Making ethics practical: User stories as a way of implementing ethical consideration in Software Engineering

Making Ethics Practical: User Stories as a Way of Implementing Ethical Consideration in Software Engineering

Erika Halme[a], Marianna Jantunen[a], Ville Vakkuri[b], Kai-Kristian Kemell[c], Pekka Abrahamsson[d]

[a]*Faculty of Information Technology, University of Jyvaskyla, Mattilanniemi 2, Jyvaskyla, 40100, Finland*
[b]*School of Marketing and Communication, University of Vaasa, Wolffintie 32 FI-65200 Vaasa PL 700, Vaasa, FI-65101, Finland*
[c]*Department of Computer Science, University of Helsinki, Pietari Kalmin katu 5, Helsinki, 00560, Finland*
[d]*Faculty of Information Technology and Communication Sciences, Tampere University, Kanslerinrinne 1, Tampere, 33014, Finland*

Abstract

Context: Shortcomings of AI systems have recently brought ethics into the spotlight in Software Engineering (SE) in the form of AI ethics. However, actually implementing ethics into practice remains a challenge in both AI ethics and SE at large. Translating abstract ethical principles into requirements and features is difficult and lacks established processes, as well as practices and methods.

Objective: In this study, we explored user stories as a way of implementing ethics in SE. Initially, we simply investigated whether user stories could be utilized for this purpose. After we began to consider this possible, we began to develop the concept of ethical user stories (EUSs) as a specific practice for this purpose.

Method: We utilized a design science research (DSR) approach to first explore the use of user stories in implementing ethics, and then to develop the concept of EUS. This process featured three DSR phases through which the concept of EUS was iteratively developed with empirical data.

Results: Over three DSR iterations, we studied 689 user stories produced in different contexts including both student and industry settings. Based on the data, we defined the concept of EUS and provided empirical validation for it.

Accepted manuscript - Information and Software Technology

Conclusions: The concept of EUS provides a novel way of tackling ethics in SE. This paper presents the concept in-depth, along with practical suggestions for utilizing EUS.

Keywords: user story, ethics, AI ethics, design science research, software engineering practice
PACS: 0000, 1111
2000 MSC: 0000, 1111

---

1. Introduction

Recently, the shortcomings of existing AI/ML systems have highlighted various ethical issues that have resulted in calls for ethical consideration in AI (Artificial Intelligence) and ML (Machine Learning). Many of these failures have made global headlines and prompted widespread public and academic discussion on AI ethics. As a result, numerous guidelines for ethical AI have been proposed, including ones from national and supranational actors such as the EU, alongside scientific ones and ones devised by companies developing AI [1]. This discussion on AI ethics has served to bring attention to ethics in Software Engineering (SE) overall, which, though it does have a long history, has not been a major topic in SE research or practice in the past.

As with ethics in SE in general, bringing this ethical discussion into practice is a recurring challenge in AI ethics as well. For example, in terms of professional ethics, the ACM guidelines for professional conduct for software professionals [2] are a widely known set of ethical guidelines in the area. Yet, McNamara et al. [3] argue that it has had little impact out in the field, and Gogoll et al. [4] posit that this is the case for codes of conduct in SE in general due to their abstract nature. This has also been the case in AI ethics, where various sets of guidelines [1, 5] for ethical AI have been introduced, and whose impact in practice has similarly been called into question [6, 7]. This has also served as the primary motivation behind this paper. While such guidelines with their ethical principles communicate important issues to address, they arguably do not provide the tools for doing so.

AI systems are ultimately still software, although they have been described to be probabilistic as opposed to the deterministic nature of conventional software [7]. To this end, various existing papers have highlighted the importance of SE in ML development [8, 9]. While ML systems contain unique ML components and processes related to them not seen in conven-

tional software systems, these ML components are ultimately just a part of the larger system. Yet, in AI ethics, SE as a whole remains a gap. A myriad of conceptual papers exist, alongside various papers focusing on technical ML solutions related to AI ethics, while papers focused on AI ethics from the point of view of SE remain few and far between. Currently, the conceptual discussion on AI ethics and the resulting principles and guidelines seem to have had little impact on SE practice in the area [7]. This is also highlighted in various conceptual papers discussing the challenges of bringing AI ethics into practice (e.g., [6]).

To facilitate the development of ethical AI systems, we consequently wish to further shift the focus towards SE in AI ethics and the *development* of these systems. In this paper, we propose a novel way of implementing ethics (and AI ethics in particular): ethical user stories (EUSs). As the name implies, EUSs build on traditional user stories, which are an established agile requirements engineering (RE) practice in SE [10]. Consequently, the idea of EUSs is to help turn ethics into tangible requirements to be implemented into systems, as one way of implementing ethics in practice in SE. While AI ethics has motivated this research, this approach to implementing ethics is not exclusive to AI ethics at all. AI ethics is simply one of the more topical ethics-related issues in SE currently, both in academic literature and out in the field.

As a distinct type of user story, EUSs highlight the importance of ethical considerations. Aside from EUS being an artefact, creating an EUS is a process that promotes ethical consideration, leading to a proactive approach to ethics. In addition to presenting the EUS concept, this paper also explores the process of devising EUSs.

The contribution of this paper is, thus, two-fold. First, we present the concept of EUS as a novel theoretical concept and a SE artefact. We then empirically demonstrate EUSs and explore the process of devising EUSs through a two-phase study featuring, altogether, 689 user stories. The empirical portion of the paper provides insights into how EUSs function in practice. The specific research questions we tackle in this paper are: **(RQ1)** What are EUSs?, and **(RQ2)** What are the traits/elements/attributes that frame the EUS concept?

The rest of this paper is structured as follows. Section 2 presents the theoretical background of this study (ethics in SE, user stories, AI ethics). Section 3 presents the EUS concept. Section 4 presents the research approach used to formulate the EUS concept and validate it in practice through empir-

ical data. Section 5 presents our findings from the empirical studies. Section 6 discusses the theoretical and practical implications of our findings and their limitations. Section 7 concludes the paper.

2. Background

In this section, we discuss the theoretical background of this paper. In Section 2.1, we discuss ethics in SE in general, adopting a more general point of view on our specific topic. In Section 2.2, we discuss the specific ethical context that we focus on in this paper, AI ethics. In Section 2.3, we discuss user stories on a general level.

2.1. Incorporating Ethics in SE

Incorporating ethics in SE has been a challenge for decades, and it became a more widely known topic in the 1970s and 1980s. Software and software systems influence individuals, as well as create new contexts, which warrants ethical consideration [11]. Some have argued that they create entirely new ethical issues instead of simply providing existing ones with new application contexts [12]. There have been two traditional approaches to tackling these issues: education (of software engineers) [11, 12], and raising the level of professionalism in the field (e.g., the works of Gotterbarn et al. [2] on the ACM code of ethics).

Over the decades, a number of practical issues have prompted discussion related to ethics in SE or, more generally, IT. Some examples of such issues include: piracy (especially in the early 2000s) (e.g., [13]), security policy compliance in organizations [14], green IT (e.g., [15]), social media use (addiction etc.) (e.g., [16]), ethics in big data [17] and more generally data privacy issues, and, most recently, the downright boom in AI ethics (which we discuss in Section 2.2). Specific problem contexts such as these have provided lenses through which ethics has been approached in SE. Instead of more general, far-reaching attempts to incorporate ethics in SE, it could be argued that much of the discussion on ethics in SE has been focused on specific problems in specific problem contexts such as these.

One example of a more general approach to ethics in software development is value sensitive design (VSD) [18, 19]. VSD is a design approach that focuses on incorporating human values in software systems and advocates tackling ethical issues through design choices. It has achieved an established status in research. Although it still struggles with practical adoption among

software developers, it has remained prominent largely within the design community. Over the years, VSD has received its share of criticism. For example, framing and prioritizing values, alignment with existing methods, and industry adaptation have all been recognized as major challenges for VSD, and therefore for incorporating ethics [20].

Whereas in the past ethics has often remained on the sidelines, especially in the eyes of practitioners, the rise of AI ethics [21] has arguably resulted in new-found interest in ethical aspects in SE in recent years. For example, various widely known conferences have recently invited papers on ethical aspects in SE.

2.2. AI Ethics

Though this paper ultimately discusses a way of implementing ethics that could arguably work in SE overall, our specific application context is that of AI ethics. In part, this is due to our own research interest in AI ethics, which has served as one motivation for this paper as well.

AI ethics is a long-standing area of research far pre-dating the more recent advances in ML in the past decade. These recent advances, however, have certainly served to notably accelerate the discussion in the area. [21] Aside from the surge of research interest in the area, this discussion has made its way into the public sphere as well. Various incidents related to AI/ML systems have made the news, resulting in an increase of public awareness of these issues. Building on the lengthy history of AI ethics and its current topics of interest, AI ethics as a concept has now been defined and applied by various initiatives [7, 22] to establish rules, principles, guidelines, and standards [1, 23].

AI ethics is often approached through principles, which define what is ethical AI in a given context. Jobin et al. [1], based a review of various AI ethics guidelines, outline the following 11 principles: transparency, justice & fairness, non-maleficence, responsibility, privacy, beneficence, freedom & autonomy, trust, sustainability, dignity, and solidarity. However, none of these principles appear in *all* the reviewed guidelines, highlighting the still on-going discussion in the area [1]. These principles each deal with various issues and risks associated with ML. 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.

Indeed, interest in AI ethics has now surged to the point where multiple papers simply reviewing the various existing guidelines on AI ethics have

been published recently [1, 24], highlighting the number of such guidelines publicly available, in addition to any internal AI ethics guidelines utilized by organizations that are not available for public review. Such guidelines also include ones published by particularly notable parties such as the IEEE [25], and even the EU [26]. These guidelines are typically built around various *principles* that describe what an ethical AI system should be like. Bringing these principles into practice, however, remains a key challenge in the area.

In other papers [7, 27, 28], we have studied the state of practice in AI ethics and have argued, based on empirical data, that little is currently done in the way of implementing AI ethics out in the field. In particular, the numerous ethical guidelines created to help implement AI ethics (Jobin et al. [1] and Hagendorff [24] review these guidelines in their papers) seem to have had little impact on practice. The numerous tools related to AI ethics discussed by Morley et al. [29] seem to also not see much use, based on the studies looking at the current state of practice. While technical ML tools can help tackle ethical issues such as bias, it seems that the motivation to use these tools may often be missing. Moreover, such tools, while helpful and practical for what they are intended to do, are narrow in scope. Far fewer solutions exist for taking into account AI ethics on a project level in SE.

As one potential solution, we have proposed the ECCOLA method for implementing AI ethics in SE [30], and Siqueira De Cerqueira et al. [31] build on ECCOLA with their RE4AI guide. However, we feel that there is still much work to be done in bringing AI ethics into practice in SE. There is still a lack of empirical research in AI ethics overall [32]. Even though the implementation of AI ethics in practice remains a challenge, there is now demand for ways to do so. Policy-makers and companies have begun to take an interest in AI ethics. Perhaps in part due to the interest in AI ethics, ethics in SE more generally speaking has recently been in the spotlight as well.

There are many open questions that still need addressing in bringing AI ethics into practice, including the question of *who* should be involved in implementing AI ethics. According to Canca [33], simpler ethical issues could be addressed by developers alone, while more complex issues could necessitate the presence of an AI ethics expert. Hallamaa & Kallikoski [34] propose that ethical issues could be approached primarily from the point of view of safety and risks, focusing on favorable outcomes and avoiding unwanted effects in order to make the discussion more practical and less theoretical. This discussion is still nascent, and this paper presents one

more SE practice-oriented contribution aimed at tackling the challenge of incorporating AI ethics and ethics in general into SE.

2.3. User Stories

User Stories are popular tools [35, 36, 37] in agile software development, and specifically in RE, for facilitating communication between the customer and the development team in relation to requirements [38] [39]. User stories act as a bridge between requirements and software features. The idea of a user story is to provide context by describing how a system would be used by a user, and in doing so, to provide guidance for developing the system. From the point of view of RE, user stories can be linked to the entire requirement engineering process from requirement elicitation and documentation to requirement validation and management [40].

There are varied ideas of what a user story should look like. However, they are typically short sentences, written on an index card, a post-it note, or a user story template and expressed via natural language, emphasizing the goals of the system and its usage [41]. Jeffries [39] describes user stories using a triple C component scheme, introduced by Wake [42], where the *cards* are a physical medium, *conversation* as the discussion surrounding the cards, and *confirmation* for tests that verify them [42].

User stories are currently often formulated on the basis of a 2001 template by Connextra [1] that progresses as follows: "As a ..., I want to ..., so that ..." [38], or some variation of this three-part template. For example, Lucassen [10] proposes a new way of conceptualizing Cohn's [38] user stories by approaching it through four core elements: format, role, means, and end (e.g., means-end as opposed to Cohn's reason-goal).

In this paper, we turn to Cohn's [38] description of user stories, using the aforementioned three-part template. This is a common way of approaching user stories in the industry, and we wish to use established practices given that our topic otherwise is novel, to make it easier to approach (as further discussed in Section 3, and as is recommended in [30]). Cohn [38] describes user story writing as a story creation process and emphasizes that the customer or user should be involved in the process from start to finish, contrary to traditional RE practices in which the user is typically only involved at the start and at the end of the process.

[1] https://www.agilealliance.org/glossary/user-story-template/

3. Conceptual Model - Ethical User Stories

In this section, we discuss the concept of ethical user stories (EUSs). In Section 3.1, we discuss the theoretical justification behind the EUS concept and our motivation for pursuing it (i.e., the 'why'). In Section 3.2, we present the concept of EUS and explain how EUS are utilized in practice (i.e., the 'what' and 'how').

This conceptual model is based on the empirical studies discussed later in this paper in Sections 4 and 5. In Sections 4 and 5, we describe the process that ultimately led to the formulation of the concept described here, through multiple sets of data. This concept is the primary contribution of this paper. However, through the empirical data, we also provide some interesting observations and lessons learned to suggest further practical and theoretical implications, in addition to using it to provide support for the concept.

3.1. Theoretical Justification and Motivation

Our motivation for exploring the idea of EUSs originally stems from our past studies in AI ethics, as we have discussed in Sections 1 and 2.2. These is currently a strong demand for ethical systems in the field of AI, following global public discussion on the topic. This is highlighted by the large number of guidelines being proposed for ethical AI [1]. There is also a more widespread interest in ethics in SE in general as well in recent years. For example, privacy is arguably a current mega-trend among users.

However, implementing ethics in practice is often challenging due to the abstract nature of the principles and values it is conceptualized through. For example, AI ethics is a field where bridging the gap between theoretical discussion and practice remains an ongoing challenge (as we discuss in [7, 27] among others). This theoretical discussion has resulted in a set of recurring ethical principles (transparency, fairness, etc.) that highlight various potential ethical issues [1, 6]. However, these principles continue to provide little guidance in terms of tackling these issues in practice. In another study, we proposed a method for implementing AI ethics in order to help bridge this gap [30]. Yet, much more work is arguably still needed in this regard, as empirical studies in AI ethics in SE remain scarce [32]. Thus, we propose EUS as one way of approaching ethics in practice in SE.

Implementing ethics in practice in SE is a challenge on a more general level as well and is hardly limited to the specific area of AI ethics. AI ethics

has simply been the most prominent driver of ethical consideration in SE in recent years. Ethical consideration is not a natural part of SE. Developers do not normally take ethics into consideration, and ethics is not present in SE practices nor in methods typically used in SE. Even when pushing developers to take into account ethical guidelines (the ACM Code of Ethics [2]), McNamara et al. [3] saw little change in their ways of working. In part, this also highlights that it is difficult to approach ethics through abstract guidelines and principles, as is also the case in AI ethics [6, 7].

Converting abstract principles into practice is challenging. Even if relevant values are recognized through principles, if such guidelines do not provide actionable ways of implementing them, the responsibility of doing so lies entirely on the developers and the organization developing the system at hand [7].

We argue that, in order for ethical principles to get implemented, ethics needs to become a part of the SE process [30]. It is ultimately the developers, and the development team more generally, whose responsibility it is to somehow implement these ethical issues into the system being developed. Though the demand for ethics may come from the customer or other project stakeholders, the technical implementation is carried out by the developers. Thus, ethics needs to be present in SE practice, either through novel practices or through existing ones. On this note, Zuber et al. [43], calling for well-integrated ethical deliberation in software development, suggest that agile processes are particularly suitable for integrating them in the field. We therefore considered utilizing user stories, a well-known agile practice, for this purpose.

We approach ethics through RE. In SE, requirements drive development. Requirements describe what the system should do and are a tool for formulating features to be implemented in practice, based on customer or other stakeholder requirements. If we wish to see ethical issues being taken into account in development, we argue that they should be converted into requirements. In this regard, we have turned to user stories, which are a well-established RE practice for formulating requirements (as we have discussed in Section 2.3).

We have opted to utilize user stories for several reasons. First, user stories have been utilized in the past to introduce new elements into SE (e.g., quality [44]) and have also been explored as an avenue for implementing ethical consideration in SE [45]. Second, user stories being a widespread practice in RE is also beneficial in that it is a familiar practice for many developers and orga-

nizations. By utilizing an existing, widely known practice, we wish to make the implementation of ethics in this manner more approachable for development teams. To this end, Zuber et al. [43] also argue that the prevalence of agile development eases the integration of ethical thinking through familiar methods, and the structures in agile development empower developers and thus foster ethical thinking. Third, devising user stories requires discussion among developers and other stakeholders involved in devising them. Ethics is not straightforward and involves deliberation. Discussion and deliberation are consequently important when looking to determine how to approach ethical issues in a given (SE) context. Finally, writing stories to illustrate requirements and features can help make ethics more approachable for developers by illustrating what types of situations can involve ethical issues in practice.

Building on the existing practice of user stories, we propose a novel concept, EUS, to help implement ethics in SE. While it is possible to devise conventional user stories that tackle ethical concerns, ethics is not a part of conventional SE practice today [6]. As a practice, user stories alone do nothing to ensure that ethics is taken into account. Therefore, we propose EUS as a subpractice of the traditional user story for the following reasons:

1. Ethics is a new issue in SE practice, and developers are generally not familiar with ethics. Distinguishing new issues as a separate, new type of user story to highlight the importance of such type of considerations has been done before in terms of usability or security, for example [46] [47].
2. Despite the growing demand, ethics does not seem to get implemented "naturally" (i.e., without actively seeking to do so) in SE today [7, 28, 3]. It is argued that processes (practices, methods) to implement ethics are required [6, 30].
3. When ethics is distinctly present in a SE process/practice, it becomes easier to determine what has actually been done to implement ethics. In this case, this could result in a separate class of backlog items (similar to what is proposed in relation to GDPR by Bartolini et al. [48]). This way, the (possible) implementation of ethics does not need to be retroactively defined by analyzing various project artefacts [49].
4. While attributes such as quality or performance are more objective and more easily measured, ethics is more contextual and needs to be defined in a given context. Ethical considerations requires conscious

deliberation, or a framework.

3.2. What is an Ethical User Story and How to Devise One

EUSs enable ethically motivated features to be implemented into a software product. They are user stories intended to formalize and help tackle ethical issues in SE, by using an ethical framework as the lens for doing so. EUSs help bring ethics into SE practice by utilizing an existing RE practice (user stories) to convert ethics into tangible requirements.

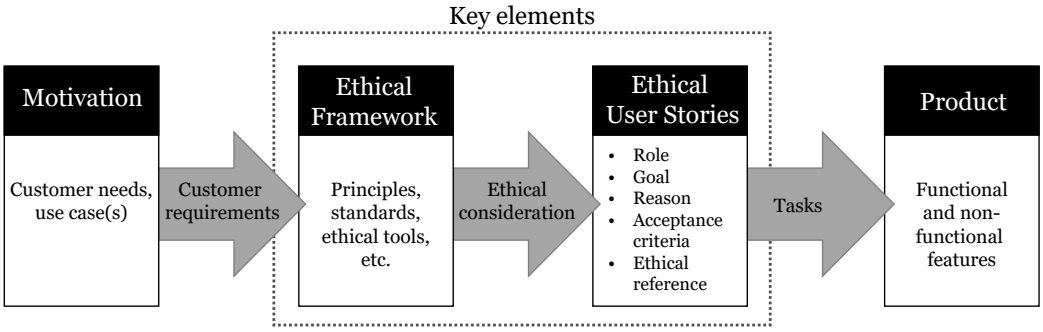

Figure 1: Conceptual model of EUS

Figure 1 describes the process of devising an EUS. In practice, EUSs are highly similar to traditional user stories, as seen in Figure 2. The main differences come from the process used to devise them (Figure 1, as well as the contents of the stories, which are related to ethical issues. Below, we cover this process in more detail by discussing the individual elements (i.e., (1) motivation, (2), ethical framework, (3) (ethical) user stories, (4) product)).

Motivation. A motivation to develop an ethical system is required as a starting point for devising an EUS. This motivation, though, can also have a practical impact on the development. For example, in our research endeavors, we have observed industrial cases where the (external) customer was particularly concerned with developing an ethical AI system when commissioning one. Thus, the customer was the one pushing the development team(s) to take into account ethical issues, and tackling these issues in practice was something the stakeholders had to handle among themselves. In this fashion, the stakeholders need to determine, together, what they wish to achieve in terms of ethics by formulating requirements for the system. To this end, motivation also involves considering *what* (kind of system) is being

developed, and this has a direct impact on what type of ethical framework should/could be selected for the project.

Ethical framework. An ethical framework is required to define relevant ethical issues in any given project context. An ethical framework provides the lens through which ethics is approached in the project. This ethical framework can be any ethical tool: a method, a set of guidelines, or even one as abstract as an ethical theory. The purpose of the framework is to help determine what is 'ethical' in the given context, and to provide direction on which ethical issues to take into account. The framework should be as closely related to the given context as possible. For example, one should use an AI ethics framework when working with AI/ML systems, as opposed to a more generic one.

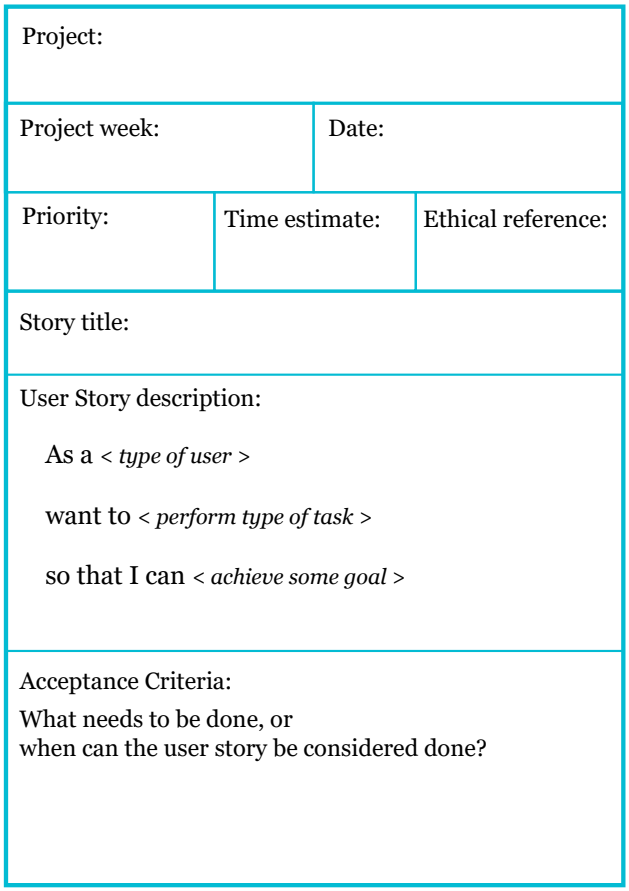

Figure 2: Example of a template for an ethical user story

(Ethical) user stories. As an artefact, an EUS closely resembles a traditional user story. EUSs contain all the elements typically present in user stories, e.g., the typical three-part-template of *who* (role), *what* (goal), and *why* (reason) [38]. This is often formulated as "As a [...], I want to [...], so that [...]". These three elements have also been included into the US process overview (see Figure 1). However, we would note that EUSs do not have to be devised following this particular template. We have simply selected it due to its widespread use, and thus, familiarity to practitioners.

In addition to these typical user story elements, we introduce an ethics-related component, *ethical reference*, that links the user story to the ethical framework at hand. The ethical reference indicates what part(s) of the framework(s) the particular user story is related to. For example, if using an AI ethics framework, one might document which AI ethics principle(s) a given EUS is related to. This also helps in producing a paper trail and documenting trade-offs, which is an important part of ethical consideration [30].

Additionally, in terms of their contents, EUSs should tackle the potential ethical issues identified based on the framework of choice. We explore this in more detail in Section 5 through empirical data.

Product. Based on the EUSs, functional and non-functional requirements are determined, which are then implemented into the product. As a result, a system that takes ethical issues into account is (ideally) developed. However, benchmarking whether a system is ethical or not cannot be done through EUSs alone. Such benchmarking is context-dependent and requires a framework, and overall, presents various challenges in AI ethics, for example. We consider this an interesting topic for future research.

Having introduced the EUS concept and covered the creation of EUSs as a process, we now elaborate further on EUS as an artefact through a practical example of an EUS. Figure 3 showcases an EUS written as a part of a real industrial case. In this case, an AI/ML-enabled smart terminal system was being developed for use in harbor environments for passenger flow management and estimation. The customer was concerned with making the system ethical, and ECCOLA was introduced as a tool for facilitating ethical consideration in the project. ECCOLA was used to direct the creation of (ethical) user stories to formulate requirements for this system.

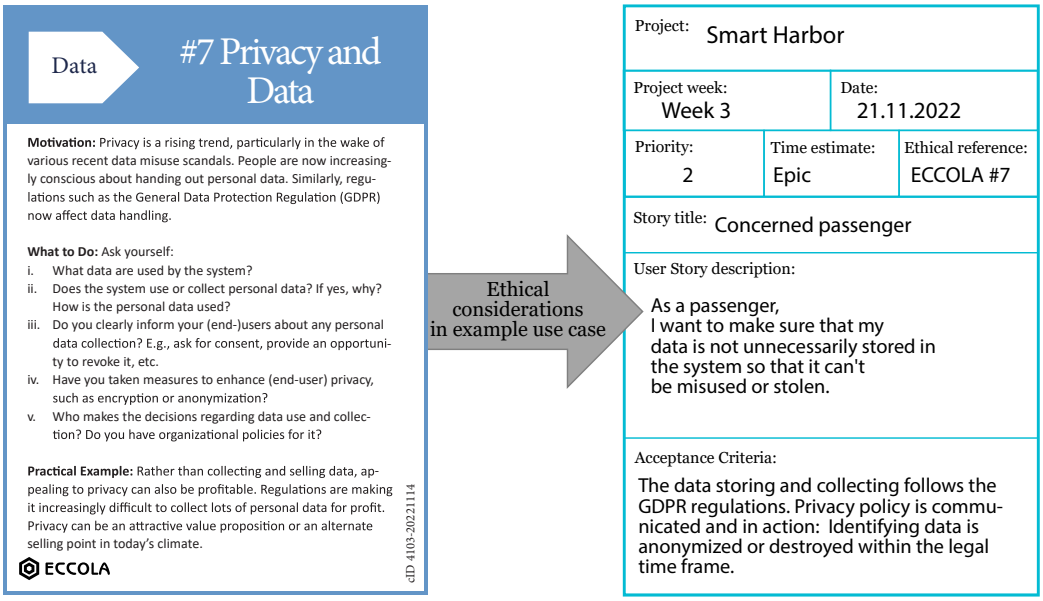

Figure 3: An ethical user story example from an industrial project, using ECCOLA as framework.

4. Research Design

While the purpose of this paper is to introduce, in-depth, the EUS concept as a novel way of implementing ethics in practice in SE, we consider it relevant to also present empirical validation for the concept. In doing so, we 1) evaluate whether it is actually possible and feasible to implement ethics through EUSs, 2) showcase how the EUS concept was iteratively formulated, and 3) further demonstrate *how* EUSs can be written and what they look like in practice. For this empirical validation, we have conducted a holistic multiple case study utilizing different sources of data. The general research approach we have used is design science research (DSR), based on Peffers et al. [50]. In this section, we describe this DSR process and its three iterations (referred to as phases).

In Figure 4, we illustrate this research process and its phases thus far. In the following process description, we cite studies we have published on the topic where applicable, although for the most part, these findings are novel. We summarize our relevant existing research (motivating studies and preliminary results) in Table 1. The empirical validation and application of the EUS concept is still an ongoing process, and we continue to explore the

use of EUSs in different contexts, including industrial ones. In this paper, we wish to present the EUS concept in-depth. In other words, what are EUSs, and how do you write them? We also discuss results and lessons learned from these existing papers in relation to this point of view.

All the EUSs we have used for the analysis are available on Figshare. [2] This repository includes all the user stories we have analyzed *that were considered EUSs*. Over the three phases, we have analyzed a total of 689 user stories, but not all of these were ultimately considered EUSs, as is discussed in the rest of Section 4, as well as Section 5.

The rest of this section is split into three subsections. Each subsection covers one phase, and thus there are three subsections, one for each of the three phases seen in Figure 4.

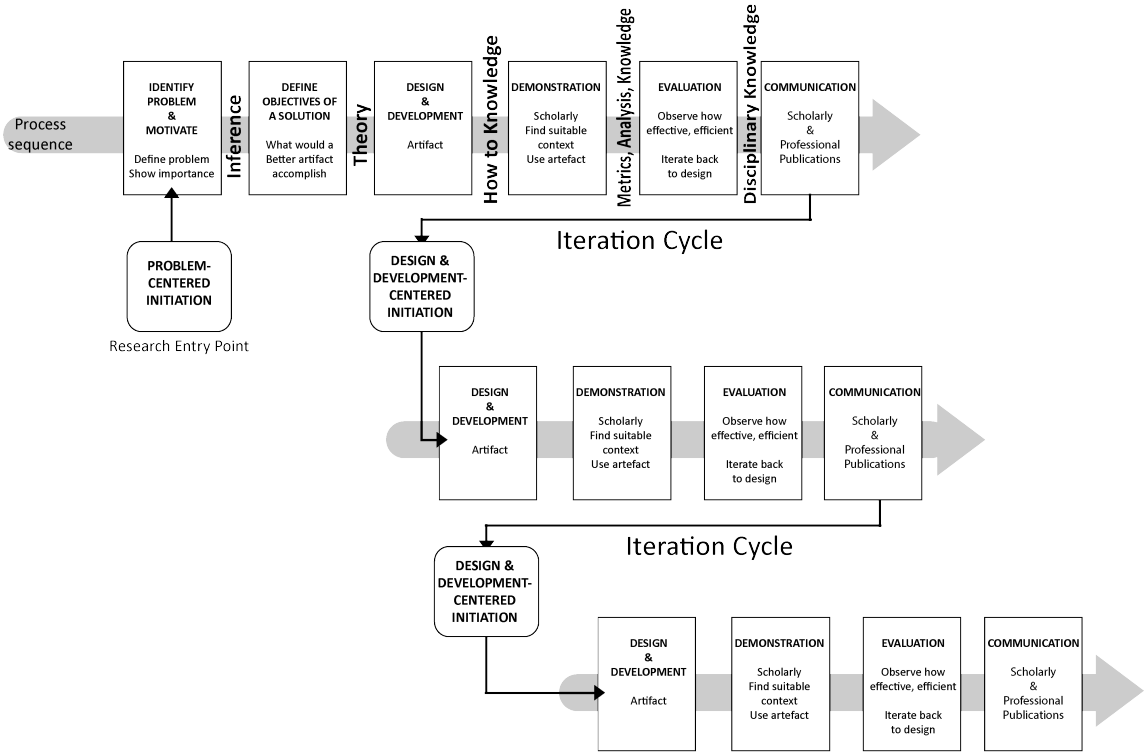

Figure 4: Process of developing the EUS concept, based on the DSRM process of Peffers et al. [50]

[2] https://figshare.com/s/723d290fd27ec98fa139

Motivation

Study	**Results**
Various studies on AI ethics practice [7, 27, 28, 30]	Motivation to pursue novel ways to implement AI ethics (led to EUS)

Preliminary results from EUS research process

Publication	Goal(s) of the publication
(Conference paper) How to Write Ethical User Stories? Impacts of the ECCOLA Method [51]	To evaluate the ECCOLA method. During the study, piloting the idea of utilizing user stories to implement (AI) ethics became a secondary goal.
(Preprint) Implementation of Ethically Aligned Design with Ethical User Stories in SMART Terminal Digitalization Project: Use Case Passenger Flow [52]	Documenting research progress in a research & industry collaboration project. Secondary goal was to publish the data set (industry EUSs) for academic review.
(Doctoral symposium paper) Ethical User Stories: Industrial Study [53]	Presenting overview of research for doctoral symposium.
(Doctoral symposium paper) Ethical Tools, Methods and Principles in Software Engineering and Development: Case Ethical User Stories [54]	Presenting overview of research for doctoral symposium.
(Extended tutorial abstract) Utilizing User Stories to Bring AI Ethics into Practice in Software Engineering [55]	To present the EUS concept for tutorial and conference participants & to receive academic feedback on EUSs.

Table 1: Relation to our existing papers

4.1. Phase 1

Identify problem & motivate. We have extensively discussed the problem context and motivation in Sections 1, 2, and 3. In brief, the lack of ethical consideration in AI/ML development is an ongoing problem that is becoming increasingly relevant the more ubiquitous such systems become. While this has been our primary motivation for this study, incorporating ethics into SE more generally is also a recurring challenge acknowledged in extant research.

Define objectives of solution. This has also been covered in this paper, particularly in Section 3. We aim to create an artefact to help incorporate ethics into SE processes. The artefact should be approachable and familiar to software developers, and thus we turned to existing SE practices (user stories) for inspiration.

Design & development. We have discussed EUSs as a concept extensively in Section 3. However, at this stage of the process, going into Phase 1, we were still simply experimenting with the idea of utilizing user stories to implement ethics in SE and had no clear artefact in mind for what an ethical user story should be like. Unsure of whether this was plausible in the first place, we conducted an initial, exploratory study into the topic, which we discuss next.

Demonstration. We first began to explore the utilization of user stories to implement ethics in SE in 2021. Data for phase 1 was collected through a practical, project-based course on information systems (IS). The participating students were master level students of IS and computer science (CS). These students formed 15 project teams in total with 3-5 members per team, and, during the course, worked as teams on a real-world case given by a client company.

The project duration was six weeks. During this time, the students received five assignments, one each week after the first week's introductory lecture. These assignments comprised two parts: non-technical and technical. The non-technical part was the focus of this study and formed the basis of data collection. User stories were discussed during the lectures to familiarize the teams with the practice of producing them. Then, during the following weeks, the teams produced user stories and worked on implementing them. The first ones were produced for the second week, and, in the following weeks, each week of the project featured user story revision, the creation of new user stories if/when applicable, and a check to see if user stories were implemented in the product.

Phase	Participants	Objective	Data and Analysis	Outcome
1	15 student teams (3-5 students per team)	Exploring the potential of using user stories in implementing ethics in SE	Data: 298 user stories. Analysis: user story quality evaluation with INVEST model and Grounded Theory coding	Emergence of the EUS concept as a high-quality, human-centric user story
2	27 individual students (participants of Phase 1)	Exploring the role of an ethical framework in EUS creation and evaluation	137 user stories analyzed were through the ethical framework to determine if they contain an ethical component	Identifying the defining features of an EUS through the addition of an ethical framework
3	Industry project members (SEA4VALUE, SMARTER), 3 workshops, a total of 51 participants (industry & academia)	Characterizing the EUS concept, exploring EUSs in an industrial context, identifying relevant stakeholders	254 EUSs analyzed with focus on the process of their creation and their defining features, using grounded theory approach and cross-referencing the ethical framework	Defining the EUS concept and identifying its components, confirming the role of the ethical framework in the creation of EUSs

Table 2: Summary of the research phases

The purpose of this study was to explore how ethical consideration could impact user story creation. Thus, we introduced an ethical framework into the process. To this end, these 15 teams were divided into two groups for the entire duration of the course: (1) the control group, which received generic instructions (lecture and handout) on how to write user stories, but no ethical framework, and (2) the ECCOLA group, who received an lecture on ethical frameworks and an ethical framework to utilize in the form of some cards from the ECCOLA method [30] in addition to instructions on how to write user stories. There were six teams in the control group and nine teams in the ECCOLA group. The purpose of this split was to see what impact the use of an ethical framework had on the user stories being produced. Altogether, the teams produced 298 user stories during the project. [3]

Evaluation. These 298 user stories were analyzed using the INVEST model. The INVEST model [56] is a tool for evaluating the quality of user stories through a scoring system. The name of the tool is an acronym that contains the list of attributes to be scored. Thus, according to INVEST, user stories should be: [I]ndependent, [N]egotiable, [V]aluable to purchasers/users, [E]stimable, [S]mall, and [T]estable. This scoring was done by three teachers and/or researchers. The purpose of the INVEST evaluation was to understand whether user stories dealing with ethical issues could be of high quality as well, as it has been argued [38] that user stories have to be of high quality to get implemented at all.

In addition to utilizing INVEST as a framework for analysing the data from Phase 1, we also looked at the contents of the user stories more generally through a grounded theory approach, which was used to code the data. This was done in order to make more qualitative observations about the data as well, given the exploratory nature of the topic.

Communication. We have published some results about this data in [51]. However, these focused on the further evaluation of the ECCOLA method [30], while the EUS concept was of secondary importance at the time. The findings of the paper simply served as an initial proof-of-concept for the idea of implementing ethics through user stories. In Section 5, we discuss further lessons learned from this data that relate to the EUS concept and their utilization in practice.

[3] These are available online in an external repository: https://doi.org/10.6084/m9.figshare.14210753

4.2. Phase 2

Design & development. Based on the data from Phase 1, we began to consider EUSs an actual possibility for implementing ethics, prompting us to conduct further studies on the topic. The DSR process thus continued iteratively, as we returned to the *design & development* stage of the process. The *problem & motivation* and the *objectives of the solution* remained the same as in Phase 1. In Phase 1, we had the pre-conception that an EUS was simply a human-centric user story. Going into Phase 2, we became more interested in the effects of the ethical framework being introduced into the evaluation process. We began to consider EUSs closely linked with the ethical framework used to devise them to the point where it seemed like a defining aspect. In other words, in order to determine whether a user story is an EUS or not, it needs to contain elements from an ethical framework. Still working to explore and define the concept of EUS, we continued to be interested in understanding how EUSs differ from conventional user stories in Phase 2.

Demonstration. The demonstration of Phase 2 was carried out in a somewhat similar manner to that of Phase 1. Master level CS and IS students were tasked with producing (E)US, and were once again split into a control group who utilized no ethical framework, and the ECCOLA [30] group that utilized one. These were students from Phase 1 who were already familiar with writing user stories. The main differences compared to Phase 1 were as follows. First, the (E)US were produced by individuals rather than teams. Second, the (E)US described requirements for a hypothetical system rather than for a real-world case. Third, we asked the students to evaluate how much time they spent on devising the user stories. Fourth, the user stories were produced independently at one's own pace, rather than in a course project environment. In this fashion, 137 user stories were produced.

In Phase 2, we were interested in any observations related to EUS as an artefact or a process, and as mentioned earlier, particularly in relation to the ethical framework and the contents of the (E)US. Additionally, we explored the time aspect: would it take longer to devise EUSs compared to traditional user stories? This would be relevant going into industrial use, and indeed, the purpose of Phase 2 was to also provide further validation for the EUS concept before moving on to industrial evaluation.

Evaluation. In Phase 2, we focused on the *contents* of the 137 (ethical) user stories. First, the contents of the user stories were evaluated through an ethical framework (ECCOLA) to determine whether they focused on ethical issues related to the system or not. We compared such user stories to other

user stories to further explore the idea of EUSs. Second, after determining which user stories were EUSs, we conducted an additional thematic analysis to further explore their contents.

The first evaluation, i.e., to determine which user stories were EUSs, was carried out in two stages. In the first stage, three separate evaluators evaluated all 137 user stories on a scale of 0 to 2 (where 0 = ethical component not present, 1 = ethical component present, and 2 = ethical component unclear). If two evaluators agreed on a score of (be it 0, 1, or 2), this was the final score for the particular user story. On the other hand, user stories with no such score consensus were discussed among the evaluators in order to reach a consensus, while reflecting on what it was that caused the conflicting scoring.

In the second stage, for further rigour, two additional evaluators worked independently to score the same 137 user stories. Once finished, all five evaluators went over the scores together and discussed any unclear cases to come to a final conclusion for each individual user story. In this second stage, the final scores assigned to each user story in the first stage were compared to the evaluations of the two individual evaluators of stage 2, so that, like in the first round, two out of three scores had to match. The purpose of this second stage was to further understand what types of elements or contents in the user stories caused disagreements in the scores. After determining which user stories were EUSs, we analyzed the remaining EUSs using a thematic analysis [57] approach. This was conducted on the EUSs to understand what characteristics they shared.

Communication. The results of this phase have not been communicated in an existing study. We discuss them in this paper in Section 5. Most importantly, the results were looked at among the research team and motivated us to move on with utilizing the EUS concept in an industrial setting in Phase 3.

4.3. Phase 3

Design & development. Based on the results of Phases 1 and 2, we began to study EUSs in an industrial context. At this stage, we officially added the *ethical reference* as a component to the EUS concept based on Phase 2. The purpose of the ethical framework is to provide a definition for what is 'ethical' in the given context when looking at the contents of a user story. Moreover, the ethical reference serves to directly link a user story to a

component of the framework, providing transparency to the process (which is considered pivotal in AI ethics, for example [1].

Demonstration. In Phase 3, we began to utilize EUSs in an industrial context after exploring it in a university course setting in Phases 1 and 2. This industrial setting was a project between ten industrial partners and five research institutes, which was led by an R&D company. It was a maritime industry project that focused on utilizing AI/ML in a smart terminal context through three use cases: (1) ship turnaround, (2) truck traffic, and (3) passenger flow. The project also concerned AI ethics and looked at ways to implement it into such solutions. To this end, we explored the implementation of AI ethics through EUSs in all three use cases. As was the case in Phases 1 and 2, we utilized the ECCOLA method [30] as the ethical framework here. Otherwise, the process was split into two stages.

In the first stage, three virtual workshops focused on devising EUSs were organized, one for each use case. In these workshops, the use case owner(s) were asked to choose the themes from the ethical framework (i.e., the ECCOLA cards) that were best suited for their use case. The selected cards were then placed on a digital whiteboard, and the workshop participants were divided into groups. These groups were divided into breakout rooms, and each room contained industrial partners and one researcher familiar with the ECCOLA method and AI ethics in order to facilitate the discussion if/where needed. The participants also received instructions on how to use the ethical framework beforehand. The cards were used to direct and facilitate ethical discussion from the point of view of the project and the system(s) being developed. The participants made 367 notes based on the discussion and their previous experiences with the project. From the point of view of the project, the purpose of the workshops was to address any further ethics-related issues that had not already been addressed earlier in the project.

In the second stage, eight researchers from the University of Jyväskylä (who worked as developers in the project) began to devise EUSs based on the workshop notes from the first stage. This work was done over the course of multiple online meeting sessions where the participants worked on a shared spreadsheet containing the notes while on a voice/video call. This continued until all the notes had been exhausted by converting them into user stories. Altogether, 254 EUSs were devised. These user stories were also cross-referenced among the researchers to ensure the quality of the user stories, as quality is important for user stories [38].

Evaluation. In Phase 3, we focused on two aspects in our evaluation:

(1) the process of devising EUSs (this time in a real-world industrial setting with various stakeholders), and (2) what do real-world EUSs look like? While the students in Phase 1 also devised EUSs for a real case project, this was a fully industrial setting with no student data. In terms of the process, we looked at, for example, how the notes translated into user stories, and whether the presence of a familiar ethical framework could guarantee that a user story became an EUS, etc. with an explorative approach. In terms of more formal data analysis, we utilized a grounded theory approach to code the 254 user stories, looking at recurring characteristics of the user stories. Additionally, we looked at each user story from the point of view of the framework (ECCOLA) and its themes (cards) to determine whether a single EUS could overlap with multiple themes of the framework, etc.

Communication. In this paper, we make further observations about the concept and process of devising EUSs based on this industrial data (367 workshop notes and 254 EUSs). We have published some preliminary results based on a subset of these data in [53]. In that paper, we only looked at 125 user stories from a single use case (passenger flow).

5. Results and Lessons Learned

In this section, we discuss our results from the three DSR phases detailed in Section 4. Sections 5.1 to 5.3 cover one phase each, while Section 5.4 provides a brief look at the current state of this ongoing research endeavor and provides some additional observations. While discussing our results, we highlight some key observations as findings that have supported the creation of the artifact (EUS) during the iterative DSR process. These findings are also highlighted due to their perceived practical and theoretical relevance beyond this DSR process.

5.1. Phase 1 Results

In Phase 1, we began to explore EUS. At this stage, our plan was to simply see whether user stories could be used as an approach to implementing ethics, and to also further evaluate the ECCOLA method (these results relating to ECCOLA were the focus of [51]). We did not plan on creating a separate concept for this initially.

As we looked at our analysis of the 298 user stories again for this paper, the idea of EUSs began to emerge. We found that some of the user stories

contained two elements we considered relevant for an EUS: (1) high quality and (2) human-centricity.

Moreover, such user stories were successfully implemented in the system being worked on in the project, based on the evaluation of the course projects conducted by the teaching staff who evaluated user story implementation, among other things (using the user story acceptance criteria as a basis). Based on these observations, we considered user stories a potential way of bringing ethical consideration into the SE process. It seemed that EUSs could be written with the same skill-set as traditional user stories, and even without an ethical prompt.

While analyzing the data of Phase 1, we placed emphasis on evaluating the quality of the user stories. Cohn [38] argues that quality is key in having user stories actually get implemented. Thus, if an EUS is not of high quality, it will (likely) not get implemented, and thus the implementation of ethics in this manner is likely to fail as well. Based on our analysis of the data of Phase 1 (presented in more detail in [51]), it seemed to be possible to write EUSs that are of high quality, tackling this concern. In brief, by scoring the user stories using the INVEST model, we identified many user stories that we considered ethical that were also of high quality on the INVEST scale. Thus, EUSs as user stories did not seem to be impractical.

However, at this point our idea of EUSs was arguably still fuzzy. We considered EUSs, for which we did not have a definition yet, to be human-centric user stories of high quality, as we attempted to look at the concept from a very general point of view. It was not until later that we began to consider the link between an EUS and an ethical framework essential (as we discussed in Section 3) in defining the concept. Nonetheless, at this stage we considered our findings encouraging in that implementing ethics through user stories seemed plausible, leading us to continue on this path.

Finding 1: *Implementing ethics through user stories is possible.*

Finding 2: *Ethical user stories can be, and need to be, of high quality.*

5.2. Phase 2 Results

In Phase 2, we placed more emphasis on the fit between the ethical framework and the user stories. In our analysis of 137 user stories in Phase 2, we looked at the the user stories through the lens of the ethical framework being used (i.e., ECCOLA [30]). As the user stories themselves did not involve any direct references to the framework, we went over the user stories, scoring

them in the manner discussed in Section 4.2 (where 0 = no ethical component, 1 = ethical component, 2 = unclear). Here, the 'ethical component' was considered a link to the framework; did the user story deal with the contents of the framework? In doing so, we were also able to look at the differences between user stories scored 0 and 1, as well as what types of user stories received a score of 2.

Based on this analysis, 54 out of the 137 user stories total were considered EUSs (i.e., were scored 1). Looking at these 54 EUSs by means of thematic analysis (the results of which are in Figure 5), we determined three elements that (at least) comprised an EUS: (1) motivation, (2) presence of ethical components (according to the framework), and (3) end-user focus. However, in Phase 3, we began to doubt whether an end-user focus was in fact required at all, as we discuss in Section 5.3. Whereas (2) is self-explanatory, motivation, here, refers to a user story being motivated by an ethical issue. In other words, even if the user story discussed a straightforward technical feature, it could be an EUS if this feature was motivated by an ethical concern.

Finding 3: *EUSs should be motivated by ethical consideration. Even a technical user story can be an EUS if it is motivated by ethical consideration.*

Finding 4: *EUSs should contain an ethical reference that denotes what ethical framework and what part of that framework the user story relates to.*

Additionally, in Phase 2, we looked at EUSs from a resource point of view. In other words, how long does it to take to write an EUS compared to a traditional user story? For this purpose, the time spent on writing user stories was self-reported by the participants. Here, we looked at the points of data where this time was reported on the level of individual user stories. In all, 21 participants reported time estimates for 102 individual user stories.

On average, based on these 102 user stories, a user story took 18 minutes to create. The shortest time spent on an individual user story was 6 minutes, while the longest time was 37 minutes. However, variances in the time spent were between individuals rather than types of user stories: some participants simply took longer than others to devise user stories, and vice versa. Based on this data, there seemed to be no difference in time spent on EUSs compared to traditional user stories.

Finding 5: *EUSs do not seem to take longer to devise than traditional user stories.*

We considered this an interesting observation going into industrial settings. Arguably, resources are always important from the point of view of companies. It is far more likely for ethical consideration to receive some

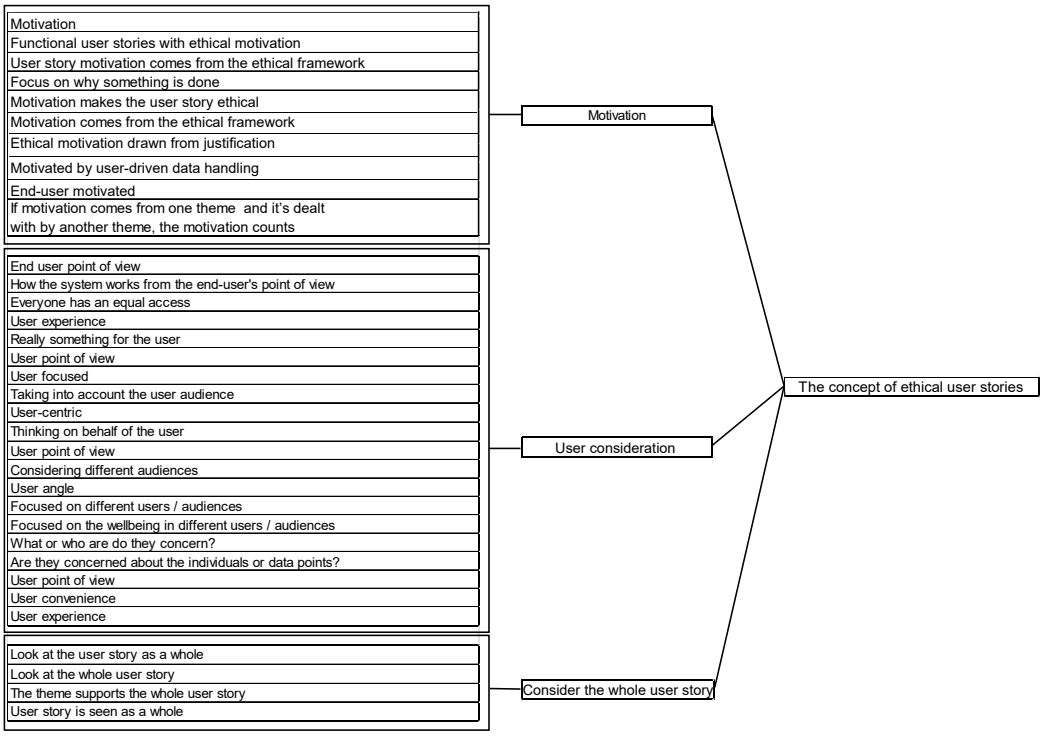

Figure 5: Results of the thematic analysis of 54 EUS in Phase 2

attention the less additional resources doing so takes from the organization. As we have discussed in Section 3, this was one reason we opted to look at existing SE practices here as opposed to introducing entirely new ones: a familiar practice should take less effort to utilize.

5.3. Phase 3 Results

Going into industrial evaluation in Phase 3, we reflected on our experiences and the data from the previous two cycles. In the design & development stage of Phase 3, based on our experiences in Phase 2, we determined that an ethical framework was necessary for defining EUS as a concept.

Finding 6: *An ethical framework is required to determine whether a user story is an EUS.*

In Phase 3, we looked at 254 EUSs devised as a part of an industrial project, as discussed in Section 4.3. Based on Phase 2, we now introduced a direct link to the ethical framework in the EUS in the form of the *ethical*

reference component. In other words, each user story should contain information on what part of what ethical framework it is inspired by. Each of the 254 user stories contained such a component, and these were a part of the analysis in Phase 3. Overall in Phase 3, we further continued to look at EUSs in order to better characterize the concept, this time in an industrial setting.

Given the real-world industrial setting, the process of devising EUSs was also approached from a practical point of view in Phase 3. In other words, which stakeholders should be involved in the process and how could it be carried out? In this case, the process was organized in the manner described in Section 4.3. Based on our experiences, it needs to be highlighted that the motivation of the relevant stakeholders to implement ethics is crucial, as much of ethical consideration is based on discussion among stakeholders. The work done with different stakeholders (e.g., use case owners) during the workshops was what enabled the developers to write EUSs. In this process, the role of the ethical framework was also crucial as it enabled the participants to have a shared understanding of ethics in the given context, and served as a way of directing the discussion towards relevant issues.

Finding 7: *Motivation to consider ethical issues is important not only on the level of individual user stories, but also on the level of the entire project.*

Finding 8: *Ethical consideration involves deliberation or discussion, and the choice of an ethical framework is important in defining relevant issues to focus on.*

Moving from the process to the EUSs themselves, we can make some observations based on the grounded theory analysis. The introduction of the ethical reference made the resulting user stories directly linked to the ethical framework. Each user story was written based on an element of the framework and conversely could be traced back to it through the direct reference.

However, as we analyzed the user stories, it seemed that a single user story could involve elements from different components of the framework. In this case, we could assign multiple ECCOLA cards to a single user story, even though it was motivated by just one. Though a potentially interesting observation, it is more related to the framework and AI ethics principles than the concept of EUS.

Moreover, during the process of writing these 254 EUSs, we characterized three types of EUSs: (1) human-centric EUSs (e.g., personnel training,

user needs, system goals, regulations), (2) EUSs with a technical focus (e.g., testing, back-ups, logistics), and (3) EUSs focused on data and information (e.g., data privacy, data handling).

Based on the data from Phases 1 and 2, we were under the assumption that EUSs needed to be human-centric, as they had all been thus far in the data. However, in Phase 3, contradicting with our initial ideas from Phases 1 and 2, we found EUSs that were not human-centered, but only ethically motivated and linked to the ethical framework. In other words, EUSs can be very feature-oriented and technical as well, largely externally resembling a conventional user story.

Finding 9: *EUSs do not have to be human-centered. While they still need to have a human-centered motivation, their contents can be purely technical.*

5.4. Results Summary & Current Situation

In Table 3, we summarize our findings from the DSR process we have presented in this section. Together, these findings have resulted in the EUS concept we present in Section 3. Some of the findings also provide practical insights into how EUSs could or should be utilized.

At this stage, we feel that the EUS concept itself has been validated and is mature enough to be presented through this paper. Our research on EUSs, however, continues, with a focus on their use in practice in industrial settings. We feel that there is still work to do on EUSs. In particular, we are still interested on understanding how EUSs are prioritized and actually implemented in industrial settings. We further discuss this in relation to future research suggestions in Section 7.

6. Discussion

In this paper, we began to tackle a practical challenge in the field of AI ethics: bringing ethics into practice in SE. AI ethics is a field where practical implementation is an ongoing challenge [7]. Despite various guidelines being proposed to help implement ethics [1], utilizing these in practice is difficult [7], and existing tools are largely either ML techniques of narrow scope or guidelines [29]. Taking ethics into account is also a challenge in SE overall [4, 45]. AI ethics is simply the most prominent field in this regard currently.

Based on our results, we argue that EUSs (see Section 3) present one way of bringing ethics into practice in SE. By leveraging an existing SE

#	**Description**
1	Implementing ethics through user stories is possible.
2	Ethical user stories can be, and need to be, of high quality.
3	EUSs should be motivated by ethical consideration. Even a technical user story can be an EUS if it is motivated by ethical consideration.
4	EUSs should contain an ethical reference that denotes what ethical framework and what part of that framework the user story relates to.
5	EUSs do not seem to take longer to devise than traditional user stories.
6	An ethical framework is required to determine whether a user story is an EUS.
7	Motivation to consider ethical issues is important not only on the level of individual user stories, but also on the level of the entire project.
8	Ethical consideration involves deliberation or discussion, and the choice of an ethical framework is important in defining relevant issues to focus on.
9	EUSs do not have to be human-centered. While they still need to have a human-centered motivation, their contents can be purely technical.

Table 3: Findings

practice with minor changes, we are able to use a familiar practice that is easily understood by developers, making it more approachable. In our DSR process, we have seen that EUSs can be produced like any other user story, even if the individuals producing them do not have a notable background in ethics. Overall, we have demonstrated that user stories can be utilized to convert ethical consideration into tangible requirements through the concept of EUSs. This concept has been developed iteratively through three phases of DSR. In doing so, we have provided a contribution that may help in bridging the prominent gap in AI ethics discussed in various extant studies (e.g., [6, 7]). This contributions also extends to implementing ethics in SE in general, as EUSs are not a practice specific to AI ethics.

While we have introduced the practice of EUS in this paper, the processes for utilizing it and for implementing ethical consideration into SE overall are still an open question. The workshop-oriented approach for industrial settings that we utilized in Section 5.3 involved both developers and AI ethics experts and comprised of two stages (the initial workshop with both parties present and the later sessions for converting the workshop notes into user stories among the AI ethics experts). This is but one way of devising EUSs. For example, Canca [33] suggests that simple ethical issues could be tackled by developers alone, while more complex ones could necessitate the presence of AI ethics experts.

Seeing as most organizations arguably do not have AI ethics experts at their disposal, utilizing practices (and processes) that developers alone could potentially utilize successfully could make AI ethics more approachable for a wider range of organizations. Utilizing familiar agile practices helps in this regard [43], which has been the case here as we build on an existing practice: user stories. In Phases 1 and 2 of the empirical portion of this paper, EUSs were successfully devised without the involvement of AI ethics experts. We consider a suitable ethical framework to be particular important in this regard, as a complex and abstract framework is likely to be more difficult to utilize without the presence of ethics experts with domain knowledge.

The choice of an ethical framework is a notable issue in and of itself. First, finding and selecting a suitable framework can be difficult without involving experts (and possibly even for experts). Second, there is a lack of established ethical frameworks both in AI ethics [6] and in SE in general, which is the root of this problem. Nonetheless, as ethics is easily a vague and broad topic without a framework to guide discussion, selecting any framework is arguably better than trying to rely on ethical assumptions and personal convictions.

Let us return to the two research questions (RQs) we outlined at the start of this paper: (1) what are ethical user stories?, and (2) what are the traits/elements/attributes that frame the EUS concept? Through the DSR process (Sections 4 and 5) utilized to define and validate the concept, we have answered both of these RQs in Section 3. In brief, an ethical user story is *a user story related to ethical issues or themes present in a given ethical framework, which transforms them into features of a software product.*

We answer the second RQ through the process of devising an EUS (Figure 1), which highlights the elements needed to devise an EUS. First (in addition to the EUSs themselves that are implemented into the product), a motivation to implement ethics, on both the project level and on the level of individual user stories, is needed. Second, an ethical framework is required to define what is 'ethical' in the given context, and to direct ethical consideration and discussion towards relevant issues. Third, the resulting user stories need to include references to this ethical framework, and their contents should reflect those of the framework in order to be considered EUSs.

Finally, EUSs are, above all, a practice for the SE of today. It is possible that, in the future, ethical consideration will have become a more established part of SE to the point where it is embedded into existing practices and routinely taken into account where relevant. However, as it is now, ethics is still considered distant.

6.1. Practical Implications

EUSs present an approachable way of bringing ethical consideration into SE through a familiar practice. The concept is presented in detail Section 3. Here, we present practical insights we have gained from our own experiences with EUSs thus far over the three DSR phases described in this paper.

Select a suitable ethical framework. The ethical framework is key in writing EUSs. It (1) defines what 'ethics' means in the given context and (2) provides you with an idea on what kind of ethical issues you should be tackling. While an ethical framework can even be a classic ethical theory, it should be something more context-specific when possible. For example, if you are interested in AI ethics, it would ideally be something related to AI ethics (e.g., guidelines [1], or the ECCOLA method [30] we have used in this paper). One more general-purpose ethical framework could be the RESOLVEDD strategy [58, 59], or the recent IEEE standard for addressing ethical concerns during systems design [60].

Discuss, discuss, discuss. As ethics remains a novel topic in the context of SE, it necessitates consideration, which is achieved through discussion. Instead of trying to go directly to tangible requirements, it would be beneficial to have a more open discussion related to the system and its use context. Use the ethical framework to direct the discussion: what issues pointed out by the framework seem relevant for *your* development context? Moreover, discussion is at the very core of ethics in general, as there are seldom any 'right' answers.

Make notes. Notes leave a paper trail (and it is important to leave a paper trail in ethics [7, 30]), but more importantly, the notes serve as a basis for EUSs. Notes are a low-effort way of bridging the gap between the ethical framework and user stories. They help you formulate your ideas, while also making them visible for others taking part in the discussion. Notes can also be iteratively combined and modified as the discussion progresses, and multiple notes can converge into one user story (as we have discussed in Section 4.3).

Write ethical user stories. Your notes serve as a natural starting point for devising EUSs. The EUSs themselves largely resemble traditional user stories: the difference is that they tackle ethical issues you have highlighted in your discussion. The only practical difference is that they should contain a reference to your ethical framework (e.g., which ECCOLA [30] card you used for that user story, if you used ECCOLA as the framework). User story quality (e.g., as measured with the INVEST model [56]) is important for EUS practice; quality in user stories usually translates into practical, actionable user stories. Even though you are dealing with ethical issues, your user stories should be something your developers can work with.

6.2. Limitations

As we have utilized different types of data collection and analysis approaches in the three DSR phases of this study, there are also various limitations that should be acknowledged. We would like to emphasize, however, that the main contribution of this paper is the concept/practice of EUS. The role of the empirical studies has been to validate the feasibility of this novel approach to ethics in SE through its use in practice. Thus, while the evaluations have their shortcomings, we argue that they nonetheless have served, at very minimum, as a proof-of-concept for EUSs.

Perhaps the most prominent limitation stems from the way we have utilize ethical frameworks to devise EUSs in this paper (the process of which we have

discussed in Section 3). In this regard, the limitation is that, in all three phases of the study, we have utilized the same ethical framework: ECCOLA. There were two reasons behind this: (1) all the empirical contexts were related to AI/ML, and we were not aware of any other suitable framework for AI ethics in particular (seeing as, e.g., AI ethics guidelines are considered to be poorly suited for practical use [6]), and (2) we developed ECCOLA, and we were interested in potentially further studying the use of ECCOLA in the process as well. Nonetheless, the framework plays a notable role in devising EUSs, as we discuss in Section 3, and thus it would be of interest to explore how different frameworks might affect the process and the resulting EUSs. For example, if the ethical framework is difficult to understand, the process of devising EUSs based on it is likely to also require more effort, and may result in lower quality user stories. We highlight this as a future research suggestion in Section 7.

Evaluating the quality of user stories presents another limitation. In Phase 1 of the DSR process (Section 5.1), we utilized the INVEST model to evaluate the quality of the user stories. We selected INVEST due to it being a tool promoted by the Agile Alliance [56], although ultimately it is but one of different frameworks for doing so. It nonetheless provided us with some metric of evaluating whether EUSs could be 'good', or practical, as many of the quality attributes of INVEST are related to the user story being practical.

Moreover, during the process, we reflected on the relevance of concept validity in this context. In other words, how to measure what is ethical, and to that end, how to determine the presence of ethical consideration in a user story. Ultimately, based on the DSR process, we chose to tackle this issue by introducing the use of an ethical framework as a requirement for devising EUSs. The ethical framework is now used to determine what is 'ethical' in a given context, and adherence to the framework needs to be evaluated while evaluating the ethical consideration in a user story. While, arguably, it is possible to write a user story that contains ethical consideration even without actively using an ethical framework to do so, we consider the systematic implementation of ethics to require intent, which is actualized through the use of an ethical framework. As ethics tends to be a vague concept, a framework provides the means to shared understanding of ethics in a given context.

As we have utilized data from different contexts over the course of the three phases, the data collection and analysis approaches also result in a num-

ber of limitations. Overall, explorative qualitative research has well-known limitations relating to descriptive, interpretive, and theoretical validity [61]. As for more specific limitations, first, we have utilized both student and industrial data. The use of student data is a recurring topic of discussion across scientific disciplines, and in SE, it is an accepted research approach. For example, Šmite et al. [62] and Falessi et al. [63] consider the use of student data to be a valid approach in SE. However, we have also utilized industrial data to build on the student data. Second, the qualitative analysis approaches used (thematic analysis [57] and grounded theory) both present some limitations to consider. In both approaches (in the case of thematic analysis, when not using a framework to generate the themes, as was done here), researcher interpretation plays a notable role and results in some subjectivity. We have attempted to tackle this issue by having multiple (2-5) researchers conduct the analysis in each phase.

Finally, in this paper, we have not focused on how the user stories have ultimately impacted the systems being developed in the different DSR phases (the user stories of Phases 1 and 3 were written for real-world systems being developed). In Phase 1, the teaching staff of the course determined whether the user stories were implemented, but in Phases 2 and 3, we had no data on the implementation. Past Phase 1, we have relied on the generally accepted assumption that high quality user stories are more likely to get implemented, and in this regard, we have confirmed that EUSs can (and should) be of high quality. This is something we wish to focus on in future studies on the topic, while the purpose of this paper was to establish the concept and practice of EUS and to provide initial empirical support for it.

7. Conclusions

The recent surge of interest towards AI ethics has also highlighted challenges in implementing ethics in practice in SE. In AI ethics in particular, bringing the various ethical principles discussed in various existing works [1] into practice remains a key challenge in the area [6, 7, 30]. Making ethics a part of SE practice is a larger challenge outside the specific context of AI ethics as well [3], however, and this recent interest in AI ethics seems to have resulted in a heightened interest in ethical issues in SE at large.

In this paper, we have discussed the implementation of ethics in SE through the use of user stories. We have developed and presented the concept of EUS (Section 3) through DSR (Section 4). EUS is a novel practice that

builds on an existing SE practice, in other words, user stories, to provide a way of incorporating ethical consideration as a part of existing SE processes. In building on an established, existing practice, our aim has been to make ethics more approachable for software developers in particular.

The concept has been developed over three DSR iterations. In the process, we have studied 689 user stories produced in different settings, using different data collection and analysis approaches to do so. In addition to using these data to build the concept of EUS, we have highlighted various observations and lessons learned related to the use of EUSs in Section 5. At present, EUSs provide one potential way of implementing ethics in SE, which we have provided some empirical support for.

Future Research

Further empirical studies into ethics in SE are still sorely needed, especially in the field of AI ethics. Existing studies highlight various problems with bringing abstract, high-level ethical principles into practice in SE (e.g., [3, 6, 7]). However, few empirical studies in general exist in AI ethics [32], and thus we, too, recommend further empirical studies in the area of AI ethics. To this end, we urge further studies into how ethics could be implemented in practice outside the specific area of AI ethics as well. In particular, we feel that further practices and methods that could make ethics a part of SE are needed. Ethical consideration, at present, is arguably not something that is conventionally seen in SE, and is likely to feel distant to developers. This is something that suitable practices and methods could help address.

We also recommend further studies into utilizing user stories as one such practice. This paper simply presents a starting point for doing so, with some empirical support. First, we would recommend utilizing different types of ethical frameworks in doing so. For example, would AI ethics guidelines, which are often considered impractical [6], function as an ethical framework in devising EUSs? Second, we personally have plans to continue exploring EUSs by looking at how they are ultimately converted into features once they have been written and are being utilized in development, which is another potential research direction. Third, future research could explore the creation of more dedicated user story templates for EUSs, as we have simply utilized the existing three-part template widely used in SE today (based on Cohn [38]), with the additions described in Section 3.2. The role-goal-reason template could be further investigated in this context. Finally, we have provided some initial insights into the resource point of view of the EUS process,

in that it does not seem that EUSs take longer to devise than traditional user stories. However, upon exploring the process of converting EUSs into features, it would be interesting to also explore whether *this* takes longer than it does with traditional user stories.

References

[1] A. Jobin, M. Ienca, E. Vayena, The global landscape of ai ethics guidelines, Nature Machine Intelligence 1 (9) (2019) 389–399.

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

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

[4] J. Gogoll, N. Zuber, S. Kacianka, T. Greger, A. Pretschner, J. Nida-Rümelin, Ethics in the software development process: from codes of conduct to ethical deliberation, Philosophy & Technology (2021) 1–24.

[5] B. Attard-Frost, A. D. los Ríos, D. R. Walters, The ethics of ai business practices: a review of 47 ai ethics guidelines, AI and Ethics (2022).

[6] B. Mittelstadt, Principles alone cannot guarantee ethical ai, Nature Machine Intelligence 1 (11) (2019) 501–507.

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

[8] S. Martínez-Fernández, J. Bogner, X. Franch, M. Oriol, J. Siebert, A. Trendowicz, A. M. Vollmer, S. Wagner, Software engineering for ai-based systems: A survey, ACM Trans. Softw. Eng. Methodol. 31 (2) (apr 2022). doi:10.1145/3487043.
URL `https://doi.org/10.1145/3487043`

[9] G. Giray, A software engineering perspective on engineering machine learning systems: State of the art and challenges, Journal of Systems and Software 180 (2021) 111031. doi:https://doi.org/10.1016/j.jss.2021.111031.
URL https://www.sciencedirect.com/science/article/pii/S016412122100128X

[10] G. Lucassen, F. Dalpiaz, J. M. E. Van Der Werf, S. Brinkkemper, Forging high-quality user stories: towards a discipline for agile requirements, in: 2015 IEEE 23rd international requirements engineering conference (RE), IEEE, 2015, pp. 126–135.

[11] J. H. Moor, What is computer ethics?, Metaphilosophy 16 (4) (1985) 266–275.

[12] D. G. Johnson, Computer ethics, Prentice-Hall, Inc., 1994.

[13] T. T. Moores, J. C.-J. Chang, Ethical decision making in software piracy: Initial development and test of a four-component model, MIS Quarterly 30 (1) (2006) 167–180.
URL http://www.jstor.org/stable/25148722

[14] D. Chatterjee, S. Chatterjee, S. Chatterjee, An empirical investigation of the influence of organizational virtues on information technology security policy compliance, in: ACIS 2021 proceedings, 2021.

[15] T. Butler, M. Daly, R. Hackney, Socio-technical transitions towards environmental sustainability through green ict, in: SIGGreen Pre-ICIS 2015 Workshop, 2015, pp. 1–12.

[16] J. R. Drake, D. Hall, J. B. Becton, C. Posey, Job applicants' information privacy protection responses: Using socia media for candidate screening, Transactions on Human-Computer Interaction 8 (4) (2016) 160.

[17] L. Perreault, Big data and privacy: control and awareness aspects, in: CONF-IRM, 2015.
URL https://aisel.aisnet.org/confirm2015/15

[18] B. Friedman, Value-sensitive design, interactions 3 (6) (1996) 16–23.

[19] B. Friedman, D. G. Hendry, Value sensitive design: Shaping technology with moral imagination, Mit Press, 2019.

[20] B. Friedman, M. Harbers, D. G. Hendry, J. van den Hoven, C. Jonker, N. Logler, Eight grand challenges for value sensitive design from the 2016 lorentz workshop, Ethics and Information Technology 23 (2021) 5–16.

[21] J. Borenstein, F. S. Grodzinsky, A. Howard, K. W. Miller, M. J. Wolf, Ai ethics: A long history and a recent burst of attention, Computer 54 (1) (2021) 96–102.

[22] V. Dignum, Ethics in artificial intelligence: introduction to the special issue (2018).

[23] H. AI, High-level expert group on artificial intelligence (2019).

[24] T. Hagendorff, The ethics of ai ethics: An evaluation of guidelines, Minds and machines 30 (1) (2020) 99–120.

[25] T. I. G. I. on Ethics of Autonomous, I. Systems, Ethically aligned design: Prioritizing human wellbeing with autonomous and intelligent systems (2019).
URL `https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf`

[26] Ethics guidelines for trustworthy ai (2019).
URL `https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai`

[27] V. Vakkuri, K. Kemell, J. Kultanen, M. T. Siponen, P. Abrahamsson, Ethically aligned design of autonomous systems: Industry viewpoint and an empirical study, EJBO - Electronic Journal of Business Ethics and Organization Studies 27 (2022).

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

[29] J. Morley, L. Floridi, L. Kinsey, A. Elhalal, From what to how: an initial review of publicly available ai ethics tools, methods and research to translate principles into practices, Science and engineering ethics 26 (4) (2020) 2141–2168.

[30] V. Vakkuri, K.-K. Kemell, M. Jantunen, E. Halme, P. Abrahamsson, Eccola — a method for implementing ethically aligned ai systems, Journal of Systems and Software 182 (2021) 111067. doi:https://doi.org/10.1016/j.jss.2021.111067.
URL https://www.sciencedirect.com/science/article/pii/S0164121221001643

[31] J. A. Siqueira De Cerqueira, A. Pinheiro De Azevedo, H. Acco Tives, E. Dias Canedo, Guide for artificial intelligence ethical requirements elicitation-re4ai ethical guide, in: 55th Hawaii International Conference on System Sciences, 2022.

[32] M. Sloane, J. Zakrzewski, 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 '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 935–947.

[33] C. Canca, Operationalizing ai ethics principles, Communications of the ACM 63 (12) (2020) 18–21.

[34] J. Hallamaa, T. Kalliokoski, Ai ethics as applied ethics, Frontiers in Computer Science (2022).

[35] S. Dimitrijević, J. Jovanović, V. Devedžić, A comparative study of software tools for user story management, Information and Software Technology 57 (2015) 352–368.

[36] I. K. Raharjana, D. Siahaan, C. Fatichah, User stories and natural language processing: A systematic literature review, IEEE Access 9 (2021) 53811–53826.

[37] W. Hudson, User stories don't help users: Introducing persona stories, interactions 20 (6) (2013) 50–53.

[38] M. Cohn, User stories applied: For agile software development, Addison-Wesley Professional, 2004.

[39] R. Jeffries, A. Anderson, C. Hendrickson, Extreme programming installed, Addison-Wesley Professional, 2001.

[40] A. De Lucia, A. Qusef, Requirements engineering in agile software development, Journal of emerging technologies in web intelligence 2 (3) (2010) 212–220.

[41] X. Wang, L. Zhao, Y. Wang, J. Sun, The role of requirements engineering practices in agile development: an empirical study, in: Requirements engineering, Springer, 2014, pp. 195–209.

[42] B. Wake, Invest in good stories, and smart tasks (Aug 2003).
URL https://xp123.com/articles/invest-in-good-stories-and-smart-tasks/

[43] N. Zuber, J. Gogoll, S. Kacianka, A. Pretschner, J. Nida-Rümelin, Empowered and embedded: ethics and agile processes, Humanities and Social Sciences Communications 9 (1) (2022) 1–13.

[44] G. Lucassen, F. Dalpiaz, J. M. E. M. v. d. Werf, S. Brinkkemper, The use and effectiveness of user stories in practice, in: M. Daneva, O. Pastor (Eds.), Requirements Engineering: Foundation for Software Quality, Springer International Publishing, Cham, 2016, pp. 205–222.

[45] P. Kamthan, N. Shahmir, On ethically-sensitive user story engineering, in: Proceedings of the 4th International Conference on Computer Science and Software Engineering, CSSE '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 71–79. doi:10.1145/3494885.3494899.
URL https://doi.org/10.1145/3494885.3494899

[46] A. M. Moreno, A. Yagüe, Agile user stories enriched with usability, in: Agile Processes in Software Engineering and Extreme Programming: 13th International Conference, XP 2012, Malmö, Sweden, May 21-25, 2012. Proceedings 13, Springer, 2012, pp. 168–176.

[47] J. Peeters, Agile security requirements engineering, in: Symposium on Requirements Engineering for Information Security, Vol. 12, 2005.

[48] C. Bartolini, S. Daoudagh, G. Lenzini, E. Marchetti, Gdpr-based user stories in the access control perspective, in: M. Piattini, P. Rupino da Cunha, I. García Rodríguez de Guzmán, R. Pérez-Castillo (Eds.), Quality of Information and Communications Technology, Springer International Publishing, Cham, 2019, pp. 3–17.

[49] F. Gilson, M. Galster, F. Georis, Extracting quality attributes from user stories for early architecture decision making, in: 2019 IEEE International Conference on Software Architecture Companion (ICSA-C), 2019, pp. 129–136. doi:10.1109/ICSA-C.2019.00031.

[50] K. Peffers, T. Tuunanen, M. A. Rothenberger, S. Chatterjee, A design science research methodology for information systems research, Journal of management information systems 24 (3) (2007) 45–77.

[51] E. Halme, V. Vakkuri, J. Kultanen, M. Jantunen, K.-K. Kemell, R. Rousi, P. Abrahamsson, How to write ethical user stories? impacts of the eccola method, in: International Conference on Agile Software Development, Springer, Cham, 2021, pp. 36–52.

[52] E. Halme, M. Agbese, H.-K. Alanen, J. Antikainen, M. Jantunen, A. A. Khan, K.-K. Kemell, V. Vakkuri, P. Abrahamsson, Implementation of ethically aligned design with ethical user stories in smart terminal digitalization project: Use case passenger flow, arXiv preprint arXiv:2111.06116 (2021).

[53] E. Halme, M. Agbese, J. Antikainen, H.-K. Alanen, M. Jantunen, A. A. Khan, K.-K. Kemell, V. Vakkuri, P. Abrahamsson, Ethical user stories : Industrial study, in: Joint Proceedings of REFSQ-2022 Workshops, Doctoral Symposium, and Posters & Tools Track co-located with the 28th International Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2022), CEUR Workshop Proceedings, 3122, 2022.

[54] E. Halme, Ethical tools, methods and principles in software engineering and development: Case ethical user stories, in: D. Taibi, M. Kuhrmann, T. Mikkonen, J. Klünder, P. Abrahamsson (Eds.), Product-Focused Software Process Improvement, Springer International Publishing, Cham, 2022, pp. 631–637.

[55] K.-K. Kemell, V. Vakkuri, E. Halme, Utilizing user stories to bring ai ethics into practice in software engineering, in: International Conference on Product-Focused Software Process Improvement, Springer, 2022, pp. 553–558.

[56] What does invest stand for? (2021).
URL https://www.agilealliance.org/glossary/invest/

[57] D. S. Cruzes, T. Dyba, Recommended steps for thematic synthesis in software engineering, in: 2011 International Symposium on Empirical Software Engineering and Measurement, 2011, pp. 275–284. doi:10.1109/ESEM.2011.36.

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

[59] V. Vakkuri, K. Kemell, P. Abrahamsson, Ethically aligned design: An empirical evaluation of the resolvedd-strategy in software and systems development context, arXiv preprint arXiv:1905.06417 (2019). arXiv:1905.06417.
URL http://arxiv.org/abs/1905.06417

[60] IEEE, IEEE Standard Model Process for Addressing Ethical Concerns during System Design, Standard, IEEE (Sep. 2021).

[61] J. Maxwell, Understanding and Validity in Qualitative Research, Harvard Educational Review 62 (3) (2010) 279–301. arXiv:https://meridian.allenpress.com/her/article-pdf/62/3/279/2110973/haer_62_3_8323320856251826.pdf, doi:10.17763/haer.62.3.8323320856251826.
URL https://doi.org/10.17763/haer.62.3.8323320856251826

[62] D. Šmite, C. Wohlin, T. Gorschek, R. Feldt, Empirical evidence in global software engineering: a systematic review, Empirical software engineering 15 (2010) 91–118.

[63] D. Falessi, N. Juristo, C. Wohlin, B. Turhan, J. Münch, A. Jedlitschka, M. Oivo, Empirical software engineering experts on the use of students and professionals in experiments, Empirical Software Engineering 23 (2018) 452–489.