From Demo to Design in Teaching Machine Learning

Karl-Emil Kjær Bilstrup
Magnus Høholt Kaspersen
Simon Enni
Ira Assent
Marianne Graves Petersen
Computer Science, Aarhus University
Aarhus, Denmark
{keb,magnushk,enni,ira,mgraves}@cs.au.dk

ABSTRACT

The prevalence of artificial intelligence (AI) and machine learning (ML) technologies in digital ecosystems has led to a push for *AI literacy*, giving everybody, including K-12 students, the necessary knowledge and abilities to engage critically with these new technologies. While there is an increasing focus on designing tools and activities for teaching machine learning, most tools sidestep engaging with the complexity and trade-offs inherent in the design of ML models in favor of demonstrating the power and functionality of the technology. In this paper, we investigate how a design perspective can inform the design of educational tools and activities for teaching machine learning. Through a literature review, we identify 34 tools and activities for teaching ML, and using a design perspective on ML system development, we examine strengths and limitations in how they engage students in the complex design considerations linked to the different components of machine learners. Based on this work, we suggest directions for furthering AI literacy through adopting a design approach in teaching ML.

CCS CONCEPTS

• **Social and professional topics** → *Computing literacy*; **K-12 education**; *Computational thinking*; • **Human-centered computing** → **HCI theory, concepts and models**; • **Computing methodologies** → **Machine learning**.

KEYWORDS

Design, AI literacy

ACM Reference Format:
Karl-Emil Kjær Bilstrup, Magnus Høholt Kaspersen, Simon Enni, Ira Assent, and Marianne Graves Petersen. 2022. From Demo to Design in Teaching Machine Learning. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3531146.3534634

FAccT '22, June 21–24, 2022, Seoul, Republic of Korea

ACM ISBN 978-1-4503-9352-2/22/06…$15.00
https://doi.org/10.1145/3531146.3534634

1 INTRODUCTION

Machine learning (ML) has become a pervasive technology, driving most of the impressive technological breakthroughs within the field of artificial intelligence (AI) in recent years. A child growing up with the current digital ecosystems will likely meet ML models in both public and private spheres, as ML systems increasingly advise doctors in medicine [18], play important roles in criminal justice and terrorism prevention [82], curate media intake and advertisements on streaming sites, social media, and news platforms [35], and even enter our homes in the form of digital assistants [48]. However, as the complexity and size of ML models increase to meet these diverse challenges, so do their opaqueness, and the resulting lack of understanding combined with pervasiveness and power is deeply problematic for those living in this ML-infused world, children and adults alike [12]. In addition, there is a growing concern regarding how the opacity and complexity of ML systems can lead to systemic racial and gender biases [72], and unfair decision processes [5].

These concerns have driven a range of research, e.g., to improve the transparency of ML models [21, 78], increase accountability [1, 22], ensure fairness [24], protect privacy [17], increase robustness [63], and to give users of ML-based systems opportunities for understanding the underlying computational processes in order to trust them and feel in control [1].

In addition, these concerns have led to an increased focus on how ML systems are part of larger social and political systems [2, 44]; how concepts such as fairness and ethics should be understood and measured in the context of these larger systems [67, 73]; and how ML system designers' decisions must be communicated and discussed to make them accountable for how their systems impact these larger systems [40].

In line with the above concerns, criticism has been raised regarding existing ML curricula in higher education for not engaging with the ethical issues of ML [6, 60] and suggestions have been made for new ways to teach implications of ML (e.g., [68]). However, having ML system designers engage with the social implications of ML is not sufficient to ensure more fair and just systems. Since we are all (in one way or another) users of, or at least affected by, ML systems we are all part of judging how and when a system is beneficial, and to whom. This calls for a new, broad *AI literacy* which develops users' critical reflection and understanding of these systems. One approach to do so is through addressing AI literacy in kindergarten to 12^{th} grade (K-12) education, and researchers have recently begun exploring what an AI literacy curriculum for K-12 education should look like [47, 74].

With the high stakes outlined above, it is clear that we need to improve the understanding of ML systems also for those using them and those who are affected by them. To address this, we look to the *Computational Thinking* [80] (CT) research community, which provides an influential perspective on the teaching of computing in general in K-12 education around the world [36]. This community has emphasized how thinking like a computer scientist is more than understanding computational *concepts*, but also involve understanding and development of computational *practices* and *perspectives* [10].

In this paper, we align with the more critical voices in the CT community (e.g., [19, 39, 41]), who argue that in order to empower and engage students, they should be taught how to *decode* the intentions, choices and trade-offs embedded by developers and corporations in technological artefacts and systems more generally, and how to *code* their own intentions and choices into such artefacts and systems [19]. In line with this, we argue that to truly understand ML systems, students must engage not just with the technical components of ML models but also with their design, the rationales behind their construction, and the context of their use and deployment.

A critical question then becomes how ML can be conceptualised to scaffold students' engagement with, design of, and reflection about ML. Existing popular tools for teaching K-12 students about ML and AI, such as Teachable Machine [15], Cozmo[1], ML4kids[2], sidestep complexity and trade-offs inherent in the *design* of ML models in favor of *demonstrating* the power and functionality of the technology. It enables students to explore the capabilities of ML systems and even to use ML to design and construct their own systems, but rarely expose the complex and difficult design questions and considerations in the process of constructing and deploying ML systems in the real world, such as designing and representing training data sets or choosing one model type and architecture over another. Thus, we argue that without involvement in the *design* of ML systems, it will be difficult for students to understand the trade-offs and rationales embedded in such systems in addition to understanding the role of ML systems in their wider context of use.

This argument is well in line with recent research in the field of Explainable Artificial Intelligence (xAI), promoting evaluation criteria which also focus on real users and tasks [21]. Notably, this re-orientation of xAI aligns well with how the field of Human-Computer Interaction (HCI), over time, has been increasingly committed towards real users, real tasks, and the broadening of use-contexts from work to everyday lives [9], and as a result has embraced design as a discipline in the field [13, 81].

This paper addresses how to develop children's AI literacy in K-12 education. We do so by presenting a *design approach* to teaching ML that complements the *demo approach* outlined above. First, we outline what a design approach means for teaching ML by presenting and discussing design literature and design models characterising the ML process. Next, we review existing literature on activities and tools for teaching ML with regards to how they engage students in design questions within ML. Based on this review, we discuss our approach and present a set of sensitising concepts[3] [8] aimed at helping K-12 educators and designers of tools and activities for teaching ML with directing their work. Thus, the paper makes the following contributions:

(1) A framing of ML development as a design endeavour and the consequences of this for teaching ML in K-12 education.
(2) Sensitising concepts for a design approach to teaching ML in K-12 education, based on a review of existing literature in conjunction with the framing above.

In summary, we propose a shift in the way we teach ML from *demo* to *design.*

2 TOWARDS A DESIGN APPROACH TO MACHINE LEARNING

Before turning to teaching ML, we take a step back and investigate more specifically what a design approach to machine learning implies. For this we first look towards the field of HCI which, over time, has embraced lessons from the discipline of design in order to strengthen its approach to dealing with users, implications, choices, and rationales in the construction process of interactive systems. Next, we look towards the field of machine learning to understand how it has approached ML as a design process.

2.1 Learning from HCI

In the following, we highlight three sets of fundamental assumptions and approaches within the discipline of design, as it has come to inform the field of HCI more generally, and which seem particularly pertinent to the challenges facing AI literacy.

2.1.1 A contextualized approach. One of the hallmarks of design is the creation of systems that are useful for people in their everyday lives [81]. Fundamentally, design is dedicated to the construction of objects and systems which really work for people, producing quality results and satisfying experiences. Moreover, design is dedicated to a holistic perspective of understanding systems as they are used in their contexts, by users with needs and values. Users are conceived of as those who directly use a system, but also as those indirectly affected by its use and those who make decisions regarding the acquisition and use of the system [25]. Thus, constructing systems from a design perspective implies a dedication towards understanding systems in their full context of use. Applying this to ML, we can term this a *contextual approach* to the construction and evaluation of ML models. This view resonates with recent arguments put forth in xAI which promotes the integration of social sciences in our understanding of the role of explanations [53] and the recent calls for applying situated and socio-technical frames in ML system design [44, 67].

2.1.2 Exploring alternatives and identifying rationales. Historically, people such as Winograd and Buxton have been highly influential in bringing the lessons from design to bear on how we build software [13, 81]. These initiatives have been motivated by concerns regarding how design of software previously focused primarily on incremental improvements of systems (getting the design right), rather than exploring alternatives (getting the right design) [13].

[1] cozmo.com
[2] https://machinelearningforkids.co.uk/
[3] See section 5 for an explanation of the term

What is often emphasised from the practice of design is the role of reflection in action - the need for designers to work in cycles of creating and reflecting [66] and rather than working towards one solution, working in an iterative process of exploring alternative solutions and evaluating these [13]. For ML, this perspective means *exploring alternatives* and *identifying rationales* for choosing one solution over another. Recent research has highlighted how ML system designers should be confronted with the normative choices in their systems [54], and how the rationales behind a ML system must be made explicit to hold the system (and the system's designers) accountable for its actions [40].

2.1.3 Exploring possible futures. Design is a matter of "intentional change in an unpredictable world" [57]. It is emphasised how the role of design is not to predict the future, but rather to *explore possible futures*, and ask what-if questions [23]. The field deploys a recurring notion of possible, plausible, probable, and preferable futures [23], emphasizing the need to investigate and articulate for whom a solution is a preferred future and in what sense. In addition, *speculative* design [23] is particularly concerned with broadening this debate beyond industry and politicians towards citizens capable of partaking in such debates. Finally, within the discipline of design, there is also a rich discourse on how the designers' intentions influence design. Verbeek [76] highlights how designers' *intentions are inscribed in technologies*, and how technologies mediate how users perceive the world and steer how they act in the world, by inviting certain uses and inhibiting others. These arguments align with recent criticism of attempts to discuss ML technologies as neutral and efforts to 'de-bias' ML models and datasets [73]. Bias and intentions are inherent entities in ML models and systems, hence they should be understood and discussed in ethical and political terms.

To summarize, what we take from the design perspective to reframe how we can engage with teaching ML includes the following three design considerations:

(1) Taking a contextualised approach to the construction and evaluation of ML systems and considering the expected context of use.
(2) Exploring alternatives and emphasizing design rationales throughout the ML design process.
(3) Being mindful of how intentions are embedded in ML systems and reflecting on future implications.

In the following analysis, we explore how these design considerations can be investigated more particularly in the process of developing Ml systems. But first, we outline the current state of the art in the intersection of ML and design and present a recent design model that will guide our analysis.

2.2 Machine Learning as Design

In this section, we look towards the field of ML for perspectives which can serve to inform a design approach for teaching ML. While ML has a strong theoretical background in statistical learning theory, making ML systems work in practice often involves a large degree of tacit knowledge and discretionary action [59], that some derogatorily have termed 'black art' [20] or 'folklore and magic spells' [14]. This critique implies that ML is often approached

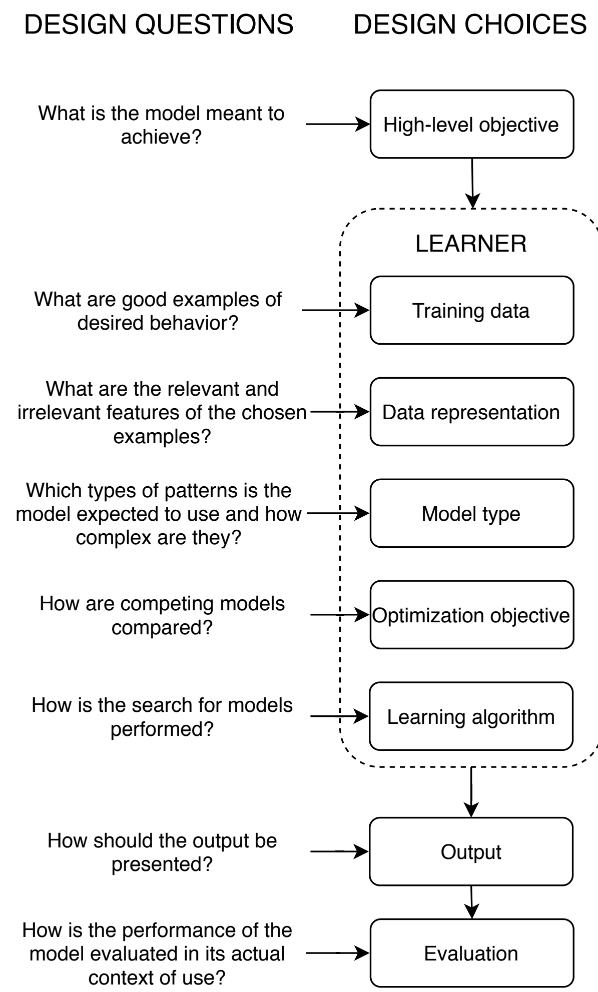

Figure 1: Design model of the different components in creating a machine learner along with the questions that guide this design. From [26].

without the critical reflection suggested by a design-oriented perspective. As a response, a number of recent papers in the ML field suggest new ways of reflecting on and documenting the ML process [45, 52, 55, 59, 61].

In one of these recent papers, Enni and Assent [26] present the design model for ML seen in Figure 1). It consists of eight separate components of machine learners[4] that have to be implemented in a ML project. This model furthermore articulates a set of design questions which are addressed (either explicitly or implicitly) in the choices made for each of these components, and invites reflection on how these choices impact the behavior of the resulting ML model. The questions emphasize how model design and -evaluation must be understood with respect to its context of use, how choices and trade-offs must be made in the particular design of the machine learner,

[4]Enni and Assent distinguish between machine *learners*, which are programs, typically coded in high-level programming languages such as Python, and the ML *models*, that the learner programs produce as they train on data, which are often represented as computer generated data files or collections of learned model parameters. [26].

and how desired behavior and relevant features are examples of intentionality being embedded in the design of the machine learner. As such, Enni and Assent's design model serves to operationalise a design perspective in terms of the particular details of creating a ML model.

Enni and Assent's design model is focused on the way machine learners are designed around knowledge of the statistical relationship between input and output data modelled by the ML model. It acknowledges that ML solves problems not by explicitly programmed computational solutions, but by designing a learner to learn a solution from examples. Thereby the design of the machine learner is the indirect design of the ML model. Thus, their design model helps articulating how ML requires not just computational thinking but also design, which is reflected in the model's design questions; each motivated by the effect it is ultimately expected to have on the ML model [26].

As we analyze the tools for teaching ML covered in the following sections we will use the model illustrated in Figure 1 as a key to investigate which parts of the ML design process these tools engage with.

3 REVIEW OF ML TEACHING TOOLS & ACTIVITIES

To investigate how the design perspective presented above can inform the design of educational tools and activities for teaching ML at a K-12 level, we have conducted a systematic literature review on existing research on such tools and activities. The aim is not to provide a comprehensive report on the state of a well-established research field but to provide a snapshot of this emerging topic which can be used to inform future directions for ML teaching. By using the design model in Figure 1 as a key perspective, and systematically analysing existing work with the design perspective, we examine:

(1) Which ML design components existing tools and activities engage students in.
(2) How existing tools and activities conceptualize the ML design questions linked to each component.

Below, we account for our review method, followed by a discussion of the results from the analysis.

3.1 Review method

Our literature search started out quite broadly using the SCOPUS database by Elsevier[5] which includes publications in AAAI (including the symposium on education) and ACM, as well as other key venues such as the International Journal of Child-computer Interaction (IJCCI) and International Journal of Artificial Intelligence in Education (IJAIED). The search was conducted in August of 2020 and yielded 1947 results. The search query can be found in Figure 2. To retrieve papers with a main focus on ML, we searched for both ML and AI in the title. Both terms were used, since they are used interchangeably in some cases. Next, we searched for mentions of K-12 students (and synonyms) in the title, abstract and keywords to find papers focusing on teaching such students. Last, since the

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

TITLE ("Machine learning" OR "ML" OR "AI" OR "artificial intelligence") AND TITLE-ABS-KEY ("children" OR "young" OR "Youth" OR "students" OR "pupils" OR "k-12" OR "kids" OR "teenagers" OR "teens" OR "youngsters" OR "adolescents") AND (LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "ENGI"))

Figure 2: Search query used in the SCOPUS database for the review of existing ML learning tools.

SCOPUS database searches for publications between several disciplines, we limited the search to computer science and engineering to find publications focusing on the design of educational tools and activities. Computer Science, as a subject category in the SCOPUS database, also covers CS Education research venues and journals such as Computers and Education, and International Journal of Educational Technology in Higher Education.

To review these publications, we underwent a systematic process following three rounds of elimination and a final step of data analysis and coding of selected papers: In step (1), we eliminated papers that were not scientific publications in English, and papers that did not focus on teaching ML to K-12 students, by reviewing the meta information (title, abstract and keywords) of the 1947 papers (153 publications remained). For step (2), we retrieved the full texts of the remaining papers and eliminated papers using the same criteria as in step 1 but based on the the full texts (76 publications remained). In step (3), we eliminated all papers that did not describe *activities or tools* for teaching K-12 about ML. This resulted in 34 publications published from 1987 to 2020. Finally, in step (4), the papers were analyzed and coded based on the questions 1 and 2 from above.

The review process was conducted by two of the authors. Following the elimination rounds, in order to reduce individual biases in the analysis, ten papers of the final 34 were divided between the two authors who separately analysed the papers. To align the analysis, four authors discussed the ten papers and reached consensus. Finally, the two authors divided the remaining 24 papers between them. The findings from the analysis are presented in the next section.

4 FINDINGS

In this section, we unfold our analysis of the tools reviewed above by going through each designed component in Enni and Assent's [26] model in Figure 1. Table 1 provides an overview of which components were addressed in the analysed papers. As can be seen in Table 1, overall, the analysis shows a tendency to emphasis the beginning and end of the ML design process. In the following, we unfold and discuss how the components are addressed in the papers. For each component we a) unfold the design question, b) analyse how design perspectives are addressed in current tools and activities, and c) reflect on how they can be addressed in future tools and activities.

ML component	n=	References
High-level Objective	16	[4, 7, 11, 15, 27, 31, 33, 56, 62, 64, 69, 70, 75, 79, 83, 84]
Training Data	21	[3, 4, 15, 16, 30, 31, 33, 37, 38, 42, 46, 49, 56, 58, 64, 69, 75, 77, 79, 83, 84]
Data Representation	9	[4, 7, 16, 27, 30, 49, 58, 62, 69]
Model Type	9	[4, 16, 28–30, 46, 49, 58, 77]
Optimization Objective	1	[28]
Learning Algorithm	1	[15]
Output	16	[3, 4, 7, 15, 34, 37, 38, 42, 49, 64, 69, 75, 77, 79, 83, 84]
Evaluation	10	[3, 15, 27, 31, 33, 34, 42, 56, 83, 84]

Table 1: Review of how existing literature engage students in in the components of ML. It illustrates how existing educational tools and activities mainly focus on components related to the input and output of ML learners and less so on the inner components.

4.1 High-level Objective

The initial design choice in creating a machine learner is the high-level objective for the resulting model [26], and designers need to ask themselves *what the model is meant to achieve?* From a design perspective, this entails investigating the context in which the model is to be deployed, whom its users and stakeholders are, and in turn what their goals are. Is ML the correct choice for this situation, or are there more appropriate alternatives?

16 out of the 34 papers address the high-level objective of machine learners. A few of these let students explore the context in which a model is to be employed, most often as part of a *design case*, in which students go through a design process for developing ML systems. This is achieved through e.g., card-based workshops, where students are given different cards representing stakeholders, data sources, interaction methods, etc., and design their own ML system [7]; or by engaging children in designing a "fair" AI librarian, through exploring the context and discussing what fairness means [70]. Other papers use predetermined objectives, where students create ML models to predict e.g., athletic moves for use in high-school physical education [84].

While we find good examples among the 16 papers, where they effectively contextualise ML systems, there is a lack of engagement with the rationales behind the systems, and with exploration of different alternatives. This would allow students to consider pros and cons of different objectives, and to make deliberate decisions as to how a potential ML system should be implemented. Last, it could support students in reflecting more generally on when and why ML is a good solution to a problem.

4.2 Training Data

Training data should provide diverse examples which adequately represent the problem space in which the model will be used. This emphasises the context in which the data is collected and the context in which the ML system will be implemented. Designers must ask themselves, given the intentions of the system, *what are good examples of desired behavior* that we can show to the learner and how do we sample this data?

Training data is explored in 21 of the papers with great diversity among activities and tools. Some papers use computer-generated data (e.g. [28]), others use mock-data tied to real-world artefacts such as screws, cookies, smileys (e.g. [46, 77]) which give some meaning to the data but without providing a context (e.g., why are we sorting cookies?). There are also examples where students generate data themselves through e.g., gestures [84].

We find examples among the papers where they take a contextualised approach and have students engage directly with generating and collecting examples of desired behavior. This provides the data with a context for students to explore, and highlights the subjectivity of the data which can support students in rationalizing about the credibility and limitations of data. However, most of the papers instruct students very specifically about which data to collect and how to do it, whereas reflections on the critical design questions in this process are rarely scaffolded in the tools and activities. We see a need for engaging students in questions such as: What are good examples of desired behavior in the given context? How can they be collected? And is the data-set representative of the problem it describes?

4.3 Data Representation

The 'examples of desired behavior' identified in the previous step must be represented in a way that highlights what is important and what is not to model the problem accurately and as intended by the designer. The designer must ask, depending on the context *what are the relevant and irrelevant features of the chosen examples*, and how can they be represented to illustrate this?

Of the nine papers which engage students in data representation, a few papers use simplified visual examples to illustrate how data can be represented in different ways, such as geometric objects, screws, or cookies (e.g [49]). Others provide more contextualized examples, like choosing the most important factors in a wildlife environment [69] or physical attributes of basketball players [4]. These activities are used to scaffold discussions about which features may be important to make accurate models and how different representations can change ML models' performance.

The nine papers do, however, put little emphasis on differences between features (other than the model's performance), how such information relates to those it represents, and how some features may be more sensitive than others. Such considerations could be included in students' choice of features and when comparing different models. One model may be preferable to another because it is built with less sensitive or more accessible data. Last, activities and tools could scaffold reflections on fundamental questions about which information is worth collecting to build a given model, and if certain application areas have features that should generally be avoided, such as race, gender, and religion when modeling human behavior.

4.4 Model Type

For this component, designers must explore different model types and architectures, asking *which types of patterns is the model expected to use and how complex are they?* Choosing to use ML to solve the problem at hand implicitly assumes statistical dependencies, and the designer must consider how the model should generalize beyond the finite training data. Choosing the wrong model type has major implications that may result in the model relying on spurious correlations that are not stable over time or can result socially discriminatory generalizations.

Nine papers engage students with model types and, in some cases, let them tinker with hyperparameters (e.g. [49, 77]). One approach, is to engage students in comparing and discussing the pros and cons of different model types [28]. Another approach, is to have students inspect data-sets and identify patterns, which they can use to make assumptions about the task [46, 77]. These papers, however, provide no or little context, and reflections and discussions are mainly centered around optimization of model performances. One paper focuses on xAI and has students creating explainable models in the form of decision tress, and inspect why their models make wrong predictions [4].

The nine papers only briefly address design implications for this component. Some of the tools and activities let students compare different model types but focus only on the technical differences between the models and do not explore advantages and disadvantages of each model type from a broader, contextualised perspective. Comparing model types can scaffold students' reflections about how different features are related, how difficult this relation is to model, and how increased complexity may increase the predictive power of a model but may also make the model overfit to the training data. A simple model that is more robust, easier to train, and more transparent may, depending on the context, be preferable to a more complex powerful model, even at a cost in performance. These considerations could constitute the basis for reflections on and discussions about the possibilities and limitations of ML models among students.

4.5 Optimization Objective

For this component, the designer must consider what a successful model is and ask; *How are competing models compared?* It is necessary to formulate an optimization problem to train a ML model, and this is typically achieved by designing a loss-function on the training data that measures the total error of the model.

We found only one paper which presents students to loss functions and let them reflect on and reason about them. In the paper, students use linear and polynomial regressions, and calculate the root mean square error, while comparing it to other definitions for calculating errors [28].

We do, however, see opportunities for engaging students in the design questions related to this component. Prioritizing one type of error over another, or setting up a hard constraint on a measure incurs trade-offs with different implications for different contexts that need to be discussed and evaluated. Take, for example, some methods for avoiding discrimination and unfairness in ML models which alter the optimization objective to favor fairness in the objective function [51]. However, such interventions typically come with a cost in predictive accuracy. A number of salient normative discussions about the intentions and contexts of the ML model could emerge as a result, which could become the focus of ML teaching exercises at the K-12 level.

4.6 Learning Algorithm

Searching for the best possible model is an important component in creating a ML model, and choosing how to do it is dependent on the choices made in the other components. The designer must ask *how is the search for models performed?* This includes setting up the algorithm properly and considering which resources are available to perform this search.

We found one paper which let students change the learning rate of a neural network in order to explore the effects on the resulting model [15].

Reasoning about the best choice among many possible learning algorithms requires competencies that cannot be expected of K-12 students. However, the merits of the different algorithms and the settings of their hyperparameters could potentially be explored in a trial-and-error setting, where students choose and tweak algorithms to observe the results in the behavior of the final model, much like engineers do in ML practice. Such tinkering could impart important lessons about over- and underfitting as well as the logic behind common optimization algorithms such as gradient descent. In a similar vein, tinkering with learning algorithms might spur discussions of the material and environmental cost of training ML models, thereby revealing important and often forgotten issues in the ML discourse [71]. Considering the increasingly dire prospects of global warming, such discussions could be an important addition to the classroom.

4.7 Output

Deploying machine learning in real world contexts is challenged by the difficulty of understanding its reasoning [1]. Thus it is important to consider the context and stakeholders who are affected by a ML model when designing its output. Opaque ML models run the risk of leaving users feeling disempowered and make it next to impossible for affected stakeholders to question their influence in decisions made about them. On the other hand, users might feel overwhelmed if confronted with the full complexity of a ML model. Thus, designers must ask themselves *how should the output be presented?*

15 papers engage students in analysing and considering the output of ML models. In many cases, students are shown only the classification result (e.g. [49, 84]), while some also expose more detailed information such as the probability of a classification (e.g. [15]). There are also more contextualized examples, where students explore how the output of ML models can be integrated in interactive systems (e.g, in [65, 69, 84]). For example, the system Any-Cubes allows users to control an actuator using image recognition by e.g., starting a motor when a toy giraffe is shown to the camera [65].

However, none of the papers engage students in considerations about the implications different output types have on how users and systems perceive a ML model's predictions in real-world deployments. E.g. some systems might actively try to hide how ML works behind the scenes to make an interaction as fluid as possible, where others might wish to make processes behind the decisions more transparent to allow for inspection. Such considerations could spur and inform discussions, among students, about the role of ML systems in society and their personal lives.

4.8 Evaluation

In this last component, designers must ask *how is the performance of the model evaluated in its actual context of use?* Deciding on evaluation metrics for determining a models performance in use is not a straightforward task. The context of use is important to determine how a model should be evaluated, e.g., if a model is used to aid medical practitioners with cancer detection, merely measuring predictive accuracy is not appropriate, as the number of negative examples (i.e. those without any cancer) vastly outnumber the number of positive examples. Furthermore, getting a testing data-set of high quality requires intimate knowledge of the application area.

Of the 10 papers which engage students in evaluations of ML systems, a few let students experience how a model performs by playing with its application afterwards, e.g playing rock-paper-scissor against the computer [43] or test an RC car's ability to navigate on a track [56]. There are also more contextualised examples, where students test a ML system in its intended use setting, for example by letting students build models for improving athletic moves which can be used in practice by the students [84].

However, the way evaluation is addressed in the 10 papers is narrow in scope as they do not engage students in considering different stakeholders, deeper reflections on what constitutes a 'good' system or in having deeper reflections on different evaluation metrics in relations to different application contexts. In general, future educational tools and activities could have more focus on how ML systems are tested in real-world contexts and how users experience the systems.

5 DISCUSSION

Here, we discuss how the findings above can constitute a design approach for teaching ML in K-12 education, what educators should be mindful of when taking this approach, and what future work is needed to further substantiate it.

Across all components from the ML model in Figure 1 existing tools and activities mostly focus on demonstrating how ML is used and on creating the best performing model possible. Furthermore, as seen in Table 1, they mainly focus on components related to the input and output and less so on the inner components such as optimization objectives and learning algorithms. While most papers focus on the different components from a technical, demonstrative point of view and skim over the social contexts and the decisions related to these (e.g. [30, 32]), we do find some examples where choices about training data, representation, etc., are contextualised, and where students explore alternatives (e.g. [4, 28]). However, they are rare and do not emphasize the design rationales and intentions behind, nor the future implications of the ML systems, which students are engaged in constructing. The examples we found that did engage students in how ML systems intervened in existing social context and the ethical choices connected to implementing such systems (e.g. [7, 70]) mostly did not engage students, or engaged them only to a limited degree, in the inner components of ML systems. As we have argued, it is crucial that students, to some degree, understand and engage with the inner components in order to decode [19] the intentions hidden deep within ML systems.

Acknowledging that developing and designing ML systems is more than understanding and applying computational and mathematical concepts, but also includes engaging with and intervening in complex socio-technical systems [2, 44] and making ethical choices [73] poses new challenges from an educational point of view. From a design perspective, all components should be addressed with attention to these concerns, as the design of every one of them includes value-judgements and considerations with implications for how the system is used and how successful it is. If students are not engaged in these technical aspects and choices in ML systems they risk not being confronted with the difficult dilemmas and choices inherently encountered when developing these systems. The challenge thus becomes how to design educational activities and tools which tie these different ends together such that they expose the full complexity of ML systems in a comprehensible manner. To address this, we argue that design has an important role to play. By combining the model in Figure 1, which splits ML development into a set of designed components, with the understanding of design as described in section2, we arrive at a design approach for teaching ML which can expose students to the concerns with ML systems described in section 1.

To help those educators and designers of educational tools and activities for teaching ML, who want to apply this design perspective, with directing their work, we present a set of sensitising concepts. They are not intended to provide a prescriptive framework along which to work, but instead to *"suggest directions along which to look"* [8, p. 148]. Table 2 presnts our sensitising concepts which synthesise the emphasis of a design approach based on the three design considerations from section 2 and formed by our analysis of current tools and activities. The design-oriented approach favors reflexive engagement with the design of ML models over demonstrating their power and use, and thus it shifts focus from describing *what* the system does to reflecting on *why* and *how* this happens while suggesting alternatives. Rather than providing answers up front, the design-oriented approach aims to ask questions that allow students to interrogate ML systems. Rather than presenting ML as the solution to a given issue, the design approach asks students to explore said issue, engage with trade-offs, and reflect on the consequences of applying ML. Instead of presenting ML models

Demo	Design	
Answer	Question	(1) Taking a contextualised approach to the construction
Abstract	Contextualise	and evaluation of ML systems and considering the
Demonstrate power	Expose limitations	expected context of use.
Describe	Suggest	
Make easy	Expose complexity	(2) Exploring alternatives and emphasising design rationales
Improve performance	Engage in trade-offs	throughout the ML design process.
Solve problems	Explore futures	
Find solutions	Make judgements	(3) Being mindful of how intentions are embedded in ML
Pursue objectives	Reflect on consequences	systems and reflecting on future implications.

Table 2: A table contrasting a demo-oriented approach to teaching ML with our suggested design-oriented approach. The design approach is described though three design considerations and nine sensitising concepts which suggests directions along which to look for educators and teachers who want to apply the approach.

as generic solutions that might be used to solve any number of issues, it suggests that focus is put on the contextual nature of constructing ML models and the many different choices engineers and ML practitioners make underway. Finally, instead of making ML "easy" by constructing scenarios in which success is a matter of improving performance, we suggest that teachers create scenarios that expose the complexity of the ML process and explore its limitations, without overwhelming students with technical details.

Importantly, the design approach that we suggest is not a departure from the demo-approach but an addition. As numerous researchers have shown in the work we have reviewed, the demo approach is effective in teaching and getting K-12 students excited about ML. Instead, the design approach is a suggestion as to how we might educate the public to take agency in dealing with the issues that has followed and will continue to follow in the footsteps of ML's success, and which we will need to deal with eventually.

With this approach, we acknowledge that is necessary to hide away, or 'black-box', aspects of ML to not overwhelm students [50], and future work should engage with where to cut the cake, so to speak. We suggest that deciding on which parts of the system to black-box should be done with reference to a ML design model such as the one in Figure 1, using the design questions to engage with the remainder of the system. Some aspects and design choices in ML systems may be more difficult to reflect on for novices and different age groups, and this may require priming students with a conceptual understanding of, e.g., computational optimization for the optimization objective and learning algorithm components or basic statistics for the evaluation component.

In summary, we suggest a *reflexive* approach to teaching ML that empowers students to engage critically with ML to complement the commonly used demo-oriented approach that presents a decontextualised and uncritical view of the power of ML. Through this strategy, the agendas of AI literacy with a critical focus on engaging students in how technology form their everyday life [19] are more likely to succeed, as students participate directly in the design and deliberation of ML development and, thereby, achieve the knowledge and perspectives necessary to engage critically with ML going forward.

6 CONCLUSION

This paper proposes a design approach to how ML is conceptualized and taught to support AI literacy in K-12 education. We recognize the value of *demoing* capabilities of ML when teaching ML, but see a lack of focus on exposing the *design* questions involved in making ML systems as well as the complexity and trade-offs inherent in real-world ML systems. Based on design and HCI literature, we have identified three design considerations that could help expose the complexity of, and concerns related to, the design of ML systems: *1)* Taking a contextualised approach to the construction and evaluation of ML systems and considering the expected context of use; *2)* exploring alternatives and emphasizing design rationales throughout the ML design process; and *3)* being mindful of how intentions are embedded in ML systems and reflecting on future implications.

With this perspective and in combination with the ML design model in Figure 1 which conceptualize the ML process as a set of designed components, we have reviewed how existing research-based tools and activities engage students in learning about ML. We found that most tools take a demo-approach that prioritises showing the power of ML over exposing students to the complex decisions that go into making, deploying and maintaining a successful ML system. We suggest a complementary approach based on design. Here, instead of demonstrating the power of ML, we aim to expose its design process. We aim to suggest rather than describe, asking why and how rather than what, question rather than answer, explore futures rather than solve problems, engage with the context rather than presenting ML systems as generic solutions, exposing the complexity of ML rather than making it easy, and exploring its limitations rather than improving its performance.

We hope that this approach will help educators and designers of tools and activities for teaching ML in empowering future generations when engaging with, using, or being used by ML systems.

ACKNOWLEDGMENTS

This work was supported by a research grant (#28831) from VILLUM FONDEN.

REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible

Systems: An HCI Research Agenda *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3173574.3174156

[2] Rediet Abebe, Solon Barocas, Jon Kleinberg, Karen Levy, Manish Raghavan, and David G. Robinson. 2020. Roles for Computing in Social Change. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 252–260. https://doi.org/10.1145/3351095.3372871

[3] Adam Agassi, Iddo Yehoshua Wald, Hadas Erel, and Oren Zuckerman. 2019. Scratch nodes ML: A playful system for children to create gesture recognition classifiers. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery. https://doi.org/10.1145/3290607.3312894

[4] Jose M. Alonso. 2020. Teaching explainable artificial intelligence to high school students. *International Journal of Computational Intelligence Systems* 13, 1 (2020), 974–987. https://doi.org/10.2991/ijcis.d.200715.003

[5] Solon Barocas and Andrew D Selbst. 2016. Big data's disparate impact. *California Law Review* 104 (2016).

[6] Jo Bates, David Cameron, Alessandro Checco, Paul Clough, Frank Hopfgartner, Suvodeep Mazumdar, Laura Sbaffi, Peter Stordy, and Antonio de la Vega de León. 2020. Integrating FATE/Critical Data Studies into Data Science Curricula: Where Are We Going and How Do We Get There?. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 425–435. https://doi.org/10.1145/3351095.3372832

[7] Karl-Emil Kjær Bilstrup, Magnus H. Kaspersen, and Marianne Graves Petersen. 2020. Staging Reflections on Ethical Dilemmas in Machine Learning : A Card-Based Design Workshop for High School Students. In *DIS '20: Proceedings of the 2020 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, United States, Eindhoven, Netherlands, 1211–1222. https://doi.org/10.1145/3357236.3395558

[8] Herbert Blumer. 1969. *Symbolic interactionism*. Vol. 50. Englewood Cliffs, NJ: Prentice-Hall.

[9] Susanne Bødker. 2006. When Second Wave HCI Meets Third Wave Challenges *(NordiCHI '06)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/1182475.1182476

[10] Karen Brennan and Mitchel Resnick. 2012. New frameworks for studying and assessing the development of computational thinking. In *Preceedings of the 2012 annual meeting of the American Educational Research Association*. Vancouver, Canada, 1–25. https://doi.org/10.1.1.296.6602

[11] Janine Bressler and Janett Mohnke. 2015. School laboratories to teach robotics, smart home and artificial intelligence: From theory to practice. In *ICSIT 2018 - 9th International Conference on Society and Information Technologies, Proceedings*. 122–125.

[12] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 3, 1 (2016), 2053951715622512.

[13] Bill Buxton. 2007. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[14] Alexander Campolo and Kate Crawford. 2020. Enchanted Determinism: Power without Responsibility in Artificial Intelligence. *Engaging Science, Technology, and Society* 6 (2020), 1–19.

[15] Michelle Carney, Barron Webster, Irene Alvarado, Kyle Phillips, Noura Howell, Jordan Griffith, Jonas Jongejan, Amit Pitaru, and Alexander Chen. 2020. Teachable Machine: Approachable Web-Based Tool for Exploring Machine Learning Classification. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3382839

[16] C. Casey, A. O. Adewumi, A. Hart, and G. O'Hare. 1987. Experience in introducing sixth formers to concepts of artificial intelligence. *Journal of Computer Assisted Learning* 3, 1 (1987), 13–22. https://doi.org/10.1111/j.1365-2729.1987.tb00304.x

[17] Daizhuo Chen, Samuel P Fraiberger, Robert Moakler, and Foster Provost. 2017. Enhancing transparency and control when drawing data-driven inferences about individuals. *Big data* 5, 3 (2017), 197–212.

[18] Po-Hsuan Cameron Chen, Yun Liu, and Lily Peng. 2019. How to develop machine learning models for healthcare. *Nature materials* 18, 5 (2019), 410.

[19] Christian Dindler, Rachel Smith, and Ole Sejer Iversen. 2020. Computational empowerment: participatory design in education. *CoDesign* 16, 1 (2020), 66–80. https://doi.org/10.1080/15710882.2020.1722173 arXiv:https://doi.org/10.1080/15710882.2020.1722173

[20] Pedro M Domingos. 2012. A few useful things to know about machine learning. *Commun. acm* 55, 10 (2012), 78–87.

[21] Finale Doshi-Velez and Been Kim. 2018. Considerations for evaluation and generalization in interpretable machine learning. In *Explainable and Interpretable Models in Computer Vision and Machine Learning*. Springer, 3–17.

[22] Finale Doshi-Velez, Mason Kortz, Ryan Budish, Chris Bavitz, Sam Gershman, David O'Brien, Stuart Shieber, James Waldo, David Weinberger, and Alexandra Wood. 2018. Accountability of AI Under the Law: The Role of Explanation. *Privacy Law Scholars Conference* (2018).

[23] Anthony Dunne and Fiona Raby. 2013. *Speculative everything: design, fiction, and social dreaming*. MIT press.

[24] Cynthia Dwork et al. 2012. Fairness through awareness. In *Proc. ACM ITCS*. 214–226.

[25] Pelle Ehn. 1988. *Work-oriented design of computer artifacts*. Ph. D. Dissertation. Arbetslivscentrum.

[26] Simon Enni and Ira Assent. 2021. Learning by Design: Structuring and Documenting the Human Choices in Machine Learning Development. arXiv:2105.00687 [cs.LG]

[27] S. D. Essinger and G. L. Rosen. 2011. An introduction to machine learning for students in secondary education. In *2011 Digital Signal Processing and Signal Processing Education Meeting, DSP/SPE 2011 - Proceedings*. 243–248. https://doi.org/10.1109/DSP-SPE.2011.5739219

[28] J. Estevez, X. Elizetxea, and G. Garate. 2019. Development of a Workshop About Artificial Intelligence for Schools. *Advances in Intelligent Systems and Computing* 771 (2019), 530–535. https://doi.org/10.1007/978-3-319-94120-2{_}51

[29] J. Estevez, G. Garate, and M. Grana. 2019. Gentle Introduction to Artificial Intelligence for High-School Students Using Scratch. *IEEE Access* 7 (2019), 179027–179036. https://doi.org/10.1109/ACCESS.2019.2956136

[30] Ignacio Evangelista, Germán Blesio, and Emanuel Benatti. 2019. Why are we not teaching machine learning at high school? A proposal. In *2018 World Engineering Education Forum - Global Engineering Deans Council, WEEF-GEDC 2018*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/WEEF-GEDC.2018.8629750

[31] Juan David Rodríguez Garcia, Jesús Moreno Leon, Marcos Román Gonzalez, and Gregorio Robles. 2019. Developing Computational Thinking at School with Machine Learning: An exploration. In *2019 International Symposium on Computers in Education, SIIE 2019*. https://doi.org/10.1109/SIIE48397.2019.8970124

[32] Á García, O Anjos, C Iglesias, H Pereira, J Martínez, and J Taboada. 2015. Prediction of mechanical strength of cork under compression using machine learning techniques. *Materials and Design* 82 (2015), 304–311. https://doi.org/10.1016/j.matdes.2015.03.038

[33] Xaioyan Gong, Yilin Wu, Zifan Ye, and Xiwei Liu. 2018. Artificial Intelligence Course Design: ISTREAM-based Visual Cognitive Smart Vehicles. In *IEEE Intelligent Vehicles Symposium, Proceedings*, Vol. 2018-June. Institute of Electrical and Electronics Engineers Inc., 1731–1735. https://doi.org/10.1109/IVS.2018.8500457

[34] Xiaoyan Gong, Liang Zhao, Renhao Tang, YaNa Guo, Xiwei Liu, Jian He, Fei-Yue Wang, Ying Tang, Wanruo Shi, Xiaojie Niu, and Xinzhu Wang. 2019. AI education system for primary and secondary schools. In *American Society for Engineering Education (ASEE) Annual Conference and Exposition, Conference Proceedings*. American Society for Engineering Education. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85078800693&partnerID=40&md5=6ab98a8b313d50d7a8e2a637bf8a5c20

[35] Mihajlo Grbovic, Vladan Radosavljevic, Nemanja Djuric, Narayan Bhamidipati, Jaikit Savla, Varun Bhagwan, and Doug Sharp. 2015. E-commerce in your inbox: Product recommendations at scale. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 1809–1818.

[36] Shuchi Grover and Roy Pea. 2013. Computational Thinking in K–12. *Educational Researcher* 42, 1 (2013), 38–43. https://doi.org/10.3102/0013189X12463051

[37] T. Hitron, H. Erel, I. Wald, and O. Zuckerman. 2018. Introducing children to machine learning concepts through hands-on experience. In *IDC 2018 - Proceedings of the 2018 ACM Conference on Interaction Design and Children*. Association for Computing Machinery, Inc, 563–568. https://doi.org/10.1145/3202185.3210776

[38] Tom Hitron, Yoav Orlev, Iddo Wald, Ariel Shamir, Hadas Erel, and Oren Zuckerman. 2019. Can children understand machine learning concepts? The effect of uncovering black boxes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, New York, New York, USA. https://doi.org/10.1145/3290605.3300645

[39] Ole Sejer Iversen, Rachel Charlotte Smith, and Christian Dindler. 2018. From computational thinking to computational empowerment. In *PDC '18 Proceedings of the 15th Participatory Design Conference: Full Papers - Volume 1*. ACM Press, Hasselt and Genk, Belgium, 1–11. https://doi.org/10.1145/3210586.3210592

[40] Severin Kacianka and Alexander Pretschner. 2021. Designing Accountable Systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 424–437. https://doi.org/10.1145/3442188.3445905

[41] Yasmin Kafai, Chris Proctor, and Debora Lui. 2019. From theory bias to theory dialogue: Embracing cognitive, situated, and critical framings of computational thinking in K-12 Cs education. In *ICER 2019 - Proceedings of the 2019 ACM Conference on International Computing Education Research*. 101–109. https://doi.org/10.1145/3291279.3339400

[42] K. Kahn, R. Megasari, E. Piantari, and E. Junaeti. 2018. AI programming by children using snap! Block programming in a developing country. In *CEUR Workshop Proceedings*, Drachsler H Drachsler H Dimitrova V Drachsler H Fominykh M. Praharaj S. (Ed.), Vol. 2193. CEUR-WS. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85053707384&partnerID=40&md5=f572adf493a473f9dbe766b2fa5945d9

[43] Ken Kahn and Niall Winters. 2017. Child-Friendly Programming Interfaces to AI Cloud Services. In *European Conference on Technology Enhanced Learning*.

566–570. https://doi.org/10.1007/978-3-319-66610-5

[44] Michael Katell, Meg Young, Dharma Dailey, Bernease Herman, Vivian Guetler, Aaron Tam, Corinne Bintz, Daniella Raz, and P. M. Krafft. 2020. Toward Situated Interventions for Algorithmic Equity: Lessons from the Field. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 45–55. https://doi.org/10.1145/3351095.3372874

[45] Joshua A Kroll. 2018. The fallacy of inscrutability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2133 (2018), 20180084.

[46] Annabel Lindner, Stefan Seegerer, and Ralf Romeike. 2019. Unplugged Activities in the Context of AI. In *12th International Conference on Informatics in Schools: Situation, Evolution, and Perspectives, ISSEP 2019, Larnaca, Cyprus, November 18–20, 2019, Proceedings*. Springer International Publishing, 123–135.

[47] Duri Long and Brian Magerko. 2020. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376727

[48] Gustavo López, Luis Quesada, and Luis A Guerrero. 2017. Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 241–250.

[49] R. Mariescu-Istodor and I. Jormanainen. 2019. Machine learning for high school students. In *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3364510.3364520

[50] Lívia S. Marques, Christiane Gresse von Wangenheim, and Jean C. R. Hauck. 2020. Teaching Machine Learning in School: A Systematic Mapping of the State of the Art. *Informatics in Education* 19, 2 (2020), 283–321. https://doi.org/10.15388/infedu.2020.14

[51] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2019. A survey on bias and fairness in machine learning. *arXiv preprint arXiv:1908.09635* (2019).

[52] Milagros Miceli, Tianling Yang, Laurens Naudts, Martin Schuessler, Diana Serbanescu, and Alex Hanna. 2021. Documenting Computer Vision Datasets: An Invitation to Reflexive Data Practices. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 161–172. https://doi.org/10.1145/3442188.3445880

[53] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (2019), 1–38.

[54] Smitha Milli, Luca Belli, and Moritz Hardt. 2021. From Optimizing Engagement to Measuring Value. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 714–722. https://doi.org/10.1145/3442188.3445933

[55] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proc FAT**. 220–229.

[56] T. Narahara and Y. Kobayashi. 2018. Personalizing homemade bots with plug & play AI for STEAM education. In *SIGGRAPH Asia 2018 Technical Briefs, SA 2018*. Association for Computing Machinery, Inc. https://doi.org/10.1145/3283254.3283270

[57] Harold G Nelson and Erik Stolterman. 2014. *The design way: Intentional change in an unpredictable world.* MIT press.

[58] E. Ossovski and M. Brinkmeier. 2019. Machine Learning Unplugged - Development and Evaluation of a Workshop About Machine Learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11913 LNCS (2019), 136–146. https://doi.org/10.1007/978-3-030-33759-9{_}11

[59] Samir Passi and Phoebe Sengers. 2020. Making data science systems work. *Big Data & Society* 7, 2 (2020). https://doi.org/10.1177/2053951720939605

[60] Inioluwa Deborah Raji, Morgan Klaus Scheuerman, and Razvan Amironesei. 2021. You Can't Sit With Us: Exclusionary Pedagogy in AI Ethics Education. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 515–525. https://doi.org/10.1145/3442188.3445914

[61] Inioluwa Deborah Raji, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 33–44.

[62] A. Rizzo, F. Montefoschi, M. Caporali, and G. Burresi. 2018. *UAPPI: A platform for extending app inventor towards the worlds of loT and machine learning.* IGI Global. 88–109 pages. https://doi.org/10.4018/978-1-5225-5969-6.ch003

[63] Andrew Slavin Ross, Michael C. Hughes Hughes, and Finale Doshi-Velez. 2017. Right for the Right Reasons: Training Differentiable Models by Constraining their Explanations. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*. 2662–2670. https://doi.org/10.24963/ijcai.2017/371

[64] A. Scheidt and T. Pulver. 2019. Any-cubes: A children's toy for learning AI. In *ACM International Conference Proceeding Series*, Doring T Alt F. Bulling A. (Ed.). Association for Computing Machinery, 893–895. https://doi.org/10.1145/3340764.3345375

[65] Alexander Scheidt and Tim Pulver. 2019. Any-cubes: A children's toy for learning AI. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, New York, New York, USA, 893–895. https://doi.org/10.1145/3340764.3345375

[66] D Schon. 1983. *The reflective practitioner. How professionals think in action.* Temple Smith, London.

[67] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) *(FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 59–68. https://doi.org/10.1145/3287560.3287598

[68] Hong Shen, Wesley H. Deng, Aditi Chattopadhyay, Zhiwei Steven Wu, Xu Wang, and Haiyi Zhu. 2021. Value Cards: An Educational Toolkit for Teaching Social Impacts of Machine Learning through Deliberation. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 850–861. https://doi.org/10.1145/3442188.3445971

[69] N. Sintov, D. Kar, T. Nguyen, F. Fang, K. Hoffman, A. Lyet, and M. Tambe. 2016. From the lab to the classroom and beyond: Extending a game-based research platform for teaching ai to diverse audiences. In *30th AAAI Conference on Artificial Intelligence, AAAI 2016*. AAAI press, 4107–4112. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85007226830&partnerID=40&md5=29344871cc4b5cf956d762cd4a4a1807

[70] Zoe Skinner, Stacey Brown, and Greg Walsh. 2020. Children of Color's Perceptions of Fairness in AI: An Exploration of Equitable and Inclusive Co-Design. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3382901

[71] Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 3645–3650. https://doi.org/10.18653/v1/P19-1355

[72] Latanya Sweeney. 2013. Discrimination in online ad delivery. *Queue* 11, 3 (2013), 10.

[73] Petros Terzis. 2020. Onward for the Freedom of Others: Marching beyond the AI Ethics. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 220–229. https://doi.org/10.1145/3351095.3373152

[74] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. 2019. Envisioning AI for K-12 : What should every child know about AI ?. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 9795–9799. https://doi.org/10.1609/aaai.v33i01.33019795

[75] Henriikka Vartiainen, Matti Tedre, and Teemu Valtonen. 2020. Learning machine learning with very young children: Who is teaching whom? *International Journal of Child-Computer Interaction* 25 (2020). https://doi.org/10.1016/j.ijcci.2020.100182

[76] Peter-Paul Verbeek. 2006. Materializing Morality: Design Ethics and Technological Mediation. *Science, Technology, & Human Values* 31, 3 (2006), 361–380.

[77] X. Wan, X. Zhou, Z. Ye, C.K. Mortensen, and Z. Bai. 2020. SmileyCluster: Supporting accessible machine learning in K-12 scientific discovery. In *Proceedings of the Interaction Design and Children Conference, IDC 2020*. 23–35. https://doi.org/10.1145/3392063.3394440

[78] Adrian Weller. 2019. Transparency: Motivations and challenges. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Springer, 23–40.

[79] Randi Williams, Hae Won Park, and Cynthia Breazeal. 2019. A is for Artificial Intelligence: The Impact of Artificial Intelligence Activities on Young Children's Perceptions of Robots. In *Conference on Human Factors in Computing Systems - Proceedings*. ACM Press, New York, New York, USA, 1–11. https://doi.org/10.1145/3290605.3300677

[80] Jeannette M. Wing. 2006. Computational thinking. *Commun. ACM* 49, 3 (2006). https://doi.org/10.1145/1118178.1118215

[81] Terry Winograd (Ed.). 1996. *Bringing Design to Software.* Association for Computing Machinery, New York, NY, USA.

[82] Tal Z Zarsky. 2012. Automated prediction: Perception, law, and policy. *Commun. ACM* 55, 9 (2012), 33–35.

[83] Abigail Zimmermann-Niefield, Shawn Polson, Celeste Moreno, and R. Benjamin Shapiro. 2020. Youth making machine learning models for gesture-controlled interactive media. In *Proceedings of the Interaction Design and Children Conference, IDC 2020*. 63–74. https://doi.org/10.1145/3392063.3394438

[84] Abigail Zimmermann-Niefield, Makenna Turner, Bridget Murphy, Shaun K Kane, and R Benjamin Shapiro. 2019. Youth Learning Machine Learning through Building Models of Athletic Moves. In *IDC '19 Proceedings of the 18th ACM*

International Conference on Interaction Design and Children. ACM Press, New York, New York, USA, 121–132. https://doi.org/10.1145/3311927.3323139