Staging Reflections on Ethical Dilemmas in Machine Learning: A Card-Based Design Workshop for High School Students

ABSTRACT

The increased use of machine learning (ML) in society raises the question of how ethical dilemmas and choices inherent in computational artefacts can be made understandable and explorable for students. To investigate this, we developed a card-based design workshop in which high school students confront and reflect upon ethical dilemmas by designing their own ML applications. The workshop was developed in an iterative process engaging four high school classrooms with students aged 16-20. Through iterations on the design of the workshop we found that a) understanding of fundamental ML served to qualify students’ ethical reflections, b) students’ design process served to them reveal to the complexity of the ethical dilemmas and tie them to the properties of the technology and to design decisions, c) while we were able to stage qualified reflections regarding ethical dilemmas, students struggled with addressing these dilemmas in their designs.

INTRODUCTION

The current focus on teaching Computational Thinking (CT) [46] across all levels of education, and the increased ease of incorporating emerging technologies such as machine learning (ML) into new applications raise the importance of investigating how ethical aspects of new technologies can be made explainable and explorable to future designers and consumers of these technologies.

In recent years, ML has changed what is possible to achieve with the use computational processing, expanding computers’ ability to understand and interact with the world. This development promises great new possibilities, but is also currently transforming the nature of social interaction, work, education, etc. [33], increasing information asymmetry [37] and making technologies less comprehensible to users [1]. This introduces a range of ethical dilemmas and issues unique to ML. These ethical issues have been addressed in both the ML community through, e.g., the ACM Conference on Fairness, Accountability, and Transparency (ACM FAT*), as well as in the HCI community [1, 39, 14], which implore that implementing ML into the world includes making ethical judgements and approaching ethical issues in the system design.

Today, companies are integrating ML into everyday technologies and infrastructures such as smart phones, maps, streaming services, etc., to provide better services to customers and to
build new business models. However, if ethics are not carefully considered (and they often are not), ML can become a tool for unethical conduct, resulting in dark patterns [11], hypernudging [47], etc. Designers and developers are navigating a myriad of different interests, balancing ethics against system efficiency, profitability, user convenience and are constantly making moral decisions and value judgements. Ethics are difficult, especially when applied in the real world, and designers and users alike are often faced with dilemmas, when designing and using ML-powered products. We argue, that to take active part in shaping the future, understanding everyday technologies is imperative, and thus youth should be able to recognise these dilemmas in technologies they interact with, and be able to reason about their consequences.

Within CT this more critical perspective is gaining momentum [23, 22, 43]. For instance, Computational Empowerment (CE), as proposed by Iversen et al. [22], advocates that students should be able to recognise the ethical choices and considerations in technology and learn to decode “the consequences of these choices for the people who will use the technology”. Iversen et al. address this by having children go through a design process which includes reflecting on the more critical aspects of their designs. This design process utilises digital fabrication techniques such as 3D printing and laser cutting, and while these are effective design tools, we argue, that to deal with specific, technology-close ethical issues such as the ones described above, students need hands-on experience with the technology in question, e.g., ML. Involving students in design processes and giving them hands-on experience has been a part of CT from the start [36]. The purpose of doing so in CT has, however, been to turn technologies and computational concepts into powerful tools for the users, enabling them to approach problems in new ways and to explore new ideas. This focus on making powerful tools rarely leaves room for ethical considerations; instead CT often operates in simplistic micro-worlds [36], which are “undisturbed by extraneous questions” [36, p. 12], where ethical problems do not exist.

Inspired by CE, we explore how to involve high school students in designing ML-systems in a way that supports exploration and reflection of technology-close, ethical dilemmas inherent in ML. To do so, we design and deploy a ML Ethics Workshop, in which high school students use a ML specific set of card decks for designing ML applications. The workshop places ethical reflections at the center of the design activities and discuss technology-close ethical issues based on their own designs of ML systems. We deploy four iterations of the workshop in four different classrooms engaging 71 high school students aged 16-20. The materials used in the ML Ethics Workshop can be found here: [LINK].

The paper contributes to the fields of HCI and CT, through a) a design rationale for an ethics-first construction kit for ML, b) experiences and insights from a Constructive Design Research process [28, 27] in which the ML Ethics Workshop was designed, and finally c) a discussion of the implications of using design processes for teaching students about ML and to reflect on and discuss technology-close ethical dilemmas of its use, based on findings from the ML Ethics Workshop.

RELATED WORK

In this section we briefly review existing literature on using design as a learning approach, on card-based design methods and on involving teenagers in design processes.

Design as a Learning Approach

Design as an approach for learning about technology can be dated back to Papert’s constructionism [36, 18]. Since then, other researchers have expanded on these ideas (e.g., [24, 12]) and today, digital fabrication, Fab-Labs and maker spaces are present at schools across the world [35, 16]. In a Scandinavian context, this approach has also been explored in recent years with a more critical approach to technology [5, 41, 22, 44]. The Fablab@school.dk project [22, 41] engages students in digital fabrication with different technologies to empower them to decode and understand their future, technology-mediated world [22]. To explore this Smith et al. presents a design process model to use in educational contexts for using "digital fabrication as a reflective and material tool for working with real-life and complex societal contexts" [41]. Eriksson et al. [5] argue for the need of a wide digital design literacy, which aims to "raise awareness about decision-making in technology design, the potential impact of technology and, ultimately, whether it contributes to meaningful relationships". The authors argue that this is best achieved in design processes, where students make design-decisions for real-world settings and are able to see the implications of these decisions.

Cards-based Design Methods

In recent years, the use of card-based design methods and tools have accelerated, and today more than 155 different card-sets can be found in design literature [38]. The use of cards in design processes has been found to enhance user-driven design processes, since they can help structure and scaffold design processes[15, 34]. Cards provide a common object of interest to participants, and thus can also help engage all participants [45]. Friedman and Hendry use their Envisioning Cards [9] to bring a more human-centered focus into the design process. The envisioning cards are designed to work across different contexts and technologies, and implore designers to e.g., consider children as possible stakeholders of a system or to consider how deliberately withdrawing from using a system might affect a (non) user’s everyday life. Situation cards [31] are cards with descriptions of realistic and problematic situations at a workplace, which participants discuss in a participatory workshop to encourage them to come up with ideas, that solves everyday problems at the workplace. Another example is inspiration cards [15], which are used in design processes with disparate participants to improve engagement and support generation of innovative and realistic design concepts with new technologies, letting the participants create posters with the cards. Card-based design methods, that focus on a specific technology are few and far between [38], but there are a few examples, e.g.: Tiles is a card-based toolkit for designing internet-of-things applications [34]. The toolkit consists of different card categories and a paper-board, where participants can organise cards into meaningful IoT applications. Tiles supports reflection on users’ creations through the
We agree with Eriksson et al. [5] that involving students as products utilising computational processes and the infrastructures in which they are implemented become more complex, the ethical issues in their design and implementation also become more complex [4, 8]. Computers’ ability to process information about individuals more efficiently than ever has many benefits, but it also introduces new ethical dilemmas which often do not have clear-cut solutions [7]. Especially, because data has become important value assets to organisations, which they build their business model on and will go a great length to collect [3, 37]. Software developers, engineers and designers in these organisations are constantly making value judgements, choosing what is morally right and wrong in every decision, counterbalancing different interests and the organisation’s priorities [8, 19, 33]. These judgements are hidden in software and complex infrastructures, making them invisible for users of the organisations’ products. This invisibility factor can be exploited for unethical abuse, but more importantly it makes ethical choices and dilemmas, inherent in computational products, difficult to reason about for users [33]. When talking about ethics in this paper, we will refer to these moral decisions made by individuals and organisations, sometimes referred to in literature as micro-ethics [19]. The implications of these moral decisions have only become magnified with the propagation of ML into everyday technologies, which has been articulated in the ML community in relation to fairness, accountability, transparency, nudging etc. [1, 47].

The ML Ethics Workshop is designed with the aim of making the moral issues and decisions understandable and explorable for students. Similar to many existing CT and fabrication tools [5, 24], our focus is to make students more capable of understanding today’s technology mediated society. However, where others focus on making computational concepts understandable [13] and make technologies into powerful tools for students or co-create meaningful futures [22], our focus different. Our aim is specifically on making technology-close ethical dilemmas in ML-based computational systems understandable, and to allow students to reflect on the choices embedded in the products and services they use everyday.

**METHOD**

The work presented here uses the Constructive Design Research (CDR) methodology [27, 10, 28] to investigate how to make ethical dilemmas and choices inherent in computational artefacts understandable and explorable for high school students. We hypothesised (in the CDR-sense, see [28, 2]), that high school students would be able to have meaningful, technology-close discussions about ML ethics, by designing, reflecting on, and redesigning their own ML systems. In CDR projects, hypotheses are instantiated through the creation of artefacts [28, 2, 42], and knowledge-creation is driven by experiments with and exploration of these artefacts [27, 2]. For this work, the center artefact is the ML Ethics Workshop, including cards and other hand-outs, which is presented below. Throughout the process, we have explored how different iterations of the workshop could be used to explore the above goal and the findings presented below come from our experiences with the different workshops. As denoted by Koskinen et al. [26] CDR may be accountable for several concerns of theory and practice. In this work we are concerned with exploring and expanding the scope of CT research, with exploring how design can be used in learning situations as well as with
producing a concrete design method for making ML understandable and explorable to high school students. As such, we see the ML Ethics Workshop as a contribution in itself and as our main way for exploring the space of using design processes for teaching ML.

MACHINE LEARNING ETHICS WORKSHOP
This section describes the current version of the ML Ethics Workshop, which guides high school students through an ethics-first design process using cards and boards guide their process. The workshop is designed to be used in all kinds of high school classrooms and it is not expected, that students have a special interest in, or knowledge about ML. It has been deployed in one and a half to three hours interventions, but we believe it can scale up to a full day workshop. In groups, students design and describe a ML application to help themselves and their peers in their everyday lives. Halfway through the process students are challenged with the ethical implications of implementing their application in a real world setting. To address these implications, they must redesign their application, dealing with ethical issues, which only become more complex as they discuss them, and weight these against the functionality and experience of their application. Throughout the workshop, we work specifically with supervised, classification-based ML.

This section will first present the rationale behind the workshop, then describe the cards and boards used in the kit and, last, thoroughly describe the current format of the workshop.

Design Rationale
The workshop guides students through a design process where they confront the moral choices and ethical dilemmas of designing and implementing their own ML applications. First, students explore, through a design process, how ML can improve their own and their peers' lives. Here, students are asked to be as specific as possible in describing what data their applications uses, and what the ML component is predicting and how it is trained. This is to ensure that later ethical discussions about their ML applications can be used when making specific choices and addressing dilemmas in the design and implementation of their application. When students have a well described ML application, which they believe can help their peers, the workshop changes character to focus on the ethical issues related to the application they just designed. By discussing questions about privacy, explainability, accountability, etc. students identify the most critical ethical aspects in the design and implementation of their application. In this discussion students will identify the ethical decisions they had already consciously or unconsciously made, and what questions they must further answer in their design process, before they can morally answer for their application. To let students experience, how these ethical questions often can lead to individual value judgements or choosing between different undesirable outcomes, they are tasked with redesigning their application to approach one or more of the current most critical issues. In this task, they will have to make choices, deciding on what is most important, striving for a functional and morally accountable application, which often turns out to be difficult. Last, students present their applications, and how they have approached ethical issues in the design and implementation, followed by a classroom discussion. In these discussions there are no external 'bad guys' to blame, only students themselves and their arguments for their design choices.

Design Artefacts
Three decks of cards and a board for describing a ML system are used by the students in the workshop. The three decks of cards consist of respectively 14 data cards, 9 ethics cards, and 26 people cards, see figure 2, all sized 6X9cm and colour-coded to communicate which deck they belong to. The data deck is used to analyse a context for possible data sources, exposing to the participants how data can be found everywhere, and to provide examples of a variety of different types of data and data sources. Each card describes a category of data sources (e.g. health data, news, users locations) and provides a few examples of specific data sources in the category (e.g. pulse, breaking news, time at a location) to help participants understand the category. The people deck is used to analyse the context in which the application will be implemented with focus on people, with each card describing a potential stakeholder (e.g. a colleague, a sibling, teacher, etc.). Participants use this card deck to identify who the application may affect, and how these people can be included in the considerations about the design of the application. The ethics deck is used for reflecting on ethical implications of implementing a ML-system into a context, asking the participants to consider ethical ML issues [1] (e.g. explainability, privacy, accountability). Each card asks the participants one or two questions to help frame a discussion around the issue in relation their design and ethics, e.g. "What happens if your system makes a bad decision? Who is accountable?"

The ML board, see figure 3, is an A3 board with a visualisation of a supervised ML-model, leaving blank fields for the students to fill in when describing the ML system in their idea, forcing them to be specific about, which data their system is using and what it is predicting. Participants first describe their idea and move on to describe the ML system: What the data in their system describes (e.g. a student, a kick to a ball, a dish), which data the system learns to predict (e.g. the students grade, the precision of the kick, calories in the dish) and up to four data sources, the system will use to make these predictions (absence, accelerometer data, weight). Last, they name their model, so they can reference to it later in the design process.

The Workshop Format
The workshop supports groups of participants to design a ML system and reflect on the ethical implications of implementing it into the world. It consists of nine steps: A short introduction to ML, a presentation of a case, six activities the groups are asked to perform, a joint presentations and discussion of the participants designs. The nine steps are described beneath:

1. ML introduction: A short introduction to supervised ML learning. Students were introduced to the predictive capabilities of ML, discussed a specific use of ML i.e. using ML in democratic elections, and gained hands-on experience with ML through an interactive ML-learning tool [25].

2. Case presentation: A short description of a narrative, which frames the design process with a context, stakeholders and
an overall goal for the products the participants will be designing. It may also restrict the product design itself (e.g., it must be an app or a wearable). It situates ML in students’ own life and let them explore, how ML can solve problems, which they find important and can provide new opportunities, which are valuable to them.

3. **Analysis of context for data:** Each group choose a few cards from the data deck and use them to analyse the case context for possible data sources.

4. **Ideation:** Based on the case description and the exploration of data sources, each group conduct a short IDEO style ideation [32], where they come up with as many ideas as possible. At the end, each group are asked to choose the best idea (or combine multiple ideas into one great idea).

5. **Description of ML system:** Each group use the ML board to describe the ML component of their idea. It helps participants become very specific about their use of data and ensures all ideas are based on ML. If it is not possible for a group to fill the board, they need to go back and revise their idea.

6. **Reflections on ethics:** Each group are dealt an ethics card and the whole Human Card-deck, which they use the to reflect on possible implications for different actors of implementing their ML application into the real world, and how it will affect different actors. They choose the one issue, they find most critical about their system.

7. **Redesign product:** Each group discuss how they can address the ethical dilemmas in the design of their products and describe in detail, through sketches, texts, etc., how their system should be redesigned to address this issue.

8. **Presentation:** Each group prepare a one minute presentation, which describes their concept, their ethical issue and how they approached it in their design. Every group make their short presentation followed by a short applause from the other groups.

9. **Discussion:** The students’ designs are used for a joint discussion about ethics, grounded by the ethical problems the participants have faced in the design process and how they tried to solve them.

All groups are given a limited amount of time to complete each activity to ensure progression in the design processes. Activities are presented and described just before they start to make participants focus on the given activity and not the whole process. The internal collaboration in the groups is left informal, but mediated by the artefacts.
DESIGNING THE ML ETHICS WORKSHOP

This section shortly describes the purpose, design rationale and the insights from each intervention that influenced the design of following version of the ML Ethics Workshop. The findings from the workshop will be presented in the following section. The ML Ethics Workshop was developed in an iterative CDR process and each iteration of the workshop was deployed in a high school classroom. A total of 71 students, aged 16-20, participated in the interventions. In total, four interventions was made in four different classrooms with different backgrounds and from different high schools. All classrooms were volunteered by their teacher and participated as part of their high-school education. Each student was, however, consented individually according to local regulations. Intervention 2 took place at a local high school. All other interventions took place in a large room at the university of the researchers. An overview can be found in Table 1.

All iterations of the workshop followed the same basic structure: students were given an introduction to ML, next they were given a design case to work on and were split into groups of 3-5 people to do so. After working through different design phases, students presented their ML application and the ethical issues they had been discussing. Table 2 provides an overview of the artefacts used in each intervention. In the following, the artefacts will be referred to by their name in the table.

<table>
<thead>
<tr>
<th>Intervention</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>3 hrs</td>
<td>1.5 hrs</td>
<td>2 hrs</td>
<td>2 hrs</td>
</tr>
<tr>
<td>Location</td>
<td>University</td>
<td>Own classroom</td>
<td>University</td>
<td>University</td>
</tr>
<tr>
<td>Background</td>
<td>Technology &amp; design</td>
<td>Social studies</td>
<td>IT</td>
<td>Informatics</td>
</tr>
<tr>
<td>No. students</td>
<td>27</td>
<td>25</td>
<td>5</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 1: Overview of the different interventions as well as the classrooms whom participated in them.

**Intervention 1: Testing An Idea**

In the first intervention, we tested the first version of the Ethics Workshop, which were our initial approach on how to design a workshop which stages ethical dilemmas. The workshop was inspired by how the Tiles Toolkit [34] uses different technology and interaction cards, a board where the cards can be organised and a playbook to guide a design process, and how Envisioning Cards Toolkit [9] stages discussions about human values through focused design activities. We brought context cards with different design context and possible directions for the design process. The context were based on students’ everyday lives, e.g. sport, dating, school. To scaffold the design process, we brought six different card decks (see figure 2), post-its and A3 boards. The main idea was, that students should come up with an idea for a ML application, that was valuable to them, use the application to discuss ethical issues of ML systems and address some of the ethical issues in the interaction design of the application.

**Insights:**

Students were engaged throughout the workshop, and seemed motivated by the many cards and the possibility to come up with their own ideas. The amount of cards did, however, seem to hinder their reflections about the activities, as students were more focused on sorting and getting through all cards in each deck, than discussing a few cards in depth. The ML cards did not support the students very well in describing a ML system; they seemed to need more structure to design meaningful ML systems. As a result the ethical discussions often lacked technical depth. Students were good at identifying overall ethical issues, but they perceived their own ML applications as either morally good or bad, based on their own intentions in the design process, with nothing much to do about it.

**Intervention 2: Describing a ML system**

In the second intervention, we provided additional scaffolding for the students’ design of an ML systems through an introduction to the ML board, which can be seen in Figure 3. It visualises a supervised ML system, which students must fill in based on their ML application. Furthermore, the workshop was simplified to only contain data, ethics and people cards to make more room for immersion and reflection.

**Insights:**

This intervention was conducted with a social studies classroom, and they struggled with the openness of the design case, which was based on the context cards (see Table 2 for an example) and the requirement of using ML. Many groups spent most of the time to decide on a idea. Few groups were able to fill the ML board correctly, but the groups who succeeded were better at reflecting about the ethical issues in their application.

**Intervention 3: Combining Design and Ethical Reflection**

The third workshop explored how to make more room for ethical discussions, and how to expose the ethical choices early in the design process. To simplify the workshop further, the people deck was removed and student groups picked only two data cards and were provided with one ethics card. This was intended to make more room for discussing each card in depth and to ensure, that each group worked with different types of data and ethical problems, making the classroom discussions more diverse. To ease the ideation and conceptualisation process for the students, the context cards were replaced with a single design case about helping lonely fellow students, providing their ML application with a specific purpose with societal relevance, and limiting the design space to mobile applications. Last, students were asked to redesign their system using wireframe sketches of their application to encourage them to become more specific in their design. They were asked to identify the most critical ethical issue in their application and sketch, how they would address this in the interface.

**Insights:**

This workshop was conducted with only five students, who worked in a single group, but this version of the workshop was also used in intervention 4 with only limited changes. The more delimited design case and fewer data cards seemed to help students generate more well-described ideas, making it easier for them to fill in the ML board. The group’s ethical discussions had more depth than any group discussions in earlier workshops, and dealt with the complexity of coming up with good solutions to ethical problems.
### Table 2: An overview of the different artefacts used throughout the different versions of the ML Ethics Workshop. The right side of the table describes in which intervention each artefact was used.

<table>
<thead>
<tr>
<th>Artefact</th>
<th>Example</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethics Cards</td>
<td>See figure 2, left</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Data Cards</td>
<td>See figure 2, middle</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Humans Cards</td>
<td>See figure 2, right</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context Cards</td>
<td>“Gaming: How can [ML] be used to... - improve you performance? - create better habits? - strengthen you friendships? ...”</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML Board</td>
<td>See figure 3</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML Cards</td>
<td>“Feature: A characteristic of the phenomenon we observe. The independant variable.”</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output Cards</td>
<td>“Emoji: Use emojis to express emotional outputs.”</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Interaction Cards</td>
<td>“Smart Watch: Use a smart watch to interact with the system.”</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

**Intervention 4: Validating The Concept**

In the fourth and final intervention, we tested the composition of intervention 3 with more groups. The only change being, that the people cards were added to the kit again to support the ethical discussion in being centred around the people, who would be affected by the introduction of students’ applications.

**Data Collection & Analysis**

Designing workshops for discussing ethics in computational products in high school classrooms is a rather unexplored area. Therefore, data has been collected and processed with an open ended approach. Throughout all workshops, data was collected using observations, sound recordings, photography, video and by collecting students’ produced artefacts; products, e.g., paper-based mock-ups of their designs. All participants in the workshops were asked for their consent, and only data about consented participants was collected. After each intervention, a write-up [30] of field-notes and observations was produced and discussed between the researchers present at the intervention. The write-ups focused on, how students used the design artefacts in the design process, how they approached each step in the workshop, how they talked about and understood machine learning, and how they discussed ethics in relation to their idea. Selected audio recordings of students internal discussions were transcribed and analysed with focus on, how students identified and talked about ethical issues, and how they came to an agreement, on how to approach ethical issues in their design. Photography and video were reviewed informally with focus on, how students collaborated around the design artefacts, and how they presented ethical issues in their application for the classroom. The artefacts produced by students were analysed with focus on, how they described their ML system using the design artefacts, and, again, how they identified and approached ethics in their designs.

**FINDINGS**

In this section, we present our findings from the interventions with regards to how they were able to engage students in reflections about ethics related to ML-based systems. These findings are synthesised from an analysis across different data types and workshops. The analysis was done collaboratively between the three authors.

**Understanding of ML Served to Qualify Reflections**

Throughout different workshop iterations, we experimented with how students could be supported in describing ML systems that conceptually and technically made sense and could, to some degree, be implemented in the real world. In Intervention 1, we gave students cards that represented specific parts of a ML pipeline (e.g. features, label, training), and asked them to annotate and combine cards to describe their systems, as seen in Figure 4. This was effective to the extend that all groups used the cards to describe their ML system, but we observed that they had a hard time describing meaningful systems, and in most cases the ML aspect of their designs were described as a magic touch that was sprinkled on top. In the example in Figure 4, a group of students designed a system for optimising how their school’s rooms are scheduled to best fit the needs of different subjects, teachers etc. The group presented the ML system as being able to “predict the best schedule for the school”, but it was unclear how exactly ML would help to solve this issue, i.e., how would previous schedules be evaluated as better or worse and what data would this be based on? In the subsequent discussions on ML ethics, this became a hindrance to the students, who were only able to shallowly discuss the issues of the system. To deal with possible issues about algorithmic decision making, their system would schedule classes and study trips “in a humane and stress free way”, which is a good intention, but does not deal with the underlying technological issues or describe, how they will achieve this. In a similar example, a group designed a ML system for helping athletes improve their workouts. The group discussed whose responsibility it was, if the system recommended an exercise that caused an injury, but their solution was to employ professional testers to prevent this from happening. It was, however, unclear how these testers would
We found that by tightly controlling the design process, students were provided with the basic structure of one, and asked them to fill it in. This approach seemed to better support the students in designing conceptually and technically meaningful systems. Not all systems were sound, and students struggled especially with the specificity of data inputs and which data to predict. In contrast to the first workshop, many students were, however, able, at least to some extend, to create meaningful ML-systems. In the subsequent discussions on ethics, these students were able to have more detailed discussions about their system’s ethical issues, e.g., in the third intervention, a group designed an app for recommending social groups to lonely teenagers based on their interests. The students were given an ethics card, which questioned the responsibility of the algorithmic decisions in their system, and based on their design, they were able to have in-depth and specific discussions about responsibility:

Student 1: "What if a group, where something is going on under the radar is added to the system by us [...] and something bad happens in that club, because we didn’t know there was a problem with them. Whose fault is it then?"

Student 2: "It is still the responsibility of the club"

Student 1: "But we brought them into our system..."

Students 3: "I think, those who are responsible for the app are to blame. And I’m thinking that there is always someone in charge. That’s a director’s job"

Student 4: "Okay, that is the Minister of Health. So it should be the job of the Minister of Health or what?" (The Ministry of Health, was the owner of the application in the design case)

Student 1: "I don’t think we should take it that far up. We need a specific division to monitor groups."

[continued discussion]

Student 3: "Then we should have a rating system or the ability to report groups, and then someone should review them afterwards [...] Someone should always be to blame."

Their discussion illustrates, how it is difficult to place responsibility once something goes wrong in ML systems, and as is discussed below, these students were able to design a system that (in a basic way) addressed this issue.

**The Design Process Tied Ethics to Design Decisions**

We found that by tightly controlling the design process, students were supported in connecting their discussions about ethics to their designs, and they were able to work on using their discussions in the redesign of their systems.

This was, however, less the case in the first interventions. In Intervention 1, a group of male students designed a ML system to rate and sort women according to their breast size. This showcased (in a rather extreme way) the approach taken by many students when designing their ML systems in the first interventions. While, we provided contexts for the students in Intervention 1 and 2, many students came up with ideas, that they found interesting for themselves, but without considering, how they would impact others, and if they (and others) would actually benefit from the idea. In this way the context of "Social Media" became the breast-application and the context of "Urban Life" became "Club Counter" where an AI would recommend clubs based on its clientele. These groups did not do well in discussing the ethical aspects with regards to their designs. They could easily identify ethical issues in their designs, but it was difficult to address them in their redesigns, as they could not account for why their application should exist in the first place. Indeed, the group wanting to rate women simply answered "No" with no further elaboration to the question: "Is it ethical to implement this?" when asked to reflect on the ethical aspects of their design.

To support students in the design of systems that could be meaningfully discussed (and were not degrading) we introduced the specific case about helping lonely fellow students in Intervention 3, and asked students to consider and sketch the ways in which their system addressed their discussions about ethics. This approach seemed to better tie together the students’ design process and their ethics discussions, e.g., in the group described above, whom designed a system for recommending groups for lonely teenagers. From their discussion, they decided to implement a system for rating and reporting groups (see Figure 5). While this is a traditional (and potentially harmful) way of dealing with responsibility, the students were able to identify the issues in their discussion (see above) and design a solution addressing the issue.

In Intervention 4, a group designed a similar app for predicting "soon-to-be" lonely people and pairing them up with each other. This group discussed the ethics of data collection, and the dilemma of the app "giving away other people’s loneliness"
by pairing them with strangers (see Figure 6). This group designed the app to include a "My Data" page, in which the user can see exactly which data is given away, and can choose to remove data; they do not want the app to include in its recommendations.

**Challenging to Address Ethical Dilemmas in Design**
Throughout the interventions we found, that students struggled with addressing the ethical dilemmas they had discussed in their design. Students, who were able to have qualified and interesting discussions about ML ethics, found it difficult to come up with good solutions to their issues, and often designed systems in which the user (themselves and their peers) were responsible for system faults or data collection. Many students were trapped in a dilemma of privacy; that their application and the quality of their solutions depended on access to personal data from their users, and for groups who had discussed the ethics of data gathering, the design solutions was often akin to traditional "Terms and Conditions" agreements that most applications and websites now have (see Figure 6). As described above, some let users control what data would be included in predictions, even if this would lead to worse recommendations. In Intervention 3, the group who reflected on accountability in their application for lonely people, also discussed privacy issues and concluded that users should be incentivised to provide as much data as possible since this would provide better recommendations or as one of the students put it: "The more lonely they are, the more motivated they become to get better recommendations". Once again, this illustrates a dilemma, choosing between different interests, that is present in real-life ML applications.

**Teenagers Test Boundaries**
Throughout the interventions, we found that students would often test our and their peers’ boundaries. The system for rating women by breast size is a glaring example, and throughout interventions, many students used rather extreme examples to illustrate their considerations, such as illustrating a system failure by suggesting that a 6-year-old user would be recommended to join a sex-club or discussing what would happen if a terrorist group accidentally ended up in their application. One immediate response could be to stop the students in their track and discipline them, but we wonder if this would be appropriate, keeping in mind our goal of staging reflections. Instead we found it helpful to provide a more specific case for the students, to help them in designing systems that would better support their discussions.

**DISCUSSION**
As argued above, many existing approaches for using design processes in CT focus on empowering students to become creators [36] or to discuss broad, societal implications of technology [22]. These approaches often neglect discussions that are both close to technology and to its implications. In this section we reflect on our workshop format, based on the findings from the interventions, and discuss how it can inform future workshops or other activities with similar objectives of staging reflections on ethical dilemmas in computational systems and products.

First, we found it helpful to frame the entire design process in a way that helps students design ethically interesting systems. In interventions 1 and 2, students were to, a large degree, responsible for framing their task and were only provided the context card (see Figure 2) to help them. Starting from workshop 3, we provided students with a case about loneliness, which is an actual issue among [NATIONAL] teenagers, meaning that students were designing for a real and vulnerable demographic and were encouraged to consider how this demographic is affected by their design choices.

Second, the focus of the design process should be moved from the possibilities of a technology and towards implications of implementing said technology in real-world settings. We saw that when we supported students in designing implementable ML systems, they seemed to have more insightful discussions...
about its implications. One reason for this, we argue, might be that design processes are typically future-oriented in the sense that we as designers are interested in what-could-be. However, for a technology-close discussion of the implications of a specific technology, it is necessary to have a look at the specific here-and-now issues of a technology. In turn, this implies a change in the criteria for success of such a design process. Our inclination as designers was to look at the product of students' design processes; were they able to successfully communicate an idea or a design, did they address the case provided to them, are they solving an actual issue? These are questions that might be used to evaluate typical design processes, but as the goal has changed, we argue, so should the questions. Instead of focusing on prototyping as an end in itself, we suggest seeing insightful discussions about ethics as the goal, and prototyping as a means to ground these discussions in something concrete. We suggest asking questions akin to the following to evaluate if the design process was successful; are students able to formulate a technically feasible system; can they identify how their idea and design choices relate to specific, technology-close ethical issues; are they able to discuss these with their own idea as the point of departure; and can they address their considerations when realising or modifying their idea?

LIMITATIONS
Since students came from different backgrounds, and many students were already interested in technology their existing knowledge of ML varied. In addition, because of the differences in length of the workshops (see Table 1), students received slightly different introductions to ML and students participating in Intervention 2 had already had a longer introduction to ML [25].

CONCLUSION
In this paper, we have explored how ethical dilemmas and choices in ML applications can be made understandable and explorable for students. To do so, we designed a workshop format, in which students confront the ethical choices and dilemmas of designing and implementing their own well-intentioned ML applications. To realise this concept, we conducted four interventions with different Scandinavian high school classes, bringing a new iteration of the workshop to each intervention.

Experiences with the workshop format illustrates how an understanding of fundamental ML and iterating on own design ideas helped to qualify students ethical reflections. In addition, the design process served to reveal the complexity of the ethical dilemmas and tie them to the properties of the technology and to design-decisions. Grounding ethical discussions in students own designs, made them accountable for their choices and illustrated, how difficult it is to come up with clear-cut solutions to these ethical issues. Based on these insights, we recommend that in order to scaffold technology-close ethical discussion, design processes should focus less on technologici possibilities and more on implications and consequences, and have presented our design workshop as an approach for doing so.

We hope, that future research in CT will address the ethics in computational products and systems and aim make technology-close ethical dilemmas and decisions explorable for students.

REFERENCES


