Balasubramaniam, Nagadivya; Kauppinen, Marjo; Kujala, Sari; Hiekkanen, Kari

Ethical Guidelines for Solving Ethical Issues and Developing AI Systems

Published in:
Product-Focused Software Process Improvement

DOI:
10.1007%2F978-3-030-64148-1_21

Published: 21/11/2020

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Balasubramaniam, N., Kauppinen, M., Kujala, S., & Hiekkanen, K. (2020). Ethical Guidelines for Solving Ethical Issues and Developing AI Systems. In *Product-Focused Software Process Improvement* (pp. 331-346). (Lecture Notes in Computer Science; Vol. 12562). Springer. https://doi.org/10.1007%2F978-3-030-64148-1_21

Ethical Guidelines for Solving Ethical Issues and Developing AI Systems

Nagadivya Balasubramaniam[1] Marjo Kauppinen[1] Sari Kujala[1] Kari Hiekkanen[1]

[1] Aalto University, 02150 Espoo, Finland
firstname.lastname@aalto.fi

Abstract. **Artificial intelligence (AI) has become a fast-growing trend. Increasingly, organizations are interested in developing AI systems, but many of them have realized that the use of AI technologies can raise ethical questions. The goal of this study was to analyze what kind of ethical guidelines companies have for solving potential ethical issues of AI and developing AI systems. This paper presents the results of the case study conducted in three companies. The ethical guidelines defined by the case companies focused on solving potential ethical issues, such as accountability, explainability, fairness, privacy, and transparency. To analyze different viewpoints on critical ethical issues, two of the companies recommended using multi-disciplinary development teams. The companies also considered defining the purposes of their AI systems and analyzing their impacts to be important practices. Based on the results of the study, we suggest that organizations develop and use ethical guidelines to prioritize critical quality requirements of AI. The results also indicate that transparency, explainability, fairness, and privacy can be critical quality requirements of AI systems.**

Keywords: AI system development, AI ethical issues, AI ethical guidelines, quality requirements.

1 Introduction

The utilization of AI technologies has unlocked significant social benefits [30]. However, the black box nature of AI technologies has raised several ethical questions among its stakeholders concerning safety, privacy, security, and transparency of AI systems [28,34]. Questions about value priorities and minimizing value trade-offs in designing AI systems have led to several studies on AI ethics [13,18].

Autonomous systems, such as autonomous cars and health robots, extend ethical issues into a domain where making imprecise recommendations could impact human lives [4,33]. To develop a responsible AI, human and ethical values need to be embodied in the system design, rather than considering them merely part of an obligatory checklist [13].

Ethical issues, such as bias, diversity, and privacy preferences need to be considered during the software engineering (SE) process [3,24]. Aydemir and Dalpiaz [3] highlighted the importance of analyzing ethical issues from the very beginning of SE process, that is, right from the requirements definition. To develop ethical AI, there are over 80 ethical AI guidelines documents and standards [23]. However, studies on applying ethical practices when developing AI are lacking.

The goal of this study was to explore what kind of ethical guidelines companies have for solving potential ethical issues and developing AI systems. First, we performed a literature review to identify the potential ethical issues of AI systems and compared the AI ethical guidelines by three expert groups. Next, we conducted the case study in three Finnish companies and analyzed their ethical guidelines for developing AI systems. The participating companies were from the retail, banking, and software consultancy domains. In this paper, we use "AI systems" to refer to intelligent systems, AI solutions, AI applications, AI services, and AI products.

This paper is organized as follows. Section 2 describes the related work focusing on the ethical issues and ethical guidelines of AI systems. In Section 3, we present the research method used in this study. Section 4 describes the results of the analysis of the case companies' AI ethical guidelines. In Section 5, we discuss how these ethical guidelines can be used during the development of AI systems. Finally, we draw conclusions based on the results of the study and suggest future research directions.

2 Related Work

First, this section summarizes a set of ethical issues of AI identified in the literature, and then it provides an overview of AI ethical guidelines defined by three expert groups.

2.1 Ethical Issues of AI

Organizations are embracing many new digital technologies, such as AI and machine learning. These developments, however, raise new ethical issues that impact their users [34]. The common ethical issues of AI are: autonomy [18,24,27,31], anonymity [14], fairness [24,27,31], privacy [18,24,31], safety [2,5,24,29,31], security [24,27,29,31], transparency [17,24,31], and trust [17,24,29].

Ethical issues related to **autonomy**, **anonymity**, and **privacy** are interrelated. For many AI systems, collecting volumes of personal data from different sources is the cornerstone of their operation. The potential misuse of data could lead to major privacy threats [24]. In some cases, privacy issues of AI are bound to produce both individual- and society-level impacts [18,31]. Likewise, ethical concerns related to autonomy include 1) the extent to which technologies can influence humans [31], 2) the level of consideration of personal autonomy, such as the surveillance of workers [27], 3) the capacity of individuals to make their own choices [24], and 4) the possibilities of man out-of-the-loop operations and their impact [31].

Data exclusion and discrimination by AI systems lead to the ethical issue of **fairness,** which is also related to public values, such as human dignity and justice [24,31]. AI technology is expected to cater to everyone without any discrimination with respect to gender, age, accessibility, etc.; it has "the moral obligation to act on fair adjunction between conflicting claims" [24]. One ethical issue AI systems run into, however, is producing unfair outcomes because of data bias, exclusion, or discrimination [31]. For example, profiling users based on their data could lead to unfair outcomes for some user groups [31].

With the influx of new technologies and smart devices, **security** issues are increasingly complex [31]. For instance, hacking a coffee machine in someone's home can help the hackers to open their front door. Similarly, AI **safety** triggers ethical and societal issues. When designing technologies such as autonomous vehicles [11], virtual-reality applications [31], and tracking technologies, such as using GPS to track elderly patients in everyday settings [29,31], the safety of users is crucial. In the AI system development, compromising users' safety is an ethical issue [29].

The lack of **transparency** in the data used for AI decision-making and the neglect of transparency rights when developing AI systems also create ethical issues [17,31]. The lack of visibility or simply the black box nature of these AI systems leaves many users confused about certain suggestions made by the AI devices. Transparency represents the significance of the stakeholders' "right to know" [17]. This enables the transparency and **trust** to go hand in hand [17,24]. Explanations of AI decisions are key to building trustworthy AI systems [17,29]. Although the lack of explanation does not stop all users from relying on the systems, explanation plays a major role in building trust [29]. Moreover, incorporating adequate measures with respect to ethical issues related to security, privacy, autonomy, and transparency during AI system development helps acquire users' trust [24].

2.2 Ethical Guidelines for Practitioners to Develop AI Systems

More than 80 public and private AI ethical guidelines documents exist [23]. In this literature review, we focus on recently published ethical guidelines by three established expert groups: European Commission's (EU) *Ethical guidelines for trustworthy AI* [19], Institute of Electrical and Electronic Engineers' (IEEE) *Ethically Aligned Design* [20], and Software and Information Industry Association's (SIIA) *Ethical Principles for Artificial Intelligence and Data Analytics* [32]. These high-level guideline documents targeted for all types of organizations, including both the public and private sectors. Table 1 summarizes the ethical guidelines of the three expert groups.

Table 1. Overview of the ethical guidelines of the three expert groups. E – ethical guideline defined explicitly in the document; I - ethical guideline mentioned implicitly in the document

Ethical guidelines	EU [19]	IEEE [20]	SIIA [32]
Transparency	E	E	E
Well-being	E	E	E
Awareness of misuse and harm	E	E	E
Explainability	E	I	E
Accountability	E	E	
Autonomy	E	E	
Fairness	E	I	I
Responsibility	I	I	E
Safety	E	I	
Privacy	E	I	
Purpose of AI system	I	E	
Competence	I	E	

Transparency, autonomy, fairness, safety, and privacy ethical guidelines are related to the ethical issues of AI discussed in the previous section. The ethical guidelines on well-being involve both societal and environmental well-being. Encompassing sustainability and monitoring social impact are key phenomena in the well-being of users [19, 20]. Examining the risks of misuse also protects and prevents AI systems from causing harm. In addition, these expert groups defined explainability, accountability, and responsibility guidelines to consider potential ethical concerns in AI system development. The ethical guidelines on purpose of AI system was defined to clarify the effectiveness and impact of AI systems [20]. Also, the necessity of having teams with diverse skills and competence to operate of AI systems effectively is highlighted in their competence guidelines [19, 20].

Floridi et al. [21] proposed an ethical framework that synthesized the opportunities, risks, recommendations, and principles to develop "Good AI Society". The framework's five ethical principles were beneficence, non-maleficence, autonomy, justice, and explicability [21]. Apart from such AI ethical guidelines documents, other tools are available, such as data ethics canvas [26] and ethics matrix [25], to aid practitioners in identifying ethical issues of AI.

3 Research Method

3.1 Research Process

The research question of this study was **what kind of ethical guidelines companies have for solving potential ethical issues and developing AI systems**. We conducted this study using qualitative methods [7] to understand companies' current situations relating on AI ethics and ethical guidelines. As a first step, we defined the objectives and questions of our interviews. Then, the interview questions were validated by senior researchers and improved based on their feedback. Afterward, we conducted two pilot interviews and three actual interviews. Finally, we analyzed the interview data and ethical-guideline documents. The data collection and analysis are described in more detail in the following sections.

The unit of analysis in our empirical research process was a company that had already defined ethical guidelines for developing AI systems. We sought companies that represented different application domains for a multiple cases study, a method that Yin [35] has recommended for exploring a relatively new issue, such as ethical AI. We selected three case companies that were recommended by an AI expert who was knowledgeable about which organizations have already invested in ethical guidelines for AI systems.

3.2 Case Companies

Company A is a software consultancy company. It designs and delivers new digital services and products and has a data science team with around 20 people who are

mainly data scientists and coders. Moreover, the company had organized AI ethics coaching for its employees and customers.

Company B is a Finnish retail company involved in the car, food, and building trades. Its AI team comprises of 25 people with different skills and capabilities. At the time of the interviews (2019), it had discussed AI ethics internally for a few years.

Company C is one of the largest financial service providers in Finland with millions of customers. It provides banking and insurance services. It had formed data science teams of 15 persons and started working on ethical AI since 2017. The company has also organized ethics training for its data scientists. Table 2 summarizes each company's number of employees and application domain.

Table 2. Overview of the case companies

Company	ID for interviewee	Number of employees	Application domain
A	P1	500	Software Consultancy
B	P2	~22 500	Retail
C	P3	~12 300	Banking

3.3 Data Collection

This paper's first author designed the interview questions using Boyce and Neale's [6] guidelines, which were then improved based on the feedback received from the three senior researchers, who are the other authors of the paper. We also tested the questions with two practitioners in order to check their feasibility and understandability. We did not make any changes to the interview questions after the pilot testing. However, these two pilot interviews were not included in this study because the first company did not have concrete ethical guidelines for the development of AI systems. The second company of the pilot interviews had recently started to develop a technology platform for helping organizations deliver explainable AI services. A representative at this second company recommended the companies and interviewees for this case study.

We organized our interview questions into two parts: the organizational context of the interviewees, and the ethics and ethical guidelines of their companies. The first two authors of the paper interviewed one person from each case company. The interviewees were deeply knowledgeable about the ethical guidelines of their companies and had closely collaborated with professionals of various backgrounds such as data scientists, designers, and developers.

We conducted the interviews in late 2018 and early 2019. The lengths of the interviews varied from 60 to 80 minutes. The interviewees agreed to audio recordings of the interviews, and we assured anonymization in results. After the interviews, the interviewees shared the ethical guideline documents of their respective companies.

Company A had designed a data ethics canvas to capture possible ethical issues and the actions needed to mitigate them. The canvas consisted of 40 questions categorized into five sections. The company also shared its AI ethics training document, which contained a set of ethical guidelines. Company B's ethical principles document had five

guideline categories and ten guidelines in total. Similarly, Company C had defined five ethical guideline categories in its document.

3.4 Data Analysis

The next step was to analyze the data we collected from the documents and interviews. The audio files from the interviews were transcribed, and the notes taken during the interviews were attached to the transcriptions. We applied the Eisenhardt research method [15] in the data analysis. We performed a within case analysis with the codes categorized based on the ethical issues of AI and created a case description report for each company. Thereafter, based on the case description written for each company, the cross-case analysis method was employed, during which evidence data from one case description was compared to the other cases to report the results [15]. Fig. 1 gives an overview of the data analysis process employed in this study.

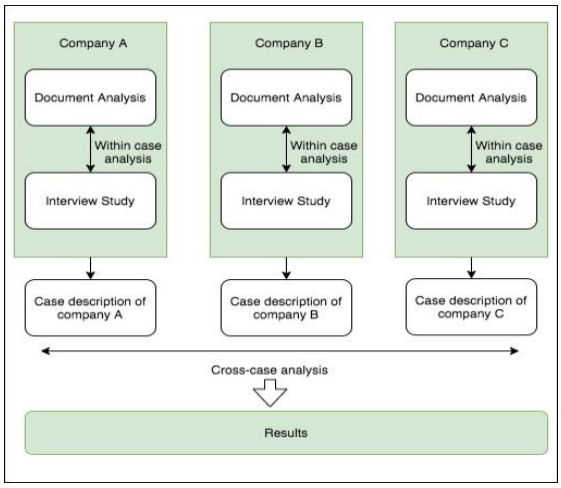

Fig. 1. Overview of data analysis

We used Charmaz's [8] grounded theory method on coding and code-comparison practices to the qualitative interview data for the purpose of analysis only. To elaborate the coding process, the first two authors of this paper first read the documents and transcripts separately. Then, they inductively applied descriptive labels (i.e., "codes") to segments of text in each document and transcript. These high-level categories of codes were formed based on the ethical issues of AI. Next, the first two researchers iteratively compared and discussed the codes and categorizations. Missing codes were added, and ambiguous codes were resolved during the iterations. In addition to the codes related to ethical issues of AI, the authors also identified codes related to ethical practices for solving ethical issues at the case companies.

4 Results

The first subsection below describes what kind of ethical guidelines the case companies have for solving potential ethical issues of AI. The second subsection describes a small set of practices that can support the use of the ethical guidelines during AI system development.

4.1 Ethical Guidelines Focusing on Potential Ethical Issues of AI

Table 3 summarizes the ethical guidelines the case companies defined for developing AI systems, which have been categorized according to potential ethical issues.

Table 3. Overview of ethical guidelines and ethical issues of AI

Ethical Issue of AI	Ethical Guidelines of AI
Accountability (A, B, C)	Decide who will be responsible and contacted if the system is seen causing harm, and decide who in the project resolves ethical issues (A) Use responsibility and security to direct the collection and utilization of the data and the creation of AI solutions and algorithms (B) Be responsible for the AI systems and the decisions they make (B) Define owners for the principles guiding the company's operations and for the algorithms the company has developed. Ensure the ethics of AI throughout its life cycle (C) Use data and AI responsibly for the good of the customers (C)
Explainability (A)	Design and build in explainability from the beginning, where it is paramount to provide justification for the outcomes of the system Do not use manipulative design – instead design for understanding
Fairness (A, B)	Avoid creating or reinforcing bias that can lead to unfair outcomes (A) Use diverse/inclusive training and test data to ensure fairness and inclusivity (A) Respect human rights and the use of AI systems must not lead to discrimination (B)
Privacy (A, B, C)	Collect, store, and use personal data safely and default to high privacy (A) Anonymize data as much as possible (A) Use responsibility and security direct the collection and utilization of the data and the creation of AI solutions and algorithms (B) Protect the data and privacy of the customers (B) Guarantee privacy and personal data protection for the individuals represented in the data used, in accordance with our data protection principles (C)
Transparency (A, B, C)	Prioritize transparency in the system and strive to increase trust in all of them (A) Go for maximum transparency and openness in the system whenever possible (A) Inform transparently to customers of where and how the company utilizes the data they have provided (B) Act openly in relation with customers, partners, and stakeholders, ensuring sufficient transparency for the evaluation of the AI the company has developed (C) Discuss the company's use of AI openly and subject the work to public scrutiny (C)

Accountability: To ensure accountability, the ethical guidelines of Company A suggest assigning a responsible person to each project to identify any possible unintended

consequences of the AI system. Then, the responsible person is in-charge of handling and controlling the harm, ethical issues, and consequences of the AI system.

At Company B, the ethical guidelines for accountability highlighted the importance of responsibility in AI systems. The guideline states that the company is responsible for its AI systems and its decisions. Moreover, the company's data collection and utilization when creating algorithms is driven by acting responsibly. The company is also responsible for AI developed outside the company and used in their systems (P2). At the time of the interview, the data scientists of the company were attempting to open AI black box and trying to understand its underlying decision mechanism, as the company is accountable and responsible for all AI services they provide to customers.

The interviewee at Company B (P2) mentioned that AI ethics is part of the company's sustainable strategy. Therefore, the company and its partners use AI systems to build a better society and a better world. The goal of their corporate responsibility is to improve social welfare, and P2 asserted that the company held itself responsible for creating values for consumers and society.

Company C's ethical guidelines on accountability required defining owners who will be responsible for guiding the company's operations and the algorithms developed by the company. They also aim to keep track of the ethics of AI system throughout its life cycle. The company also focuses on using data and AI responsibly with people-first approach, and its ethical guidelines underscore that the company's choices when applying AI should always be responsible. P3 expressed a high-level viewpoint on corporate responsibility that comprised company responsibility, societal responsibility, and customers' responsibility. Currently, Company C prioritizes societal responsibility, which P3 supported: "*We have to act sustainably in society and take good care of the society…. People expect the bank to do right things."*

The accountability-related ethical guidelines in our case companies associate accountability with developing responsible AI. Companies A and C mentioned defining a person or owner who would be responsible for identifying the unintended consequences and ethical issues of AI. However, Company B portrayed its own responsibility for the decisions of AI systems. In particular, Companies B and C emphasized developing responsible AI as a part of their societal responsibility.

Explainability: The ethical guidelines of Company A comprehended the importance of designing explainability from the beginning of AI system development, which the adoption of the General Data Protection Regulation (GDPR) in the EU pushed them to do (P1). To guide the development teams with respect to explainability, the following basic questions were defined in the data ethics canvas: 1) Is explainability needed? 2) Who needs it? and 3) How much are you able to explain the system? The interviewee also mentioned that considering the explainability from the beginning helps in choosing the right algorithmic model for the AI system.

According to the interviewee P1, explainability is also about understanding. The biggest problem with understandability is that it is difficult to explain an AI system's results, even for the people who built it. Furthermore, P1 said "*It is important to design easy systems to use; that is, whoever is using the system must also know what happens under the hood*". The problem, however, is that explanations provided by the system are difficult for the users to understand, so they do not add any value (P1).

According to the interviewee at Company B (P2), it is an important, necessary skill for the data scientists to explain the AI systems they create. They should understand what is inside the AI black box, that is, what kind of data there are and how the AI system uses them (P2). Company B highlights explainability as a key part of building responsible AI.

To summarize, the companies represented their ethical guidelines on explainability as a branch of transparency. The interviewees at Companies A and B indicated that explainability was doable by understanding the AI system, and they mentioned a lack of skills in explaining the AI as a hurdle for implementing explainability guidelines.

Fairness: Company A's ethical guidelines on fairness aim to remove bias from its data. Fairness in AI systems is closely related to the inclusion and exclusion of the data, which were described as either overrepresentation or underrepresentation or as bias and gaps in the data. Their ethical guidelines require representation of diverse set of people as possible during AI system design. Likewise, our interviewee (P1) pointed to inclusion as a hot topic in the industry.

One of Company A's ethical guidelines on fairness advocates for avoiding creating or reinforcing bias that produces unfair results. According to P1, "*There is no right way to erase or handle bias, because it is always context sensitive."* The interviewee also mentioned that each team analyzes and decides how to make results less biased, based on their particular project.

The interviewee at Company C pointed to fairness in relation to customers' trust. P3 highlighted it by saying "*To gain the trust of the customer, they have to trust that the bank is doing fair, right and so forth*." P3 added that acting fairly is vital to ethical banking. To elaborate, the key point related to acting fairly is that customers can trust that analysis and decisions that concern them are made correctly by following the correct process (P3). According to the interviewee, fairness concerns not only individuals but also society, so it is ethical to work openly, friendly, and equally to do right things that people expect from the company (P3).

In summary, the companies' fairness ethical guidelines aimed to eliminate bias and discrimination from their data. Company A showcased the need for inclusion in data to achieve fair outcomes. Likewise, Companies B and C described how their AI systems which do not discriminate, helped protecting human rights and contribute to common good to the whole society.

Privacy: The privacy ethical guidelines of Company A focused mainly on personal data protection. The GDPR was a key reason for prioritizing data privacy ethical guidelines. Data anonymization is one way to ensure privacy in AI systems (P1). To handle personal data safely, it is essential to reveal to the users on how their personal data is utilized by the AI system. Their ethical guidelines also emphasized collecting minimal sensitive data from users.

From the Company A interviewee's perspective, privacy and safety are important for any project that deals with data. Furthermore, P1 highlighted that "*AI projects introduce new vulnerabilities, so that we have to be extra careful with data security and privacy*." Privacy, security and safety were mentioned as tightly coupled concepts by our interviewee. P1 also commented about users' right to check their data privacy in compliance with the GDPR. It is important to make it as easy as possible for users to exercise their rights to data privacy (P1).

The ethical guidelines of Company B on privacy highlighted data protection and the privacy of customers. The interviewee explained that "*it is really important to understand privacy that we are using customer data heavily*." so, the company needs to ensure what is their rights with respect to customer data (P2). The interviewee also pointed out the influence of the GDPR when creating the data platforms. The GDPR affects how the privacy and security of customer data is handled (P2).

A key concept mentioned by the interviewee was about permission. That is, permission from users to use their data and to combine them with data from other sources. Customers have a right to their data, and they can decline permission to use them (P2). According to P2, the company must be aware of why it is collecting the data, and what kind of permissions it has from its customers.

Company C has ethical guidelines on privacy protection that aim to guarantee privacy and personal data protection to the individuals represented in their data. P3 emphasized that "*banking and financial services are as sensitive as health issues from an ethical point of view and privacy point of view*." The company also provides data protection and ethics courses which mainly focus on privacy, privacy issues, privacy regulations, data security, and ways to use data in compliance with the GDPR. Because the company complies with privacy laws in the EU (P3), one important question in its training material is about trust and privacy: "*Can customers trust that the information that concerns them stays private*?"

According to the interviewee, the company is trying to protect its rights to utilize the data it collects, especially according to the latest version of privacy regulations. P3 also added that there are many constraints on what the company can do with data, and that the company is cautious about using data. The foremost thing Company C does is to look at it from customers' perspective. This is because, according to the privacy laws, the customers have extensive rights to know and control what is done with their data. In addition, the company has proposed employing data anonymization so that they can observe customers' behavior without pinpointing individuals.

In summary, the case companies' ethical guidelines on privacy illustrated the contribution of the GDPR. Data privacy, personal data protection, data security, and data anonymization are the foundation of defining ethical guidelines in the case companies. In the same way, the companies exhibited their compliance with privacy laws and regulations. In addition, Companies B and C highlighted the importance of privacy ethical guidelines in order to ensure customers' trust.

Transparency: The ethical guidelines of Company A highlighted the importance of transparency, one of its values - in addition to trust, caring, and continuous learning (P1). Therefore, its ethical guidelines recommended that development teams to prioritize maximum transparency and openness in their AI systems whenever possible. However, maximizing transparency is not straightforward. The interviewee mentioned that there can also be several reasons, such as business secrets and privacy, that development teams have to manage with transparency.

According to the interviewee, transparency starts with technical transparency, which means exposing the data and algorithms of AI systems. The interviewee's definition of transparency also included explainability: "It *doesn't need to be lengthy explanation, but it needs to be understandable, otherwise it is not transparency*." P1 also explained that transparency is doable despite problems such as lack of data (or inclusive data) and data plus algorithm opacity.

The ethical guidelines of Company B feature transparency, along with the responsibility and security ethical guidelines. The guidelines focus on informing the customers on how the company utilizes their data. Being open and transparent is the goal of Company B. The interviewee mentioned privacy as a prerequisite for transparency. That is, if the customer has given permission to use their data, then it is easy for the company to implement transparency (P2).

At the time of our interview, the company was discussing how to be more transparent about the AI behind its recommendations and results. The interviewee mentioned that the company needs to improve transparency and the related technical things. One of the proposals to ensure transparency is to open algorithms (P2).

Company C's ethical guidelines on transparency are associated with openness. The guideline highlights that it is crucial to act openly with customers, partners, and stakeholders. This means discussing the use of AI openly and opening the work of the company to review and public scrutiny.

To summarize, all the case companies aimed to work openly and transparently. They see opening their data and algorithms as a starting point to ensure transparency. It is also critical to ensure that the data and analytics are correct. Being transparent with customers gains their trust, which the companies portray as a key value.

4.2 Practices Supporting the Use of the Ethical Guidelines of AI

This section describes the following three practices that can support the use of ethical guidelines in the development of AI systems:

- Defining the purpose of the AI system
- Analyzing the impacts of the AI system
- Using multi-disciplinary teams

Defining the Purpose of the AI System. All our case companies focused on defining the purpose of the AI system clearly. Company A emphasized that it is important to ensure the system that the company designs and builds have a clear purpose so that the system will be trusted to behave adhering to its purpose. In addition, the key factors like what is the expected result of the system and how it will be measured need to defined.

In their ethical guidelines, Company B highlighted their objective of creating solutions that are useful for customers. The company focused on the practice of placing the needs of the customers first and creating an AI system that is useful to customers. Furthermore, the company only uses their customer data for purposes for which their customers have given permission. Customers should control the data that the company can use (P2). Altogether, the company aims to achieve the best customer experience.

Company C's viewpoint is that, when defining the purpose, the objectives guiding the use of AI need to be determined clearly. This objective can be refined if necessary, based on changed data, technical possibilities, and the work environment. In addition, the company also considers things from customers' perspectives, which is also the focus of Company B.

Analyzing the Impact of the AI System. Company A's analysis of the impact of its AI system is supported by the following two guidelines:

- Respect and be mindful of the impact on people affected by the system
- Consider the impact of the system beyond its users, and consider the positive and negative consequences of the system

Company B mentioned that when analyzing the impact, the data-driven insight enables the company to provide added values to the everyday lives of its customers. This adheres to their guideline of placing customers needs' first. Company C's approach to analyzing its impact is carefully studying the effects of their choices on the company, customers, and the society.

Using Multi-Disciplinary Teams. Companies A and B highlighted the importance of multi-disciplinary teams when developing AI systems. According to one interviewee (P1), the key thing for Company A is to make sure that development teams are truly multi-disciplinary. The reason to have designers, social scientists, and domain experts in addition to data scientists is to ensure that all the relevant viewpoints are taken into account and explanations provided by AI systems are understandable to people (P1). The interviewee from Company B (P2) explained that they had "*different kinds of backgrounds in all teams so that they can really give new kind of value to the team"*

5 Discussion

5.1 Ethical Guidelines of AI for the development of AI systems

In this section, we first compare the ethical guidelines of the case companies with the ethical issues reported in the existing literature and the ethical guidelines defined by the three expert groups. Then, we discuss how organizations can use their ethical guidelines to convert potential ethical issues into quality requirements for AI systems.

The analysis of the ethical guidelines of the case companies revealed that their guidelines focused on solving the potential ethical issues of AI systems, such as accountability, explainability, fairness, privacy, and transparency. When comparing these results with the ethical issues of AI systems that we identified in the existing literature, we observed several similarities. The ethical guidelines of the case companies covered the following ethical issues reported in the literature: anonymity [14], fairness [24,27,31], privacy [18,24,31], security [24,27,29,31], transparency [17,24,31], and trust [17,24,29].

The ethical guidelines of the case companies on accountability, explainability, fairness, privacy, and transparency also corresponded to the ethical guidelines recommended by the three expert groups [19, 20, 36]. Also, the case companies' guidelines on accountability related to the well-being and responsibility guidelines specified by the expert groups. Furthermore, both the ethical guidelines of the expert groups [19, 20] and the interviewees of our study recommend defining the purpose of AI system.

The existing literature already offers many ethical guidelines. However, it can be difficult for companies to know which one of these guidelines they should select to be

used in their AI development. In addition, existing guidelines can be too comprehensive and broad for development teams to apply, especially in agile projects. The case companies in this study have defined company-specific guidelines that are compact. These guidelines bring up a set of potential ethical issues that the AI development teams can focus on.

Based on the results of this case study, we propose that organizations define a set of ethical guidelines for handling potential ethical issues during the development of AI. We also recommend that organizations use their ethical guidelines to identify and prioritize the critical quality requirements of the AI systems.

All our case companies had defined guidelines on transparency. The transparency guidelines of the case companies emphasized the need for being open to the use of AI and informing customers what data is used and how it is used in the AI system. The companies also highlighted prioritizing transparency when developing AI systems to gain users' trust. Cysneiros et al. [11, 12] have also proposed transparency as a key quality requirement of AI systems, and Horkoff's [22] study on the quality requirements of machine learning revealed transparency as a key quality.

The findings of our study indicate that the explainability guidelines complement and support the transparency guidelines. The companies' ethical guidelines related to explainability pointed out the importance of justifying the outcomes of AI systems and building in explainability from the beginning. Similarly, Chazette and Schneider [9, 10] highlighted explanation as an option to mitigate the lack of transparency of a system and portrayed explainability as an emerging quality requirement of systems.

The fairness guidelines of our case companies focused on two goals: avoiding bias and respecting human rights. First, it is critical to avoid discrimination, which could lead to unfair results. In addition, one of the solutions to tackle fairness issues is to adopt inclusivity in data used to train AI. The literature [9, 10, 22] indicates that fairness is a quality requirement.

The privacy guidelines of the case companies covered anonymity, security, and data protection aspects. All the interviewees especially pointed out the importance of ensuring the privacy of personal data because of the GDPR. Cysneiros et al. [11, 12] have also highlighted privacy as a potential quality requirement of AI systems.

The accountability guidelines of the case companies recommended assigning owners for the AI algorithms the company has developed. In addition, the guidelines suggested naming responsible persons who will be contacted when the AI system is causing harm and who will resolve ethical issues. Based on these accountability guidelines, we deduce that the companies perceive accountability as a characteristic of professional conduct rather than a system characteristic i.e. quality requirement of the AI system.

In addition to the ethical guidelines of AI systems, the case companies recommended practices that can support the application of these guidelines to development projects. For example, two of the case companies recommended using multi-disciplinary teams when considering different viewpoints on ethical issues. The interviewees also highlighted the importance of defining a clear purpose for AI systems by focusing especially on customer needs. The ethical guidelines of one of the companies even emphasized the necessity of considering the impacts of the system beyond the user, and consider any positive and negative consequences the system might have for society. The

analysis of potential negative consequences relates closely to misuse cases, which is a requirements engineering practice to be used for identifying use cases with hostile intent and negative scenarios [1].

5.2 Limitations of this Study

Generalizability. One important question related to case studies is: to what extent can the results of the study be considered representative of a broader range of organizations? We acknowledge that the number of our case companies was rather low. They represent, however, three different application domains. These companies were also recommended by an AI expert who is knowledgeable about which organizations have already invested in ethical guidelines of AI systems. According to that expert, these case companies are forerunners. Therefore, we believe that other organizations can learn from the ethical guidelines described in this paper.

Reliability. The data collection and analysis processes also left room for researcher bias. To avoid misinterpretation and bias in the interview questions, three senior researchers reviewed the interview questions, which we also tested with two practitioners. In addition, two researchers conducted the interviews together. The first author asked the interview questions, and the second author asked follow-up questions based on the interviewees' answers. Likewise, researcher triangulation was used in order to avoid researcher bias in the data analysis.

Construct validity. The number of interviewees in our case study was also low, as we interviewed only one person from each company. To handle this limitation, the triangulation of data sources was used. In order to gain more knowledge about the ethical guidelines, we were also able to analyze the ethical guideline documents of all the case companies. Furthermore, the study was done at company level. Therefore, we are not able to report how the ethical guidelines are applied in AI projects. We believe that this kind of company-level study provides valuable knowledge about the commitment to and support of companies for the ethical development of AI systems. We also understand that ethical guidelines are valuable when they are used in development projects.

6 Conclusions

The goal of this study was to analyze what kind of ethical guidelines companies have defined for solving potential ethical issues of AI and developing AI systems. The ethical guidelines of the case companies focused on solving ethical issues, such as accountability, explainability, fairness, privacy, and transparency. Based on the results of this study, we suggest that organizations develop and use their ethical guidelines to identify and prioritize critical quality requirements of AI. The results also indicate that transparency, explainability, fairness, and privacy can be critical quality requirements of AI systems. In addition, defining the purposes of their AI systems clearly and analyzing impacts of their AI systems can assist multi-disciplinary development teams in solving ethical issues during the development of AI systems.

One important direction in our future research is to conduct case studies and investigate how ethical guidelines are used in AI projects. Our goal is to compare the ways development teams have applied the ethical guidelines of their organizations, the challenges they have faced, and their positive experiences using those ethical guidelines. Our particular research interest is to understand how critical quality requirements of AI systems are defined and tested in practice.

References

1. I. Alexander, "Misuse Cases: Use Cases with Hostile Intent", *IEEE Software,* pp. 58-68, 2003.
2. T. Arnold and M. Scheutz, "The "big red-button" is too late: An alternative model for ethical evaluation of AI systems," *Ethics Inf. Technol.,* vol. 20, no. 1, pp. 59-69, 2018.
3. F. B. Aydemir and F. Dalpiaz, "A roadmap for ethics-aware software engineering," *ACM/IEEE International Workshop on Software Fairness (FairWare'18),* pp. 15-21, 2018.
4. V. Bonnemains, S. Claire, and C. Tessier, "Embedded ethics: Some technical and ethical challenges," *Ethics Inf. Technol.,* vol. 20, no. 1, pp. 41-58, 2018.
5. N. Bostrom and E. Yudkowsky, "The ethics of artificial intelligence," in *Cambridge Handbook of Artificial Intelligence,* K. Frankish and W. M. Ramsay, Eds., Cambridge University Press, 2011, pp. 316-334.
6. C. Boyce and P. Neale, "Conducting in-depth interviews: A guide for designing and conducting in-depth interviews," *Evaluation,* vol. 2, no. May, pp. 1-16, 2006.
7. C. Cassell and G. Symon, *Essential Guide to Qualitative Methods in Organizational Research.* SAGE Publications, 2012.
8. K. Charmaz, *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis.* SAGE Publications, 2006.
9. L. Chazette, O. Karras, and K. Schneider, "Do end-users want explanations? Analyzing the role of explainability as an emerging aspect of non-functional requirements", *RE2019,* pp. 223-233, 2018.
10. L. Chazette and K. Schneider, "Explainability as non-functional requirement: Challenges and recommendations," *Requirements Engineering,* 2020.
11. L. M. Cysneiros, M. A. Raffi, and J. C. S. P. Leite, "Software transparency as a key requirement for self-driving cars," *RE2018,* pp. 382-387, 2018.
12. L.M. Cysneiros and J.C.S.P. Leite, "Non-Functional Requirements Orienting the Development of Socially Responsible Software," *BPMDS 2020,* pp.335-342, 2020.
13. V. Dignum, "Ethics in artificial intelligence: Introduction to the special issue," *Ethics Inf. Technol.,* vol. 20, no. 1, pp. 1-3, 2018.
14. T. Doyle and J. Veranas, "Public anonymity and the connected world," *Ethics Inf. Technol.,* vol. 16, pp. 207-218, 2014.
15. K. M. Eisenhardt, "Building theories from case study research," *Acad. Manage. Rev.,* vol. 14, no. 4, pp. 532-550, 1989.
16. K. M. Eisenhardt and M. E. Graebner, "Theory building from cases: Opportunities and challenges," *Acad. Manage. J.,* vol. 50, no. 1, pp. 25-32, 2007.
17. J. Elia, "Transparency rights, technology, and trust," *Ethics Inf. Technol.,* vol. 11, pp. 145-153, 2009.
18. A. Etzioni and O. Etzioni, "AI assisted ethics", *Ethics Inf. Technol.,* vol. 18, no. 2, pp. 149-156, 2016.

19. European Commission, *Ethics Guidelines for Trustworthy AI,* Available: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines [Accessed January 24, 2020].
20. IEEE, *Ethically Aligned Design,* First Edition, Available: https://ethicsinaction.ieee.org/. [Accessed November 24, 2019].
21. L. Floridi, J. Cowls, M. Beltramatti et al., "AI4people: An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," *Minds and Machines,* vol. 28, pp. 689-707, 2018.
22. J. Horkoff, "Non-functional Requirements for Machine Learning: Challenges and New Directions," *International Requirements Engineering Conference,* pp.386-391, 2019.
23. A. Jobin, M. Lenca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nature Machine Intelligence*, Vol. 1, pp. 389-399, 2019.
24. S. Jones, S. Hara, and J. C. Augusto, "eFRIEND: An ethical framework for intelligent environment development," *Ethics Inf. Technol.,* vol. 17, pp. 11-25, 2015.
25. B. Mepham, M. Kaiser, E. Thorstensen et al., "Ethical Matrix Manual", 2006.
26. Open data institute, *Data Ethics Canvas,* Available: https://theodi.org/wp-content/uploads/2019/07/ODI-Data-Ethics-Canvas-2019-05.pdf [Accessed June 24, 2020].
27. E. Palm, "Securing privacy at work: The importance of contextualized consent", *Ethics Inf. Technol.,* vol. 11, pp. 233-241, 2009.
28. A.R. Peslak, "Improving software quality: An ethics-based approach", *SIGMS'04*, pp. 144-149, 2004.
29. W. Pieters, "Explanation and trust: What to tell the user in security and AI?" *Ethics Inf. Technol.,* vol. 13, pp. 53-64, 2011.
30. I. Rahwan, "Society-in-the-loop: Programming the algorithmic social contract," *Ethics Inf. Technol.,* vol. 20, no. 1, pp. 5-14, 2018.
31. L. Royakkers, J. Timmer, L. Kool, and R. V. Est, "Societal and ethical issues of digitization," *Ethics Inf. Technol.,* vol. 20, pp. 127-142, 2018.
32. SIIA (Software and Information Industry Association), *Ethical Principles for Artificial Intelligence and Data Analytics*, pp. 1-25, 2017.
33. Stanford University, One Hundred Year Study on Artificial Intelligence (AI100), "Artificial Intelligence and Life in 2030*," Stanford University.* Available: https://ai100.stanford.edu/ . [Accessed December 15, 2019].
34. P. Vampley, R. Dazeley, C. Foale, et al., "Human-aligned artificial intelligence in a multi objective problem," *Ethics Inf. Technol.,* vol. 20, no. 1, pp. 27-40, 2018.
35. R. K. Yin, Case Study Research Design and Methods. Sage, 2013.