What Is the Cost of AI Ethics? Initial Conceptual Framework and Empirical Insights

Kai-Kristian Kemell[1(✉)] and Ville Vakkuri[2]

[1] Department of Computer Science, University of Helsinki, Pietari Kalmin Katu 5, 00560 Helsinki, Finland
kai-kristian.kemell@helsinki.fi

[2] School of Marketing and Communication, University of Vaasa, Wolffintie 32 FI-65200 Vaasa PL 700, 65101 Vaasa, Finland
ville.vakkuri@uwasa.fi

Abstract. AI ethics has become a common topic of discussion in both media and academic research. Companies are also increasingly interested in AI ethics, although there are still various challenges associated with bringing AI ethics into practice. Especially from a business point of view, AI ethics remains largely unexplored. The lack of established processes and practices for implementing AI ethics is an issue in this regard as well, as resource estimation is challenging if the process is fuzzy. In this paper, we begin tackling this issue by providing initial insights into the cost of AI ethics. Building on existing literature on software quality cost estimation, we draw parallels between the past state of quality in Software Engineering (SE) and the current state of AI ethics. Empirical examples are then utilized to showcase some elements of the cost of implementing AI ethics. While this paper provides an initial look into the cost of AI ethics and useful insights from comparisons to software quality, the practice of implementing AI ethics remains nascent, and, thus, a better empirical understanding of AI ethics is required going forward.

Keywords: Ethics · Machine learning · Cost estimation · Software engineering · Artificial intelligence

1 Introduction

Despite AI ethics being increasingly discussed both on the academia and now out on the field as well, it remains of secondary importance in practice [13, 15]. While companies are becoming aware of the potential importance of AI ethics, its practical implementation is still an on-going issue. In research, this continues to manifest as a lack of empirical studies on the topic. While some companies show interest towards AI ethics and even release statements about their commitment to developing ethical software systems, little is known how this is done in practice, given the lack of empirical studies on AI ethics [13].

As little is known about the practical implementation of AI ethics, it is also difficult for companies to evaluate the resources and costs required for doing

S. Hyrynsalmi et al. (Eds.): ICSOB 2023, LNBIP 500, pp. 247–262, 2024.
https://doi.org/10.1007/978-3-031-53227-6_18

so. Indeed, especially from a business point of view, AI ethics remains an open question. While the potential benefits of implementing ethics are becoming more clear for software companies (through the potential cost of ignoring ethics, if nothing else), and few companies would go on record to say ethics is not a priority for them, the cost of AI ethics remains unclear.

Ethics encompasses the entirety of the development process, from design to operations. At different points of the process, ethics manifests in different ways in SE practice [16]. Early on, design decisions shape the system, and ethical issues can arise from major decisions such as the business logic or the very nature of the system [4]. During development, ethics includes issues from data to end-user involvement (e.g., as seen through the plethora of tools included in tool review of Morley et al. [10], and as highlighted by the ECCOLA method [16]). During operations, ethics may necessitate new metrics to monitor; there are some examples of issues in AI systems having recently been uncovered through bad publicity on social media (e.g., a chatbot giving unauthorized diet advice for users seeking help for eating disorders[1]).

Ethics is more than just minimum compliance to laws and regulations. At worst, ignoring ethical issues can lead to a system being pulled from production. Because ethics encompasses the entire development process, fixing issues stemming from poor design decisions early on can be highly costly and difficult in production. The ease of fixing issues early on in the development process is an acknowledged phenomenon in software quality [11], as well as, arguably, software development overall.

In this paper, we provide an initial look at AI ethics from the point of view of business by (1) discussing its relevance for business, and (2) discussing it from the point of the resources needed for implementing ethics. It is established in extant literature that there are still prominent gaps to be addressed in the practical implementation of AI ethics, and the business and resource point of view is one of them. We build this discussion on both existing literature and data from three empirical cases. By utilizing existing literature on software quality, we propose a high-level cost framework for ethics in SE. Then, through the example cases, we provide some initial insights into what types of activities, and thus, costs, are associated with implementing ethics in practice in SE.

While this paper is specifically motivated by *AI ethics*, this discussion is relevant for ethics in SE overall. For example, issues such as green IT are a part of AI ethics but also relevant for software organizations overall. We have chosen AI ethics as the context for this paper due to its timeliness and due to nature of the data we have collected.

The rest of this paper is structured as follows. Section 2 presents the theoretical background of the paper by discussing existing literature. In Sect. 3, we discuss the cost of (AI) ethics by building on existing literature on software quality and utilizing an existing cost framework for software quality. In Sect. 4, we provide some initial insights into the cost of (AI) ethics by utilizing past data we have originally collected for other research purposes (specifically, to develop

[1] https://www.theguardian.com/technology/2023/may/31/eating-disorder-hotline-union-ai-chatbot-harm.

the ECCOLA method [16]). In Sect. 5 we discuss the theoretical and practical implications of this paper. Section 6 concludes the paper.

2 What and Why Ethics

In Sect. 2.1, we provide a general overview of ethics in relation to SE, and more specifically AI. In Sect. 2.2, we expand on this discussion by adding a business focus.

2.1 Ethics, Ethics in SE, and AI Ethics

Ethics can be described as a philosophical field of study. In particular, ethics is the study of morality. In this paper, we discuss *applied ethics*, specifically in the context of both business ethics and ethics in SE, and more specifically, AI ethics. Applied ethics examines real-life situations, which are often unclear or debatable, in order to understand what would be the right or wrong action to take with the given set of values. E.g., why should software companies care about the environment (green IT)? Additionally, applied ethics can be thought of as 'ethics as practice' [1,18], examples of which are guidelines and codes of conduct in SE or AI ethics.

The current discussion on AI ethics stems from the tradition of computer ethics where ethical discussion includes the ethics of system development and use, among other topics (see, e.g., [7]). Over the decades, this discussion has included topics such as piracy, green IT, cybersecurity, automatization and, more recently, AI ethics. The current discussion on AI ethics also draws from the various past discussions on ethics in SE, including topics such as business and the societal impacts of IT.

AI ethics is often approached through principles. Jobin et al. [8], based on their extensive review of AI ethics guidelines, outline the most commonly discussed principles: transparency, justice and fairness, non-maleficence, responsibility, and privacy. For example, fairness deals with issues related to bias and discrimination, which manifest in practice as, e.g., issues in ML system outputs and training data. However, bringing these principles into practice remains an on-going challenge in the area, as the guidelines seem to not have had a notable impact on industrial practice [13,17] based on empirical studies, supporting the argument of Mittelstadt [9] about the ineffectiveness of principles alone. In fact, the practical implementation of AI ethics in general remains a topical challenge in ML development, and empirical studies remain scarce [10,12]. While in addition to numerous conceptual papers, a number of papers discussing the technical implementation of, e.g., fairness (Fairness 360 etc.) exist, reported industry use cases and empirical studies are lacking.

2.2 Why (AI) Ethics?

While some organizations may still be pondering the business relevance of ethics, especially in the field of AI, ethics has gained mainstream attention. Ethical failures and potential ethical issues have been extensively discussed in mainstream

media, and companies developing ML solutions have attempted to react to this discussion by, for example, publishing their own guidelines for AI ethics (see Jobin et al. [8]) in order to signal commitment to the values within. Though good or bad publicity is a large motivator for companies to consider AI ethics, there are arguably various potential benefits for doing so. These (may) include: (1) brand equity[2] (2) consumer adoption, (3) social acceptance, (4) employee satisfaction[3] (5) investor relations (ESG reporting), (6) market entry requirements (EU GDPR; upcoming EU AI Act), (7) proactive approach to laws and regulations (e.g., upcoming EU AI Act), (8) avoiding costly changes in production [16], and (9) a systematic approach to ethics over an an ad hoc one [16].

Brand equity refers to good or bad publicity. There have been various ethical failures that have made the news, resulting in bad publicity and typically necessitating actions taken to correct the situation. Similarly, *consumer adoption* can be negatively impacted by ethical issues. Users are becoming increasingly conscious about issues such as data privacy and fairness, and tackling such topics in an ethical manner can become a selling point in ML. In a more general sense, *social acceptance* becomes important when developing particularly disruptive technologies that impact society or an organization on a larger scale, outside the scope of just their users. For example, autonomous vehicles impact traffic as a whole, rather than just their passengers ("drivers"). Aside from external stakeholders, consideration of AI ethics can also improve *employee satisfaction* in a similar manner to improving consumer opinion. If your values strongly conflict with those of your employees, it may lead to conflicts or resignations. Moreover, *investor relations* (ESG: Ecological, Social, and Governance), can also be improved via attention to AI ethics.

Market entry requirements, in this case, refers to the relevant laws and regulations. In particular, the European Union with its GDPR and the upcoming AI Act that are directly related to AI ethics, may necessitate more ethical consideration than the local region of the company. To this end, AI ethics can foster a *proactive approach to laws and regulations* can help companies adapt to the changing regulatory landscape for ML systems, with new regulations and laws constantly discussed (e.g., recently for Large Language Models (LLMs) and Generative AI) across the globe.

Ethics, like quality [11], is arguably easier to implement early on in software development, and thus, doing so can help in *avoiding costly changes in production*. Ethics encompasses the entire development process from design to production [16]. Finally, by actively pursuing AI ethics, companies are able to utilize a *systematic approach to ethics over an ad hoc one*. Even when ethics is not implemented actively, values still make their way into the product nonetheless.

[2] E.g., the capital of Finland, Helsinki, advertising their commitment to ethical AI: https://www.hel.fi/fi/uutiset/helsinki-laati-periaatteet-datan-ja-tekoalyneettiselle-kaytolle,.

[3] E.g., https://www.wired.com/story/google-brain-ai-researcher-fired-tension/,.

3 Research Framework: Cost of Quality, and the Relationship of Quality and Ethics

In this section, we present and justify our approach to discussing the cost of AI ethics. We make a comparison to quality, which, as we argue in Sect. 3.1, shares some (historical) similarities with the current state of (AI) ethics. In Sect. 3.2, based on existing literature, we present an overview of the types of costs associated with quality and, building on it, propose a similar cost structure for AI ethics.

3.1 Is Ethics Just Another Quality Feature?

We argue that we are currently seeing various parallels between the current state of AI ethics and the historical evolution of software quality assurance. Software quality was, in the past, often overlooked in favor of more immediate business concerns such as time-to-market or simple profitability. Over time, it evolved to be an integral and integrated part of the SE process. To some extent, we currently are seeing similar developments in AI ethics. Despite the discussion on the growing importance of AI ethics, it is still typically largely overlooked in practice [13,15]. Though companies are increasingly becoming aware of ethics-related issues such as fairness, the industry still seems to lack systematic frameworks and processes for implementing AI ethics, or at least it fails to utilize them.

In this paper, we approach ethics from the point of view of quality, to provide a point of comparison with an existing, well-established phenomenon in SE. While ethics is *not* simply quality and the two are not fully analogous, we nonetheless make this comparison due to the various similarities they do share:

- *Overlooked importance.* Historically, software quality was seen as a secondary objective, much like ethics currently. Its importance was acknowledged after initial failures, but making it a part of SE practice took its time. This has also been the case in AI ethics, with its importance largely now acknowledged but its practical implementation still a challenge [13].
- *Long-term consequences for software.* Both ethics and quality can result in severe negative impacts for the system(s) being developed if overlooked. Much like how bugs can render a system unusable, unforeseen ethical issues can result in an ML system being pulled from production (e.g., as was the case with the chatbot mentioned in Sect. 1).
- *Interdisciplinary nature.* Much like how Quality Assurance (QA) requires the involvement of various stakeholders other than just software developers, implementing AI ethics is also a multidisciplinary effort. While developers (and ML experts) are the ones bringing ethics into practice, the process still involves other stakeholders as well (e.g., ethics committee, experts, users...).
- *Maturing over time.* Software quality has evolved over time from simple debugging to formal QA processes and a continuous SE process (CI/CD). AI ethics seems to also be moving from a minimal regulatory and legal compliance to the development of ethical frameworks (e.g., ECCOLA [16]) and processes, although this is still on-going [14].

– *Relevance of organizational culture.* The implementation of ethics, like quality, is unlikely to succeed if it is an afterthought or a tacked-on process. AI ethics needs to become a part of organizational culture, and to this end, it needs to become a natural part of SE (e.g., as professional norms [6]).
– *Harder and costlier to implement in production.* Quality is cheaper to implement earlier on in the SE process [11]. This is also arguably the case for ethics as well. As we discuss next, ethics also encompasses system design and business logic. A system where the core (ethical) issue stems from the very goal of the system is difficult to fix in production, to say the least.

This comparison between ethics and quality is not a novel thought of ours. Existing literature has made similar observations. For example, in the literature review of Giray [5], AI ethics topics such as fairness are explicitly referred to as new types of *quality requirements* for ML systems. Indeed, it can be argued that, if quality is about assuring that the system works as intended, ethics shares the same goal on a conceptual level: assuring that the system works as intended (from the chosen ethical point of view).

However, AI ethics is not just software quality, especially not as it is conventionally understood. While some AI ethics principles such as *predictability*, which focuses on ensuring the system produces intended outputs or results reliably, are closely related to conventional software quality goals, AI ethics also encompasses system *design* and *business* in addition to software development [16]. A technically sound system that is of high quality can still be unethical. E.g., widespread AI-based surveillance using facial recognition is typically considered unethical as a concept (e.g., in the draft of upcoming AI act such systems are labelled as being of 'unacceptable risk') – and yet the use of such systems in contexts such as airport security would be considered acceptable by many, highlighting the complex nature of AI ethics.

As opposed to seeing (some parts of) ethics as quality issues, an argument could be made that it is in fact quality that is a part of ethics in SE. The ACM Code of Ethics and Professional Conduct discusses quality as a part of the job responsibilities of a software professional. It remarks that one should "strive to achieve high quality in both the processes and products of professional work" [6]. Regardless, this further provides justification for the parallels we draw between the two in the context of this paper.

3.2 The Cost of Ethics

Based on Sect. 3.1, we argue that quality offers a familiar point of reference (in SE) for initially approaching ethics from a cost point of view. According to Slaughter et al. [11], costs of quality consist, on a high level, of conformance and nonconformance. *Conformance* refers to the costs associated with developing quality products (i.e., 'doing' quality). *Nonconformance* refers to the costs resulting from failures resulting from poor quality (i.e., not 'doing' quality).

In more detail, Slaughter et al. [11] split the costs of conformance to prevention and appraisal costs. *Prevention costs* are associated with "preventing

defects before they happen", which "include the costs of training staff in design, methodologies, quality improvement meetings, and software design reviews" [11]. *Appraisal costs*, on the other hand, include "measuring, evaluating, or auditing products to assure conformance to quality standards and performance. For software, examples of appraisal costs include code inspections, testing, and software measurement activities" [11].

Costs of nonconformance are further split into internal failure costs and external failure costs by Slaughter et al. [11]. *Internal failure* costs "occur before the product is shipped to the customer. For software these include the costs of rework in programming, reinspection, and retesting." [11] *External failure* costs "arise from product failure at the customer site. For software, examples include field service and support, maintenance, liability damages, and litigation expenses." [11]

In practice, from the point of view of the SE process, they assign these costs to three phases:

1. Software Quality Investment (SQI). The initial investment of doing quality. This includes "the initial expenses for training, tools, effort, and materials required to implement the quality initiative." [11]
2. Software Quality Maintenance (SQM). Maintaining the processes set up during SQI. Ongoing expenditures "for meetings, tool upgrades, and training that are required the maintain the quality process." [11]
3. Software Quality Revenues (SQR). Any resulting revenue. Revenues derived from "projected increases in sales or estimated cost savings due to the software quality improves." [11]

Based on this, we propose a similar typology for the cost of AI ethics. We propose the following phases for AI ethics from a business point of view:

1. Ethics Investment. The initial investment for incorporating ethics into SE. This includes a wide variety of costs, such as: recruiting new experts, adopting new methods or other SE tools, modifying existing SE processes or creating new ones, more systematic project documentation, training, materials, etc.
2. Ethics Maintenance. Costs of maintaining the processes established in the first step. These include salaries of any new hired experts, meetings and other recurring tasks, etc.
3. Ethics Revenues. Any resulting revenue originating from the previous steps, such as increases in sales, brand equity, cost savings from failure prevention, etc.

Arguably, this is still a very nascent area of research. Because the practice of AI ethics overall is still poorly understood compared to software quality, the latter of which has decades of history of practice behind it by now, the associated processes are still being shaped out on the field. Thus, providing a comprehensive and detailed framework for the cost of AI ethics at this stage is not feasible. However, past simply proposing this typology on a conceptual level, we also

provide an initial look at the cost of AI ethics in practice. In Sect. 4, we focus especially on the first phase, the initial ethics investment, through empirical insights from three past cases we have worked on.

4 Empirical Examples

In this section, we use empirical data to provide an initial look at what types of processes are required to implement ethics and what kinds of activities result in costs when doing so. In Sect. 4.1, we describe the cases that the examples are from. In Sect. 4.2, based on these cases, we discuss the practicalities of implementing ethics from the point of view of resources and costs.

4.1 Cases and Data Description

To illustrate what the cost of implementing AI ethics means in practice, we build on three cases. Each case organization worked on a project where ethics was considered one of the key requirements. One of the projects was a blockchain project and the other two were ML development projects. The cases are illustrated in (table below 1)

Through these cases, we provide an initial look at the cost of implementing (AI) ethics, focusing on the initial *ethics investment*, as well as some early insights into *ethics maintenance* (Sect. 3.2). We utilize multiple types of data for each project, including interviews, project documentation, notes from workshops with developers, observation, etc. We feel that the use of a varied set of data lets us better explore a novel phenomenon such as this by giving us a clearer picture of what kinds of resources were needed to actively tackle ethics in a software development project. The types of data for each case are detailed in Table 1.

This data is used to illustrate what types of activities are associated with implementing AI ethics into practice, which are then discussed from the point of view of the types of costs discussed in Sect. 3.2. Thus, in terms of analysis, our focus is simply on *what* was done in the project to implement ethics, and what resources were needed to do so. As empirical studies in AI ethics are still lacking (see e.g.[10,16]), our understanding of what types of processes are needed to do so is consequently lacking as well. Through these cases, we are able to provide an initial look at the cost of AI ethics by looking at what types of activities may be involved when implementing AI ethics in practice. These cases let us evaluate the feasibility of the framework before further data collection.

Moreover, in this paper and these three cases, we approach ethics through specific ethical frameworks, which vary by case. As the study of Jobin et al. [8] highlights, there is a lack of a clear understanding of what exactly AI ethics is, or should be, with different principles being used in different contexts to approach AI ethics. By utilizing existing ethical frameworks, we (and the case organizations, more importantly) are able to clearly define *what* ethics means in the context of each case. This important as it also helps define what an ethical system should look like, and thus helps define what actions should be taken to reach that goal, directly affecting *how* ethics is implemented in each case.

Table 1. Overview of cases and data.

#	Context	Data sources	Data types
1	Blockchain	1 developer	Interviews, project documentation, developer notes
2	ML (predicting tool)	1 development team	Project documentation, notes from workshops with developers
3	ML (voice recognition)	Client company & 1 development team	Project documentation, notes from workshops with developers

4.2 Case 1

Case 1 summary:

- **Project context:** Data from a single developer working in a research-industry collaboration blockchain project.
- **Who implemented ethics:** As the project progressed, involving ethics into the project became the responsibility of a single developer. The developer discussed matters with an external ethics expert as needed.
- **Ethical framework used:** EU Guidelines for Trustworthy AI [3] & Prototype of ECCOLA [16], which was being developed at the time.

Project activities related to ethics (time spent) [stakeholders involved] in case 1:

- Decision to implement ethics made in a design meeting (2h). [Project management and developers]
- Initial training with ethics expert (1h). [Ethics expert and developer]
- Ethics as a part of biweekly iteration planning (1–2h x n) [Developer and scrum master]
- Use of ethical tool during development (?h) [Developers]
- Additional ethics documentation (1–5 sheets per iteration) [Developer]
- Expert hotline (?h) [Developer and ethics expert]
- Internal presentations documenting the implementation of ethics in the project [Developer and project management]

Case 1 Observations. In case 1, we observed most resources spent on ethics being spent early on in the project (i.e., on **ethics investment**). As the project progressed, although ethics resulted in recurring resource investments (expert hotline; role in biweekly planning), the investment was largely frontloaded. Simply defining *what* the investment (i.e., ethics) is takes resources, as ethics in SE is a novel phenomenon that requires clarification in each project context.

In this regard, one challenge was the project context: the project as a blockchain project, and no ethical frameworks for that particular project context were identified at the time. As a result, frameworks for AI ethics were utilized and had to be tailored to suit the project context based on discussion within the project (expert hotline; notable focus on ethics in biweekly meetings). This

highlights the importance of a suitable framework, as it saves resources by providing a clear(er) way of approaching ethics in the project context. Otherwise this requires internal effort.

In terms of the activities related to implementing ethics, ethics seemed to ultimately become a part of various project activities, blending in with other project activities, as opposed to being a tacked-on extra responsibility. However, some novel activities remained, such as the expert hotline with an AI expert, which would translate into ethics maintenance costs going forward. In addition, we noted that the implementation of ethics resulted in extra project documentation related to ethics. In part, this extra documentation was a result of ethics being a foreign topic for most stakeholders and necessitated in-depth explanation within the project.

4.3 Case 2

Case 2 summary:

- **Project context:** Data from a proof-of-concept ML project in a software company. Predicting tool for the educational domain. Project customer was interested in exploring potential ethical issues in the project.
- **Who implemented ethics:** Entire development team (4). The development team discussed matters with an external ethics expert on a weekly basis. Attendance in these meetings varied from 1 developer to the entire team.
- **Ethical framework used:** The ECCOLA method for implementing AI ethics [16].

Project activities related to ethics (time spent) [stakeholders involved] in case 2:

- Decision to implement ethics made in a design meeting (1h). [Project management, developers, ethics expert]
- Training workshop on using the ethical framework (ECCOLA) (1,5h). [3 ethics experts, entire development team, and 6 potential end-users]
- Ethics kickoff meeting (2,5h). [Ethics expert and entire development team]
- Use of ethical tool during development (?h) [Developers]
- Weekly check-up meetings with ethics expert (1h) [Ethics expert and 1 to 4 development team members]
- Additional ethics documentation and end reporting (1–2 sheets per iteration) [1–4 developers]

Case 2 Oservations. Compared to case 1, the decision to implement ethics in case 2 proceeded in a more straightforward manner. As the project was an ML project, it was possible to utilize a method for AI ethics (ECCOLA [16]). This made it easier for the stakeholders to approach ethics in the project context in various ways. I.e., *what* is going to be done and *how*. Consequently, early on in the project, actions related to ethics could be defined more accurately.

However, this did not result in ethics taking notably less resources. In fact, ethics seemed to take up *more* resources, compared to case 1, especially because the implementation of ethics involved more stakeholders in case 2.

Following the larger initial **ethics investment**, the implementation of ethics then proceeded more systematically. Whereas in case 1 the discussion on ethics continued throughout the project between the developer and the ethics expert, in case 2 the implementation proceeded as planned initially.

Going into **ethics maintenance**, the recurring, distinct ethics-related activities were the weekly check-up meetings with the ethics expert. However, as the project progressed, these focused more on the reporting progress rather than guiding discussion. The additional ethics documentation and reporting also continued, although this was not out of necessity, but because the company itself was curious how ethics was being handled in the project. Otherwise, ethics had become a part of the normal development activities of the company.

4.4 Case 3

Case 3 summary:

- **Project context:** Data from project where a design agency (client company) commissioned software from a consultant company. Project customer specifically requested ethical software.
- **Who implemented ethics:** 3 developers and product manager (senior dev.); 4 developers in total.
- **Ethical framework used:** The ECCOLA method for implementing AI ethics [16].

Project activities related to ethics (time spent) [stakeholders involved] in case 3:

- Decision to implement ethics made in a design meeting (1h). [3 client company representatives and ethics expert]
- ECCOLA tutorial, initial training for the used ethics framework (1,5h). [3 ethics experts, entire development team, and 5 customer representatives]
- Ethics kickoff (2,5h). [Ethics expert, entire development team, and 3 client company representative]
- Use of ethical tool during development (?h) [1–4 developers; varied by iteration]
- Weekly project meeting. Ethics was handled like any other requirement in the backlog (1h) [Entire development team and client company 2–5 representatives]

Case 3 Observations. Case 3 followed a similar pattern as the other cases in terms of the initial **ethics investment**. A notable initial investment was required to define what to implement. As the project then progressed, ethics, like in case 2, was incorporated into existing practices (e.g., discussing ethics in weekly project meetings as opposed to separate ethics-related meetings).

However, as the project began to draw to a close, resource optimization was carried out, and as a result, specifically ethics-related activities were cut. This seems to imply that ethics was nonetheless not completely embedded into any

existing processes and some **ethics maintenance** costs remained that warranted cutting. The customer, who had initially requested ethical software, ultimately considered it a secondary priority. It would, thus, seem that the potential **ethics revenues** were not considered worth the resources at this stage of the project.

5 Discussion

This paper furthers the AI ethics body of knowledge through empirical insights. As the field is lacking in empirical studies [12,13], our understanding of *how* AI ethics is implemented in practice is also lacking, which is considered to be a key issue in the area [9]. Through the practical insights from the three cases, we provide an initial look at the practice of AI ethics from the novel point of view of resources and costs, furthering this understanding. By providing an initial look at the cost of ethics in SE, we hope to motivate further interest on the practical questions of AI ethics.

To begin understanding the cost of ethics in SE, and AI ethics specifically, we turned to a software quality cost estimation framework [11], which we tailored for the context of ethics (Sect. 3). In this initial study, we approached the phenomenon through the project activities undertaken to implement ethics, in order to understand what requires resources when implementing ethics. While the framework provided a basis for this initial discussion, more detailed cost estimation frameworks specifically designed for the purpose of (AI) ethics could be developed going forward, if cost estimation becomes an active concern in ethics in SE.

Further on the note of our comparison to quality, akin to the past software quality experts, the implementation of ethics in SE, at this stage, seems to require an investment in ethics experts, external or internal. In all our cases, ethics experts were present throughout the project and actively leveraged for their expertise by the project staff. Canca [2] also argues that an ethics expert is required in the process so that developers can contact them when faced with challenging ethical issues (in this case, 'challenging' as defined by the tool they are proposing). A similar process was seen in our example cases, and especially case 1. Ethics experts, in this case external ones, were included in the project and provided assistance as needed. It would, thus, seem that ethics indeed requires a continuous investment (ethics maintenance).

5.1 Practical Implications

Ethics takes effort (resources). Ethics is still new in SE, and especially the ethical discussion on AI has made ethics a common topic of discussion recently. Implementing ethics into practice is still challenging and established practices and processes are lacking, making resource estimation difficult. This paper provides an initial look at what implementing ethics could mean in practice as far

as project resources are considered, highlighting that ethics requires resource commitment, with a focus on the initial investment.

However, as the practical implementation of (AI) ethics is still an emerging area of research and practice, the practices and processes required to do so may vary greatly between organizations and project contexts. In this regard, we would recommend the use of an ethical framework to guide the implementation of ethics. This can be a set of guidelines or a method, or any other suitable artefact that helps you define *what* is ethics in your project context. If no suitable framework exists for your application context, either use a more generic one (e.g., business ethics) or consider developing one yourself. By having a shared understanding of what ethics means for your project, you can start planning *how* to develop an ethical system.

Values will get implemented in a service whether it is done systematically or not. By actively looking to tackle AI ethics, it is possible to make a conscious, informed decision on which values to implement. Through nonconformance, it is left up to the developers and other stakeholders working on the system to implement their own values as they see fit, consciously or subconsciously.

5.2 Limitations

As these cases were proof-of-concept projects, we are not able to provide insights into *ethics revenues* and only some initial ones into *ethics maintenance* based on this data. Though our data from the three cases was collected over time, a more systematic, longitudinal approach would be required for a more comprehensive study looking at all three types of costs (ethics investment, ethics maintenance, and ethics revenues). In this regard, we also highlight that these are the results of our limited observation access; it is possible that the cases included more activities related to ethics we were not able to document. Nonetheless, given the novelty of the phenomenon, we feel that this paper provides a starting point for investigating AI ethics from a new point of view that is especially of interest to companies looking into AI ethics.

The use of an ethical framework, we argue, is pivotal in implementing ethics in practice in SE, also from a resource estimation point of view. A framework, such as a set of guidelines or a method, helps us define *what* ethics is in the given project context, giving us clear boundaries within which to work. Otherwise, notable effort is spent on defining the relevant concepts before starting, although such work may be required when operating in novel application areas. However, such frameworks arguably impact what is being done to implement ethics or how ethics is implemented as well. Our findings only serve to provide initial insights into what types of activities and resources *may* be needed when implementing ethics, but given the emergent nature of the area, these may vary greatly by project, based on the ethical framework being utilized, among other factors. E.g., guidelines may only contain sets of principles but little practical guidance, while a method might provide a process to utilize.

Finally, this paper simply provides an initial look at the phenomenon. The data we have utilized was not originally collected to evlauate the implementation

of (AI) ethics from a resource point of view, but to develop the ECCOLA method [16]. While we feel that it nonetheless serves as a starting point for studying this phenomenon, it is hardly a comprehensive look at the process of implementing ethics from a resource and business point of view. Some of the projects may have included activities related to the implementation of ethics that we were not able to document based on our data. On the other hand, as we were not explicitly investigating the resource point of view through our observation and other data collection, it could be argued to not have biased the results by motivating a more extensive investment. Ultimately, the goal of this data was simply to demonstrate the otherwise conceptual points of this paper.

6 Conclusions

In this paper, we provided initial insights into the cost of AI ethics. The current state of AI ethics is reminiscent of how software quality was approached in the early 2000 s. Often overlooked at the time, quality still had long-term consequences for software, was costly to implement in production, and was an interdisciplinary endeavor involving various stakeholders, much like AI ethics today. Over the decades, quality evolved from simple quality assurance to a continuous process embedded into organizational culture. Only time will tell whether AI ethics will also mature in the same way.

We adapted a framework for software quality cost estimation into the context of (AI) ethics after drawing parallels between AI ethics and software quality to justify doing so. Based on the framework, we proposed a similar cost framework for the implementation of AI ethics. We then utilized empirical data from three cases to elaborate on the proposed framework by providing an initial look at what types of activities result in the associated costs. Based on the empirical examples, ethics in SE seems to require a notable initial *ethics investment* (e.g., initial training and planning), followed by *ethics maintenance* (e.g., due to the continued involvement of ethics experts). However, the project activities related to ethics may vary between projects, and especially depending on the ethical framework used to guide the process, as tools such as methods may propose specific practices in SE, while tools such as ethical guidelines may necessitate internal effort to devise relevant processes and practices.

As for future research, the practical implementation of AI ethics remains a challenge. Overall, we urge further empirical studies into AI ethics in general, especially ones focusing on practices, methods, and processes for bringing AI ethics into practice. While we certainly urge further studies into the cost of AI ethics as well, for which this paper lays some initial groundwork for, we feel that a better understanding of *how* AI ethics is implemented is also required in this regard. It is arguably far easier to conduct resource estimation for a clear process than it is to do so for ad hoc implementation of AI ethics.

Acknowledgments. This work was partly funded by local authorities ("Business Finland") under grant agreement ITEA-2020-20219-IML4E of the ITEA4 programme.

References

1. Barraquier, A.: Ethical behaviour in practice: decision outcomes and strategic implications. Br. J. Manag. **22**, S28–S46 (2011)
2. Canca, C.: Operationalizing AI ethics principles. Commun. ACM **63**(12), 18–21 (2020)
3. Ethics guidelines for trustworthy AI (2019). https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
4. Friedman, B., Kahn, P.H., Borning, A., Huldtgren, A.: Value sensitive design and information systems. In: Early engagement and new Technologies: Opening up the Laboratory, pp. 55–95 (2013)
5. Giray, G.: A software engineering perspective on engineering machine learning systems: state of the art and challenges. J. Syst. Softw. **180**, 111031 (2021)
6. Gotterbarn, D.W., et al.: ACM code of ethics and professional conduct (2018). https://www.acm.org/code-of-ethics
7. Van den Hoven, J.: Moral methodology and information technology. In: The Handbook of Information and Computer Ethics, pp. 49–67 (2008)
8. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**(9), 389–399 (2019)
9. Mittelstadt, B.: Principles alone cannot guarantee ethical AI. Nat. Mach. Intell. **1**(11), 1–7 (2019)
10. Morley, J., Floridi, L., Kinsey, L., Elhalal, A.: From what to how: an initial review of publicly available ai ethics tools, methods and research to translate principles into practices. Sci. Eng. Ethics **26**(4), 2141–2168 (2020)
11. Slaughter, S.A., Harter, D.E., Krishnan, M.S.: Evaluating the cost of software quality. Commun. ACM **41**(8), 67–73 (1998)
12. Sloane, M., Zakrzewski, J.: German AI start-ups and "AI ethics": using a social practice lens for assessing and implementing socio-technical innovation. In: 2022 ACM Conference on Fairness, Accountability, and Transparency, FAccT 2022, pp. 935–947. Association for Computing Machinery, New York, NY, USA (2022)
13. Vakkuri, V., Kemell, K., Kultanen, J., Abrahamsson, P.: The current state of industrial practice in artificial intelligence ethics. IEEE Softw. **37**(4), 50–57 (2020)
14. Vakkuri, V., et al.: Time for AI (ethics) maturity model is now. arXiv preprint arXiv:2101.12701 (2021)
15. Vakkuri, V., Kemell, K.-K., Jantunen, M., Abrahamsson, P.: This is just a prototype: how ethics are ignored in software startup-like environments. In: Stray, V., Hoda, R., Paasivaara, M., Kruchten, P. (eds.) XP 2020. LNBIP, vol. 383, pp. 195–210. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-49392-9_13
16. Vakkuri, V., Kemell, K.K., Jantunen, M., Halme, E., Abrahamsson, P.: ECCOLA-a method for implementing ethically aligned AI systems. J. Syst. Softw. **182**, 111067 (2021)
17. Vakkuri, V., Kemell, K., Kultanen, J., Siponen, M.T., Abrahamsson, P.: Ethically aligned design of autonomous systems: Industry viewpoint and an empirical study. arXiv preprint arXiv:1906.07946 (2019)
18. Watson, T.: Reputation and ethical behaviour in a crisis: predicting survival. J. Commun. Manag. **11**(4), 371–384 (2007)