30 employees) SW company focusing in Maritime logistic	1 Project owner 2 developers
Company B	5&6	Micro (<10 employees) SW company focusing in data-driven solutions	1 Project owner, 2 developers, 2 consultants
Company C	6	Medium Multinational (>250 employees) SW consulting company	1 Project owner, 2 developers

2019) in a student setting. Out on the field, the baseline largely seems to be that ethical aspects are currently ignored (Vakkuri et al., 2020; Vakkuri et al., 2020). However, the effects of ECCOLA on ethical consideration should be further looked into in future studies. This could be done by e.g. studying whether ECCOLA helps fulfill the requirements of one particular set of guidelines, as we have discussed above.

The second goal has been based on the method-agnostic philosophy of the Essence Theory of Software Engineering (Jacobson et al., 2012). Industry organizations use a wide variety of methods, from out-of-the-box ones to, more commonly, tailored in-house ones (Ghanbari, 2017). ECCOLA is not intended to replace any of these. Rather, ECCOLA is a modular tool that can be added to existing methods and used in conjunction with them, lessening the barrier to its adoption. Though ECCOLA is still being studied in industry settings and we are still collecting data from these cases, so far none of the companies have discussed any issues incorporating ECCOLA into their existing ways-of-working.

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 iterative process was not a suitable approach due to being too heavy. The users of the method opted out of adhering to the process and used the cards in a modular fashion despite the instructions asking them to repeat the full process every time. 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. Based on our data, the users of the method prefer this approach, and it seems to work in Agile development as the companies utilizing it are all Agile and have had no issue incorporating it into their way-of-working.

On the other hand, we do not know whether this is detrimental from the point of view of implementing ethics. Do the users of the tool make informed decisions about which cards to exclude? Would advising them to go through a full process (or e.g. all the cards in each iteration in this case) result in more ethical consideration? However, as this is a question of whether ECCOLA helps implement ethics (and to what extent), this is more related to the first goal discussed above.

In designing ECCOLA, we have also turned to VSD (Section 2.3) for some inspiration. First, as already mentioned, we have also chosen a gamified approach in the form of a card deck for ECCOLA. Secondly, both VSD and ECCOLA are iterative methods that can be used in conjunction with SE methods. Thirdly, both methods take on a proactive perspective to ethical consideration in the design or development process. Fourthly, there is some overlap in ethical themes in the methods (e.g., privacy, stakeholder analysis, etc.). On the other hand, they differ in their theoretical backgrounds (SE vs. IS), how ECCOLA is far more focused on the perspective of SE and developers, and how ECCOLA is an AI/S-specific method as opposed to a general design method.

Overall, ECCOLA is intended to become a part of the agile development process in general. Ethics should not be merely an afterthought. Ethics should be another set of non-functional requirements, as well as a part of the user stories for the system. ECCOLA is a tool for developers and product owners. Ethics cannot be outsourced, nor can ethics be implemented by hiring an ethics expert (Vakkuri et al., 2020). AI ethics should be in the requirements, formulated in a manner also understood by the developers working on the system.

As governments and policy-makers have already begun to regulate AI systems in various ways (e.g., bans on facial recognition for surveillance purposes,[7] 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 next to none. For the time being, only some commercial methods have already been proposed for AI ethics (e.g.,[8,9]).

7. Threats to validity

In this section, we discuss the limitations of the study through validity threats. These threats are split into four categories as follows: reliability, construct validity, internal validity and external validity.

7.1. Reliability

First, reliability. The research approach chosen here, action research, on its own already presents threats to reliability. As the research approach influences the research target (organization), changing it and producing unreliability, it is not possible for subsequent studies to carry out the same study in the same context.

We have had separate plans for data collection in each stage. The types of data collected are detailed in Table 3. Most of the data used to develop ECCOLA has either been user notes on ECCOLA cards or unstructured interview data. However, in the later stages while working with companies, we have collected increasing amounts of informal discussion data as e.g. meeting notes.

While collecting data, we have mostly kept our distance as researchers, maintaining a distinct role and doing our best to only collect data while avoiding advising or leading the participants on into any direction. However, in the workshops, academic and company ones, we have occasionally involved ourselves in the group work as facilitators while trying to not provide any answers to the workshop participants. In analyzing our data, we have had multiple researchers (two or three) involved in the analysis process in an attempt to limit researcher error and bias.

Additionally, in action research, an audit trail is recommended by some authors. We would highlight our past publications in the area as one type of audit trail in this regard. We published our results from testing the RESOLVEDD method in the context of AI ethics (Vakkuri et al., 2019), we published an earlier version of ECCOLA in another paper (Vakkuri et al., 2020), and we have studied the gap in the area in existing studies (e.g. Vakkuri et al. (2020) among others).

[7] https://www.bbc.com/news/technology-51148501.
[8] https://www.ideo.com/post/ai-ethics-collaborative-activities-for-designers.
[9] https://www.33a.ai/ethics.

7.2. Construct validity

The construct validity of this study has three primary threats as we see them: 1) the research strategy, 2) the construct of method, and 3) the construct of ethics. Cyclical action research is a typical SE research approach. Additionally, in designing our research strategy, we have utilized existing studies that have proposed methods in SE in designing our strategy in more detail (e.g. Fagerholm et al. (2017)). In terms of data collection and use, we looked at another study that has proposed an Agile method in the past (Abrahamsson et al., 2004). We have described our research strategy in detail in Section 4.

As mentioned in the background section, ethics and values can mean different things to different individuals (Friedman et al., 2013), and different cultures may have different ethical theories. To tackle this potential threat to validity, ECCOLA tries to be agnostic in terms of ethical theories and the definition of ethics. ECCOLA presents potential issues that should be tackled, but leaves it up to the users of the tool to decide on how to tackle them. It asks questions but does not provide the answers directly. Admitted, values such as privacy are not equally important to everyone, and as such ECCOLA does take on a stand to some extent in terms of which AI principles it includes. However, these principles are grounded in existing research and white and gray literature in the area.

Another threat to construct validity is related to the construct of method. Methods in SE describe ways of working. They consist of techniques (IS) (Tolvanen, 1998) or practices (SE) (Jacobson et al., 2012) which together describe how work should be carried out by an organization. Past studies have argued that developers prefer simple and practical methods, if they use any at all (Abrahamsson and Iivari, 2002). Moreover, organizations tend to tailor methods into in-house ones better suited for their specific context (Ghanbari, 2017), which is also something Essence encourages (Jacobson et al., 2012). To make ECCOLA desirable to the industry, we have 1) made it modular to let organizations tailor it, 2) designed it to be used on conjunction with existing SE methods, and 3) to make it more practical. The industry-as-a-lab approach (Potts, 1993) we have used in the later stages of ECCOLA's development is intended to ensure that ECCOLA is practical.

7.3. Internal threats to validity

The main threat to internal validity so far is that we cannot ascertain that ECCOLA produces ethical AI systems, and thus we do not claim that it does. This is not only a challenge in the data we have utilized, but also on a more general level: there are, as far as we know, no benchmarks or measures for ethical AI. On the other hand, we have argued that ECCOLA helps implement AI ethics and produces more ethical consideration during development, compared to a situation where no ethical method is used. Our data indicates that using ECCOLA results in ethical consideration. However, what actions are taken as a result of the ethical consideration is ultimately up to the developers and the organizations.

The wide variety of data we have utilized here presents both internal and external (discussed next) threats to validity, having been collected from different contexts and using different data collection methods. Most of the data we now have on ECCOLA has been collected after influencing the subjects in some way (as opposed to having both before and after data). We wanted to avoid asking questions beforehand so as to not direct the subjects into any particular line of thinking in relation to AI ethics. Instead, we wanted to have our subjects work as usual while additionally utilizing ECCOLA to be able to see how they use the tool. This has, however, made it difficult to measure any changes in attitudes in the subjects, or any other such changes that could be measured based on data collected both before and after utilizing ECCOLA. To this end, wanting to primarily focus on improving the method based on user experiences, we have not utilized control groups in the earlier stages to further ascertain its impacts.

Aside from what we can say based on our data on the use of ECCOLA, we would also again highlight other ethical tools discussed earlier in this paper, namely the RESOLVEDD strategy (Pfeiffer and Forsberg, 1993) and the Tripartite Method and the associated Envisioning Cards (discussed in Section 2.3). In designing ECCOLA, we have studied these existing approaches for involving ethics in broader business and development contexts, which have been argued to increase ethical consideration, and adopted similar elements as a part of ECCOLA. We would argue that ECCOLA, being founded on these approaches, should have retained some of their effectiveness in increasing ethical consideration when used.

7.4. External threats to validity

As we have utilized a wide variety of data while working on ECCOLA (data from students, companies, conference workshops, and interviews, notes, observation, etc.), these different data collection and analysis approaches present an equally wide variety of potential threats. We have, especially early on, utilized student data from classroom settings. We felt that having students utilize the method in its early stages would still provide us with data on, e.g., whether the AI ethics principles in the method were understandable and whether the process suggested by the method made sense. This let us make even large changes to the method without inconveniencing any industry organization using it, as it was still confined to a student setting. We had a large number of students use the method, giving us ample data to work with early on. However, in this case, the student setting is quite different from an industrial one (e.g. in a student project, the shortcomings of an immature ECCOLA would not result in a project manager getting into trouble).

On the other hand, when working with companies, we have thus far relied on a low number of cases, e.g. 1–3 case projects at a time. Moving forward, we wish to widen the industrial testing (and use) of ECCOLA, but while developing the method, we wanted to get more in-depth feedback from fewer cases to improve the method while working in closer cooperation with the involved parties. This presents a threat to validity as data from a low number of companies makes it less generalizable. We would turn to Eisenhardt (1989) who argues that for novel research areas (in case study research), such a low number of cases can be an acceptable number. While Eisenhardt speaks of case studies in particular, the issue of generalizability is still present in other research approaches as well. Empirical studies in AI ethics are currently few in number, and there seems to be a gap in the area (Vakkuri et al., 2020). In particular, studies on methods such as ECCOLA in the area hardly exist. In this light, we would argue that even a few cases is better than none in moving forward in this novel research area.

8. Conclusions

In this paper, we have presented a method for implementing AI ethics: ECCOLA. It is an approach intended to make AI ethics more practical for developers and organizations. Whereas guidelines can seem abstract to developers, methods are a typical approach to software engineering. To this end, ECCOLA is intended to help organizations develop more ethical AI systems by making AI ethics issues a part of the development process.

The method takes on the form of a card deck, as we discussed in more detail in Section 3. These cards from a modular method which can be tailored according to the use context. For example, one sprint may only feature a handful of cards. The method supports iterative development and can be used in conjunction with existing SE methods. Indeed, ECCOLA is not a novel approach to SE but a tool for better involving AI ethics into the development process, to be used with existing methods.

ECCOLA has been developed iteratively using the Cyclical Action Research approach (Susman and Evered, 1978) and continuous experimentation (Mikkonen et al., 2018). During its development thus far, we have gone through a number of stages, discussed in Sections 4 and 5. In each stage, we have collected data, with a focus on empirical data on the use of ECCOLA. In the process, we utilized both student data and project data from industry projects, as well as feedback from academic workshops. Though ECCOLA is still being developed further, we have reached a state of maturity where we wish to share the method with the scientific community, as well as the industry.

The use of ECCOLA in practice is discussed in Section 3.1 of this paper. The materials for using the method (cards, instructions) can be downloaded from (https://doi.org/10.6084/m9.figshare.12136308).

Declaration of competing interest

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

Acknowledgments

This research is partially funded by Business Finland (businessfinland.fi) research projects: Sea4Value-Fairway & Stroke-Data and ITEA4 (itea4.org/project/mad-work.html) research project: Mad@Work. The authors are grateful for the founders for their support.

References

Abrahamsson, P., Hanhineva, A., Hulkko, H., Ihme, T., Jaälinoja, J., Korkala, M., Koskela, J., Kyllönen, P., Salo, O., 2004. Mobile-d: An agile approach for mobile application development. In: Companion To the 19th Annual ACM SIGPLAN Conference on Object-Oriented Programming Systems, Languages, and Applications. Association for Computing Machinery, New York, NY, USA, pp. 174–175. http://dx.doi.org/10.1145/1028664.1028736.

Abrahamsson, P., Iivari, N., 2002. Commitment in software process improvement - in search of the process. In: Proceedings of the 35th Annual Hawaii International Conference on System Sciences. pp. 3239–3248. http://dx.doi.org/10.1109/HICSS.2002.994403.

Ananny, M., Crawford, K., 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media Soc. 20 (3), 973–989. http://dx.doi.org/10.1177/1461444816676645.

Biffl, S., Aurum, A., Boehm, B., Erdogmus, H., Grünbacher, P., 2006. Value-Based Software Engineering. Springer Science & Business Media.

Davis, J., Nathan, L.P., 2015. Value sensitive design: Applications, adaptations, and critiques. In: Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains. Dordrecht: Springer Netherlands, pp. 11–40.

Davison, R., Martinsons, M.G., Kock, N., 2004. Principles of canonical action research. Inf. Syst. J. 14 (1), 65–86. http://dx.doi.org/10.1111/j.1365-2575.2004.00162.x.

Dignum, V., 2017. Responsible autonomy. arXiv preprint arXiv:1706.02513.

van der Duin, P., 2019. Toward "responsible foresight": Developing futures that enable matching future technologies with societal demands. World Futur. Rev. 11 (1), 69–79.

Eisenhardt, K.M., 1989. Building theories from case study research. Acad. Manag. Rev. 14 (4), 532–550. http://dx.doi.org/10.2307/258557.

Fagerholm, F., Sanchez Guinea, A., Mäenpää, H., Münch, J., 2017. The RIGHT model for continuous experimentation. J. Syst. Softw. 123, 292–305. http://dx.doi.org/10.1016/j.jss.2016.03.034.

Friedman, B., Kahn, P., Borning, A., 2002. Value Sensitive Design: Theory and Methods. Tech. rep., University of Washington technical report.

Friedman, B., Kahn, P.H., Borning, A., 2008. Value sensitive design and information systems. In: The Handbook of Information and Computer Ethics. Wiley Online Library, pp. 69–101.

Friedman, B., Kahn, P.H., Borning, A., Huldtgren, A., 2013. Value sensitive design and information systems. In: Early Engagement and New Technologies: Opening Up the Laboratory. Springer, pp. 55–95.

Ghanbari, H., 2017. Investigating the Causal Mechanisms Underlying the Customization of Software Development Methods (Ph.D. thesis). University of Jyväskylä.

Gotterbarn, D., Brinkman, B., Flick, C., Kirkpatrick, M.S., Miller, K., Vazansky, K., Wolf, M.J., 2018. Acm code of ethics and professional conduct. 2019, (18.3.), https://www.acm.org/code-of-ethics.

Hagendorff, T., 2020. The ethics of AI ethics: An evaluation of guidelines. Minds Mach. 1–22.

HLEG, 2019. Ethics Guidelines for Trustworthy AI. EU, https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

IEEE Global Initiative, 2019. Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, first edition. https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html.

Jacobson, I., Ng, P.-W., McMahon, P.E., Goedicke, M., et al., 2019. The Essentials of Modern Software Engineering: Free the Practices from the Method Prisons!. Morgan & Claypool.

Jacobson, I., Ng, P.-W., McMahon, P.E., Spence, I., Lidman, S., 2012. The essence of software engineering: the SEMAT kernel. Commun. ACM 55 (12), 42–49.

Jobin, A., Ienca, M., Vayena, E., 2019. The global landscape of AI ethics guidelines. Nat. Mach. Intell. 1 (9), 389–399.

Johnson, B., Smith, J., 2021. Towards ethical data-driven software: filling the gaps in ethics research & practice. In: 2021 IEEE/ACM 2nd International Workshop on Ethics in Software Engineering Research and Practice (SEthics). pp. 18–25. http://dx.doi.org/10.1109/SEthics52569.2021.00011.

Leikas, J., Koivisto, R., Gotcheva, N., 2019. Ethical framework for designing autonomous intelligent systems. J. Open Innov. Technol. Mark. Complex. 5 (1), 18.

Manders-Huits, N., 2011. What values in design? The challenge of incorporating moral values into design. Sci. Eng. Ethic. 17 (2), 271–287.

McNamara, A., Smith, J., Murphy-Hill, E., 2018. Does acm's code of ethics change ethical decision making in software development? In: Proceedings of the 2018 26th ACM ESEC/FSE. In: ESEC/FSE 2018, ACM, New York, NY, USA, pp. 729–733. http://dx.doi.org/10.1145/3236024.3264833.

Mikkonen, T., Lassenius, C., Männistö, T., Oivo, M., Järvinen, J., 2018. Continuous and collaborative technology transfer: Software engineering research with real-time industry impact. Inf. Softw. Technol. 95, 34–45. http://dx.doi.org/10.1016/j.infsof.2017.10.013.

Mittelstadt, B., 2019. Principles alone cannot guarantee ethical AI. Nat. Mach. Intell. 1–7.

Morley, J., Floridi, L., Kinsey, L., Elhalal, A., 2019. From what to how. An overview of AI ethics tools, methods and research to translate principles into practices. arXiv preprint arXiv:1905.06876.

Pfeiffer, R.S., Forsberg, R.P., 1993. Ethics on the Job: Cases and Strategies. Wadsworth Publishing Company.

Pichai, S., 2018. Ai at google: our principles. Accessed: 2021-04-30, https://www.blog.google/technology/ai/ai-principles/.

Potts, C., 1993. Software-engineering research revisited. IEEE Softw. 10 (5), 19–28. http://dx.doi.org/10.1109/52.232392.

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 1 (5), 206–215.

Rusanen, A.-M., Nurminen, J., Raisanen, S., Tarkoma, S., Halmetoja, S., 2021. Ethics-of-ai mooc. Accessed: 2021-04-30, https://ethics-of-ai.mooc.fi/.

Shilton, K., 2018. Values and ethics in human-computer interaction. Found. Trends Human Comput. Int. 12 (2).

Susman, G.I., 1976. Autonomy At Work: A Sociotechnical Analysis of Participative Management. Praeger, New York.

Susman, G.I., Evered, R.D., 1978. An assessment of the scientific merits of action research. Adm. Sci. Q. 582–603.

Tolvanen, J.-P., 1998. In: Tolvanen, J.-P. (Ed.), Incremental Method Engineering with Modeling Tools : Theoretical Principles and Empirical Evidence (Ph.D. thesis). In: Jyväskylä studies in computer science, economics and statistics, University of Jyväskylä.

Turilli, M., Floridi, L., 2009. The ethics of information transparency. Ethics Inf. Technol. 11 (2), 105–112. http://dx.doi.org/10.1007/s10676-009-9187-9.

Vakkuri, V., Kemell, K.-K., 2019. Implementing AI ethics in practice: An empirical evaluation of the RESOLVEDD strategy. In: International Conference on Software Business. Springer, pp. 260–275.

Vakkuri, V., Kemell, K.-K., Abrahamsson, P., 2019. Ethically aligned design: an empirical evaluation of the resolvedd-strategy in software and systems development context. In: 2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). pp. 46–50. http://dx.doi.org/10.1109/SEAA.2019.00015.

Vakkuri, V., Kemell, K.K., Abrahamsson, P., 2020. Eccola - a method for implementing ethically aligned AI systems. In: 2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). pp. 195–204. http://dx.doi.org/10.1109/SEAA51224.2020.00043.

Vakkuri, V., Kemell, K.-K., Jantunen, M., Abrahamsson, P., 2020. "This is just a prototype": How ethics are ignored in software startup-like environments. In: Stray, V., Hoda, R., Paasivaara, M., Kruchten, P. (Eds.), Agile Processes in Software Engineering and Extreme Programming. Springer International Publishing, Cham, pp. 195–210.

Vakkuri, V., Kemell, K., Kultanen, J., Abrahamsson, P., 2020. The current state of industrial practice in artificial intelligence ethics. IEEE Softw. 37 (4), 50–57.

VSD Lab, 2021. Value sensitive design lab. Accessed: 2021-01-05, https://vsdesign.org/.

Winkler, T., Spiekermann, S., 2018. Twenty years of value sensitive design: a review of methodological practices in VSD projects. Ethics Inf. Technol. 1–5.