Supporting Critical Data Literacy in K-9 Education: Three Principles for Enriching Pupils' Relationship to Data

Children grow up in a data economy but grasping how their data are part of this eco-system and how it becomes valuable to others can be difficult. This work explores how to design tools and activities which support children's critical data literacy for K-9 education. We bring together two strands of work; First, insights from a co-design process where teachers and researchers designed tools and activities for teaching critical data literacy which they deployed in a lower secondary education classroom. Second, insights from didactic theories from maths education about working with multiple and rich representations of complex and intangible phenomena. Based on this we contribute three principles for enriching pupils' relationship to data in order to inform future research into how students can be scaffolded in forming richer relationships to the data-driven technologies in their everyday lives in order to retain agency in a data-driven world.

CCS Concepts: • **Social and professional topics** → **Computing literacy**; **K-12 education**; • **Security and privacy** → *Human and societal aspects of security and privacy*; • **Human-centered computing** → *Empirical studies in interaction design.*

Additional Key Words and Phrases: critical data literacy, K-9 education

ACM Reference Format:

. 2018. Supporting Critical Data Literacy in K-9 Education: Three Principles for Enriching Pupils' Relationship to Data. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

This paper contributes to current CCI research on promoting critical data literacy (e.g., [12, 36]) by offering three principles for designing educational tools and activities that support *an enrichment* [34] of K-9 pupils' relationship to the data they produce in their everyday life, and how it is used.

Today, children grow up in a *data economy* [5, 23, 29], where consumers often pay with their personal data for products and services instead of with traditional currency. These products and services collect data about how and by whom they are used, often in ways that are opaque and impossible to opt out of. Combined with new business models, which are centered around collection and analysis of data, personal data have become value-assets which are traded and sold [21, 26]. Users transfer ownership of their personal data to corporations in exchange for personalization of online services, information retrieval, entertainment, etc., often to put themselves in a position where even more data can be collected about them [28]. These issues have been targeted through political regulations, such as the European Union's General Data Protection Regulation (GDPR), which give consumers more control over their personal data, and through social activism such as the *open data movement* that distributes raw data to empower citizens [3].

However, as argued by Gebre [13], giving users more control and making data resources available does not necessarily mean that users will take action. This should be accompanied by data literacy, i.e., educating users in how data are collected and analysed and engaging them in the complications around its usage to enable them to question status quo and demand changes. Furthermore, such literacy should include critical perspectives, raising awareness about the manifestations of power embedded in data-heavy, computational technologies and promote more reflective ways of using them. Recent literature has argued for emphasising the role of data in young people's lives [7, 13, 32] and including more critical perspectives in data literacy [5, 13, 28]. However, most literature is conceptual, and examples of educational tools and practical interventions are limited [13]. Existing examples mainly pay attention to how pupils can utilise data to engage critically with urgent topics and communities (e.g. [6, 13, 14, 35]), but pay less attention to how pupils on an everyday basis exchange their own personal data.

Recent CCI research have started to pay attention to these personal perspectives and documented how young people (aged 13-18) are generally critical of how personal data are sold, aggregated and used, but also feel unable to counteract it [12], and how children need to understand the technology behind data privacy issues [36]. While this research focus on data tracking and collection technologies (e.g., cookies, pixel tags, digital fingerprinting, etc.), it does, however, not expose algorithmic data-analysis technologies used for aggregating large amounts of collected data. Other CCI research have engaged children and young people in the data-analysis processes (e.g. [30, 38]), but do not engage with the personal and critical perspectives.

In this paper, we examine the challenge of designing educational tools and activities which engage pupils in both understanding how algorithmic processes can turn otherwise harmless-seeming data into valuable information through aggregation of large data-sets, and in reflecting on the critical perspectives around these practices. To conduct this examination, we analyse the design and deployment of four educational tools and activities. They were designed in a co-design process where ten researchers collaborated with nine secondary school teachers to to design and develop tools and activities for teaching data literacy and deployed them in an intervention in an 8^{th} grade classroom. Furthermore, we include theories and perspectives from Math education which which have a tradition of working with the concepts of abstract and concrete [1, 8, 31, 34]. These concepts act as a lens to analyze and a standpoint from which to talk about the tools and activities, emphasising some aspects of the educational designs while leaving out other perspectives. We have chosen this body of theory as we see it as having a potential to inform the design of education on pervasive, computational technologies and infrastructures surrounding big data, including the implications linked to these. These technologies and implications easily become abstract to pupils, making it crucial to investigate how we can support pupils in forming a more concrete relationship to them.

Based on this examination, we contribute three principles to researchers, designers, and educators on how to enrich pupils' relationship to the data they produce in their everyday life, and how it is used: *1) Make data personal and close to support emotional engagement 2) Provide rich, contextualised representations of data to support reflections 3) Connect different modes of engagement with computational modelling to support sense-making*. This paper document how these principles are grounded in the design and deployment of educational tools and activities from a co-design process and didactic theories from maths education on enriching pupils' relationship to concepts/objects.

2 RELATED WORK

In this section we present work related to developing children's critical data literacy as well as learning tools and activities for teaching about data and practices related to them.

2.1 Critical Data Literacy

Multiple studies have explored how young people understand and use data, pointing out that most pupils understand data as something *scientific* instead of personal, and that those who do recognise its personal aspects have difficulty acting on their concerns [7, 13, 35]. Based on a survey with secondary school children, Gebre [13] finds that pupils have a rather narrow conception of data, which seems to be mainly based on experiences from laboratory experiments and from worksheets with structured data and conclude that this implies *"that there is a need to raise students' awareness about their own role as sources of data and how that affects them in everyday life."* [13]. Similarly, Bowler et al. [7] find that, when questioned about what data are, teens reference numeracy, statistics scientific inquiry and visualisation tools to interpret quantitative data, while few reference their personal digital traces on online platforms. Williams et al. [35] ground their approach in the open data movement and develop high school students' data literacy, by having students explore the lottery and their local community through interviews with players and retailers, and by analysing lottery data through an interactive map. Hautea et al. [14] have 700 young developers (median of self-reported age: 12) work with social and behavioural meta-data in the online Scratch community. They finds that the young developers' reflections have similarities to existing critical discourses pertaining data, but that the young developers do not know how to translate these into insights about data practices in the services they use in their everyday life. Dowthwaite et al. [12] explore young people's comprehension of personal data in online services but explicitly "black-box" the algorithmic processes behind, calling for age appropriate education on how these processes work and the issues linked to them. Last, Yap and Lee [36] propose a physical-digital-hybrid interactive book for teaching children about privacy concerns and highlight the need for children to understand the technology behind data privacy issues.

While the above research mainly focus on pupils' perception of data, and how they can be engaged in understanding and manually analysing data, it do not dive into computational processing of data. The next section will account for how research have tried to make these processes more concrete.

2.2 Making Computation More Concrete

Making computational concepts more concrete is not new. In an educational context, it has been explored how *digital manipulatives* [24] can support children in constructing knowledge though *concrete operations* (*tools to think with*), and how children through expressive and creative activities with construction kits can explore computational concepts [20, 25]. From an Interaction Design perspective, research on tangible interaction has explored how Tangible User Interfaces (TUIs) can couple (building on Dourish's notion of Embodied Interaction [11]) computational concepts to the experienced world [18].

More recent studies use embodied data collection in tools which exposes machine learning (ML) processes to children and young people [2, 15, 16, 33, 37, 38]. In most examples, students train ML models for gesture recognition, e.g., Agassi et al. [2] have students train an ML model to recognise different gestures and implement this model in a Scratch application as a gesture interface. Zimmermann-Niefield et al. [38] take a more contextualised approach. They have students use wearable sensors to train a ML classifier to recognise if a throw, kick or other physical activity is well or badly performed. The study has a good starting point for discussing students as data entities and the implications of such tools in a sports context but leaves this notion for future work. Ossovski and Brinkmeier [19] and Lindner et al. [17] design *unplugged activities* (see [4]) to teach pupils about data and ML. In these activities *"[L]earners handle data on their own as a computer could do with machine learning, but do not use a computer for this purpose"* [19]. Ossovski and Brinkmeier [19] have pupils organising screws by plotting data on a bulletin board and use this physical representation as a ML model, using a wooden strip as a linear classifier. Lindner et al. [17] present, among other, an activity where pupils act as different layers in a neural network (a type of ML) drawing and passing sketches between them to classify images.

Looking at these examples, we see a tendency to either focus on making the computational concepts concrete while not emphasising the implications, or doing the opposite; Contextualise the technology to make its implications concrete, but not engage pupils in the computational concepts. We argue, that both computational *and* social and personal aspects of data must be made concrete for pupils in order for them to develop a rich relationship to their data and to scaffold a critical data literacy in them. In the next section, we will elaborate on what we mean by *more concrete* and *rich relationship*.

3 ENRICHING PUPILS RELATIONSHIP TO DATA

In this section we investigate how didactic theories from maths education speaks into the design rationale behind, and the challenge addressed with, the educational tools and activities presented in the next section. based on this investigation, we present three *modes of activities*, which are used to analyse the findings from the intervention.

During the co-design process (which is elaborated in section 4), different themes for the intervention were discussed between researchers and teachers. Teachers argued that it can be difficult to engage eighth grade pupils in topics that are not about their own lives or close communities. They were concerned that topics such as data privacy, predictive policing, data analysis in political campaigns, etc., would be too *abstract* to engage the pupils, and that they would struggle to see its importance and relevance in their everyday lives. To engage pupils in critical reflection, the teachers argued, that pupils must *feel* the implications of personal data, instead of only rationalising about it; The topics would have to be embedded in, and linked to, pupils' *concrete* experiences. One teacher framed this as moving *"from private data to personal, embarrassing information"*. This shift from abstractly discussing data privacy to having pupils concretely experience and feel the implications of sharing personal data which can lead to embarrassing information became a frame which, in the co-design process, led to the design of several educational tools and activities.

Abstract and concrete became central concepts in the intervention, but they are used in many different contexts, and their meanings quickly become diluted. Therefore, we here take a step back to reflect and elaborate on what we actually mean, when we say we want to make big data and personal data more concrete to pupils.

Turkle and Papert [31] criticise computer science for overemphasising abstraction and formalism over concrete arguments which stay in touch with the inner workings of the argument and attach them to the specific context. Drawing on feminist theory, they call for a *revaluation of the concrete* in depicting between abstract and concrete arguments. They argue that some moral problems require *"a mode of thinking that is contextual and narrative rather than formal and abstract"* [31]. This is indeed the case for data privacy, which is difficult to discuss from an abstract perspective; Everybody agrees that data privacy is important and that AI systems should not discriminate, but does that help us in discussing the difficult questions about how the technologies should be implemented in our personal lives and society? We argue, it is difficult to discuss personal data in an abstract manner as its applications are intertwined into our social lives and various infrastructures, why its moral implications are linked to the responsibilities in our everyday life, to conflicting interests, and to the maintenance of our digitised society.

Wilensky [34] argues that what makes objects/concepts concrete are our relations to them, and this concreteness depends on the number of representations, modes of interaction, and different context we can relate to the object/concept. Thus, *"thinking concretely is seen not to be a narrowing of the domain of intellectual discourse, but rather as opening it up to the whole world of relationship"*. There have been multiple attempts to make data and AI concrete and graspable (e.g., [16, 17, 19, 38]), but with this definition in mind, making something concrete does not mean making it physical and manipulable. This can be one step in making something more concrete to a pupil, but providing something with a physical body, does not, necessarily, help the pupil form a richer relationship to the concept, e.g., being able to relate to similar concepts and put it into different contexts. Clements elaborates further on this notion of concrete and makes a distinction between *sensory-concrete* and *integrated-concrete*: respectively, the need to use sensory material to make sense of an idea and combining many separate ideas in an interconnected structure of knowledge (how meaningful they are) [8]. Making data (integrated-)concrete to pupils, hence becomes a question of supporting pupils in creating relations between different concrete representations of data in different contexts, and in using these different representations together to create a richer relationship to data.

Picard et al. [22] builds on Wilensky but emphasise pupils' emotional relationship to the process and content of learning as playing an important role in facilitating different modes of thinking. Therefore, learning activities should include personal perspectives, enable pupils to use personal identification, and account for the social side of learning [22]. Personal data are part of pupils' everyday life and a deeply personal matter, and thus pupils' emotional responses are an important factor to account for and instrument of use, when teaching about it.

To summarise, what we take from these theoretical perspectives can be synthesised into three modes of activity, which might indicate pupils forming a richer relationship to their content of learning (in this case personal data), making it more concrete for the students to engage with and reflect on:

- Pupils making arguments which engage with the concrete context.
- Pupils combining different representations to make sense.
- Pupils emotionally engaging with and reflecting on the process and content of learning.

Below, we present the details of the co-design process and intervention; the methodology behind, the data gathering, and the data analysis drawing on these modes of activity. Following this, we account for the design of the learning tools and activities used in the intervention as they were developed through the co-design process, and finally we use the three modes of activity to analyse the findings from the intervention.

4 METHODOLOGY

The empirical data used in this paper are a result of a co-design process in which a group of researchers and elementary and lower-secondary school teachers, designed an carried out an intervention in an 8^{th} grade classroom with a broad aim of teaching pupils about emerging technologies. Drawing on insights from didactic theories from maths education about working with multiple and rich representations of complex and intangible phenomena, we analyse the collected data from the classroom deployment. The experiences from the co-design process and the analysis result in three principles for Enriching Pupils' Relationship to data presented in section 7. The data collection and analysis are elaborated beneath. Both the co-design process and the intervention were carried out in accordance with local guidelines regarding the Covid-19 pandemic.

4.1 Data Collection

Following the co-design process, a core team of three teachers from the co-design process and five researchers carried out and intervention in an 8^{th} grade classroom (aged 14-15) of 27 pupils in which one of the participating teachers taught regularly. All pupils and their parents were consented leading up to the intervention. The intervention was carried out in late November 2020 and lasted five days. In this paper, we focus specifically on educational tools and activities from day two and three. See Schaper et al. [27] for an account of the whole intervention. Through the intervention, data were collected in several ways. The pupils were divided into eight groups of three to four pupils by the teachers in the class. Following each day, all eight groups recorded video diaries of their experiences answering questions about what they learned and what they found difficult during the day. In addition, four groups were selected by the teachers to film and comment on their work during the intervention. Finally, a post interview was made with all eight groups, each conducted by one of the nine teachers participating in the co-design workshops based on a short questionnaire developed by the core team of teachers and researchers. All videos were filmed using GoPro action cameras with an external microphone. Between three and five researchers were present during the intervention and collected data through field notes, observations, photos, and video. Last, pupils filled-in different templates throughout the intervention to document their work, which were collected by the researchers at the end of the intervention.

4.2 Data Analysis

The co-design process was characterised by a constant negotiation between all participants, and thus the content of the intervention is a compromise and contains several possible narratives to report on. Following the invention four of the authors began unravelling these narratives by discussing the data material from the co-design process; the material produced, the video documentation and the descriptions of the intervention including its objectives and timetable, identifying themes running through the material. Themes about *how pupils experienced the implications, the different framings and representations of data, the embodiment and contextualisation of computational models*, and *the connections between the physical and digital representations* emerged from this discussion.

Three authors used these themes as deductive codes to analyse the video data collected during the intervention (described above). In order to reduce individual biases, one author coded a diverse selection of video material which were discussed among the three authors to reach consensus. The remaining video material were divided equally among all three authors. Last, the three authors applied the three *modes of activity* presented in section 3 to, collaboratively, sort and further analyse the coded data. The results of this analysis are presented in section 6. Prior to that, we present the set of tools and activities that were deployed in the intervention.

5 DESIGN

This section account for the design of three tools and one activity used in the intervention and how they emerged from a co-design process. As described above, the educational tools, activities and tuition plan in the intervention were designed through a co-design process with nine teachers and ten researchers. The aim in the co-design process was to make a shift in how personal data are taught and discussed in the classroom moving *"from private data to personal, embarrassing information"*.

The three tools are a paper template (the Data Compass), a Google Form questionnaire (the Pop Quiz) and a web application (the KNN Exploration Tool). The activity is an instructional embodied activity with the whole classroom. The activity and the tools were combined in a teaching plan lasting over two days in the intervention as illustrated in Table 1. Here we account for how the three tools and the activity were designed through the co-design process and how they were implemented in the intervention. We focus here on the design of three tools and one activity in the intervention, which all were designed to engage pupils in their personal data, and how they can be (and is) used in computational models.

5.1 Pop Quiz - Sharing Personal Data

To illustrate to pupils how data are collected in the background when using digital services, the idea spurred, in the co-design process, to 'disguise' the data collection as a pop-quiz. One teacher brought a collection of youth magazines with different quizzes, which researchers and teachers used to collaboratively design questions for a pop-quiz, which

Day	Time	Activity
Before day 1	N/A	Collect data from pupils: Pop Quiz
Day 1	08:00-08:20	General introduction
	08:20-08:40	Introduction to data-set from pop quiz
	08:40-09:40	Reflection: The Data Compass
Day 2	08:00-08:15	General introduction
	08:15-09:40	Understand modelling: Embodied KNN
	09:40-10:10	Break
	10:10-10:40	Discussions of Embodied KNN
	10:40-11:10	Explore modelling: KNN exploration tool
	11:10-11:35	Reflection: The Data Compass

Table 1. Detailed intervention schedule of the two days in the classroom.

was meaningful to the pupils and which answers could be used as data in the intervention. It resulted in the pop-quiz: *"Who from Paradise Hotel* [1] *are you?"*

The pop-quiz is an online form, created in Google Forms[2], that mimics the quizzes pupils see on social platforms and in magazines, which promise insights about themselves or their peers if they answer a few questions. It consists of 18 questions ranging from *their age* and *how far they have to school*, to questions about *how important it is for them to earn a lot of money* and *how their friends would describe them in one word*. In the intervention, the teacher asked his class to answer the pop-quiz as homework, and the data collected from the pop quiz were used as a data-set in the following activities.

5.2 The Data Compass - Reflecting on Data

To scaffold contextualised discussions and reflections about personal data, we designed *the data compass*. One of the teachers proposed an exercise where pupils could rate data from boring to really embarrassing to scaffold pupils in talking about data as a personal thing you choose to share. We further developed this with the perspective on data as value-assets pupils pay with on an everyday basis in exchange for access to services, personalisation, etc. The two perspectives were combined in a coordinate system, where pupils could rate data according to their willingness to share it and how they perceived the data's value to others. In this manner, pupils work with data from a personal, experienced perspective and a broader perspective simultaneously and compare and discuss the two perspectives.

The Data Compass, see Figure 1, is a coordinate system with two axes printed on an A3 paper and data types (age, favourite subject, etc.) on small snippets of paper, that can be glued onto the A3 paper with the compass. The horizontal axis describes how willing pupils are to share data and the vertical axis describes how valuable they think it is to the receiver. E.g., pupils ask themselves and discuss: *"how willing am I to share how many hours I use on my smartphone a day?"* and *"how can the ones I share it with make use of this information?"*. In the intervention, pupils plotted data types from the *pop quiz* in the compass at the beginning of day 1 so their plots could be used in classroom discussions about sharing personal data, facilitated by one of the teachers. At the end of day 2, the pupils repeated the exercise to see how they had changed their perspectives on data.

5.3 Embodied KNN - Understanding Modelling

During the co-design process, we discussed in depth how seemingly harmless information can be a proxy to more valuable information (e.g., an address can be used to guess a person's income), and how an individual data point may

[1] Paradise Hotel is a popular, local reality TV show
[2] Note that pupils had access to the Google Educational Suite which is compliant with local legislation on data collection, and that all data collected were thoroughly anonymised before leaving this platform.

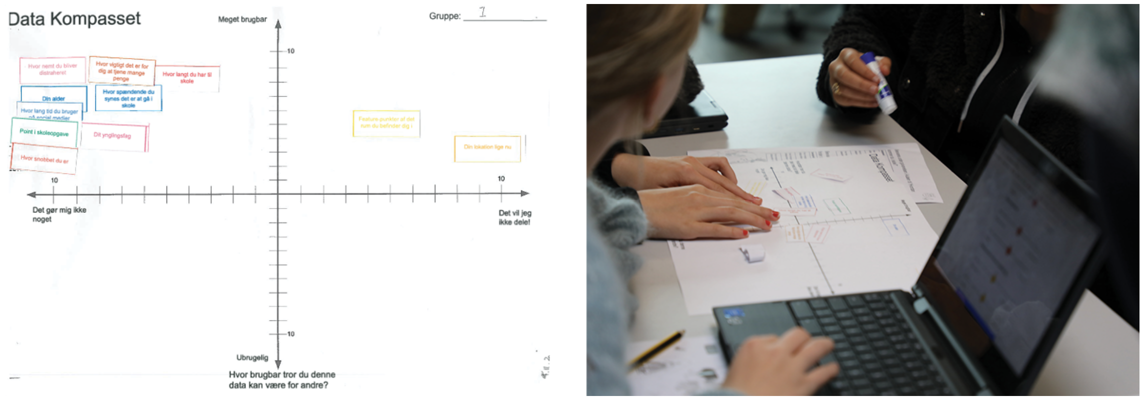

Fig. 1. Left: The *Data Compass* filled out with the *pop quiz* data by a group of pupils. Right: a group filling out the *Data Compass*

carry little value in itself but when aggregated with other data, it can be used to deduct insightful information. Providing pupils with a fundamental understanding of these concepts, and allowing them to explore this phenomenon, can help them understand how their personal data are used, and why it can be more valuable than they first think. However, for pupils to understand these concepts, they must have a basic understanding of computational modelling, which some teachers feared would be too difficult. One teacher argued that explaining a single scatter-plot graph to the class could be a whole lesson in itself, and having pupils explore multiple data entities might not be feasible. Different computational models were discussed among researchers and teachers, and one of the teachers came up with the idea for an embodied data modelling activity, where pupils themselves become data points in a K-nearest Neighbour (KNN) ML model[3]. Pupils arrange themselves in a spatial mapping on a floor according to their own data and identify their closest neighbours (i.e., classmates) to experience how the model aggregates their data with others to gain insights about them.

This became the the *Embodied KNN* exercise, an *unplugged activity* [4] where pupils embody a two-dimensional KNN model, acting as data points representing their own personal data. A coordinate system is drawn on the floor in a large room and pupils arrange themselves in the coordinate system according to the axis and, simulating the KNN algorithm with different values of K, asking their two, four and six closest classmates questions. In the intervention, the teachers chose to make a model which predicted sport a pupil prefers (e.g., dancing, handball, or e-sport) by looking at their height and hours of stationary activity each week. The whole class arranged themselves in the coordinate system according to their height and hours of stationary activity, asked their two, four and six nearest neighbours about which sport they prefer, and noted if the model were able predict their preferred sport correctly with the different K-values. A teacher photographed the pupils' arrangement from above, as seen in Figure 2 (to the left), while pupils hold coloured papers representing their preferred sport (e.g., pupils who prefer e-sport has a blue-coloured paper), which was then used for classroom discussions about KNN models, as seen in figure 2 (to the right).

5.4 KNN Exploration Tool - Exploring Data and Modelling

While the Embodied KNN activity can help pupils understand how data can be aggregated and analysed, it leaves little room for further exploration of data and computational models beyond the 2 dimensions allowed on a floor. It is very

[3]A KNN model is a relatively simple supervised ML model, which categorises new data entries by comparing them to their k nearest neighbours.

Fig. 2. Left: Pupils in the coordinate system of the *embodied KNN* activity. Right: the teacher using the image from the exercise as a basis for discussion

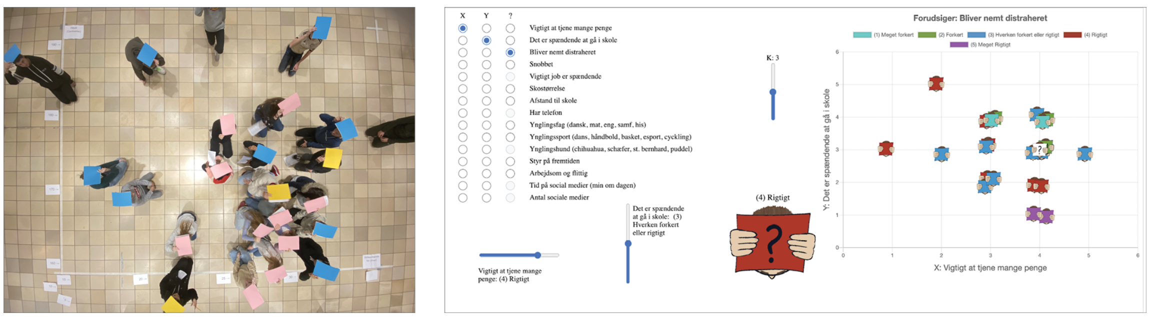

Fig. 3. The picture taken from above in the *embodied KNN* is referenced in the interface of the *KNN exploration tool*.

instructional and each change to the model takes a lot of coordination between pupils and teacher. To complement the Embodied KNN activity, we designed the KNN Exploration Tool; a digital tool, which scaffolds pupils' in further exploring the class' data (from the pop quiz) and build different KNN models. The teachers worried that the tool introduced too many new concepts for the pupils at once. Thus, the interface was designed to resemble the embodied KNN Activity, see Figure 3, to take advantage of pupils' prior experiences with this. When pupils use the tools, they can draw on their experiences from the embodied activity to make sense of the visualisation and computations in the tool.

The KNN Exploration Tool is a web application, which let groups of pupils explore how a KNN model can provide insights about themselves and the class by exploring the data-set and quickly creating and iterating on different KNN models. The interface visually references the *Embodied KNN* exercise with pupils holding up colored papers in a coordinate system as seen in Figure 3. In the intervention, pupils were presented with a fictive design case: *The School for Gifted Children.* Here, pupils built a KNN model to select which pupils to admit to a school, who are only interested in the very best pupils. The tool used the data from the *pop quiz* and pupils worked in groups using the tool on their laptops to determine what constitutes a good pupil and which data are useful for predicting this. At the end of the activity groups presented their new models and the rationales behind them to each other. After using the KNN Exploration Tool, each group documented their model in a paper template, noting the models' inputs, output, k-value, and the rationales behind their choices.

6 RESULTS FROM DEPLOYMENT

To analyse data from the intervention, we use the three modes of activity identified in section 3. Thus, we look for pupils making arguments which engage with the concrete context, pupils combining different representations to make sense, and pupils emotionally engaging with, and reflecting on the process and content of learning, as evidence for pupils' relationship with data becoming more concrete. The pupils were divided in the same eight groups throughout the activities which we will refer to as group 1-8. The findings are presented below.

6.1 Making arguments which engage with the concrete context

The Data Compass supported pupils in reflecting on how they specifically shared data in their everyday life, and to use these concrete experiences to rationalise about the implications of sharing this data. Through reflecting a the very specific types of data and using their own lives as context, they became creative in imagining, and rationalising about, how their data could be used by others.

Generally, pupils seemed to understand that data are collected about them, and they were knowledgeable about how their data could be valuable to different stakeholders. Many of their rationales were centered around personalising ads, e.g., group 1 discussed how 'distance to school' could be used to sell a bicycle or an electric scooter and group 2 discussed, if it could be used to sell bus tickets. Pupils also discussed how seemingly dull data could become valuable if used as a proxy for something else, e.g., group 8 discussed if *"how snobbish you are"* can say something about how good you are at working together with other people. The Data Compass encouraged pupils to articulate why they saw some data as more or less valuable, as they had to explain their choices, which were often challenged by other group members. E.g., two pupils in group 7 agreed that 'time on social media' was not useful and justified it with *"it is just good that you have spent a lot of time on their media"* before they rated it as a minus two on the 'usefulness' scale. Another group member was sceptical and proposed *"it could also mean [..] you have a lot of time that you do not spend on it. We can give you this app where you can spend a lot of time."*. In the post interviews both group 4 and 7 emphasised how they had learned that data are more valuable than they initially thought. One pupil from group 4 explained: *"Some very simple data can actually be very valuable [..] you can learn a whole lot about a lot of different people"*. They used the exercise with 'The School for Gifted Children' as an example: *"[..] The school really wants to know it [school assignment grades] because they want all the good ones [pupils]'*.

Pupils' discussions during the intervention also illustrated, why talking about themes on an abstract level may not address the real issues and decisions pupils make in their everyday life. Even though, they understood how data can be valuable and potentially be used against their interests, this often did not impact their willingness to share it, and many pupils seemed to be aware of this contradiction. A pupil from group 2 commented, as she looked at their Data Compass: *"We don't care about the things we share. We have already clicked accept to everything, so we can't do so much about it"*. In the post interviews 3 groups also addressed this. E.g., a pupil from group 1 stated: *"As a starting point I do not want to share so much data about myself, but we do it anyway even though we don't like it so much. Because then we can use the apps that exists"*. Even though, they understood the trade, they did not perceive themselves as having other options than accepting it.

6.2 Combining different representations to make sense

Throughout the intervention, pupils encountered multiple representations of their own data and different ways of engaging with the data. In the Data Compass exercises they reflected on how they 'trade' their personal data on a everyday basis, in the Embodied KNN exercise they experienced how data are aggregated and analysed with other data, and in the KNN Exploration Tool exercise they used the results from these analyses in a decision-making process (who should be admitted to the school for the special ones).

In the Embodied KNN exercise, pupils, simultaneously, represent both themselves as an individual standing among peers and a data point in a computational model through their spatial position in the coordinate system. The spatiality of the exercise seemingly supported pupils in understanding how the modelling worked. During the exercise, we observed that pupils were engaged when moving around the room. They were attentive to the position of each other, especially when asked to distribute themselves based on their height. This also came up in a classroom discussion following the activity, in which one girl proclaimed that *"some people sort of overrated their height"*, leading to discussions about how this affected the output of the model. In the post-interviews, when pupils were asked which activities, they learned the most from, five groups mentioned the Embodied KNN activity, highlighting that they found the activities in which they were active participants to be more engaging. However, some pupils also found the activity rather difficult and had their nearest classmates help them position themselves correctly in the coordinate system. One pupil got so frustrated during the activity, that she needed special attention from a teacher. Group 8 highlighted: *"It was a little hard to place yourself in that.. ehm... system"* and a pupil from group 8 described their takeaway from the activity as: *"You could see how some things was created from a pattern"*, which could indicate only a superficial understanding of what was happening in the exercise. Therefore, it was crucial, that the teachers spent time afterwards following up on the activity. They used the images from the exercise, see Figure 2, to explain how the models worked and frame a discussion around computational models and personal data. Especially the social aspects of being placed next to certain classmates due to the value of a certain data-point engaged pupils in the following discussions.

Connecting the two representations (pupils in a classroom and data points in a KNN model) seemingly engaged pupils by (according to the beneath mode of activity) spur emotional responses and made the new KNN models drawn on the whiteboard afterwards more concrete and richer to the pupils as they could connect them to the experience they just had. The teachers actively called back to the embodied activity to make the pupils perceive the data points as representing their peers in the classroom. A pupil from group 5 mentioned in the post interview that the discussions following the exercise were much more interesting since they where a part of the models being discussed.

Some pupils were also able to connect the embodied model with how computational models are applied in consumer services. For instance, a pupil from group 8 compared the groups of close neighbours in the Embodied Exercise to how YouTube's recommendation system might work, grouping users with similar interests to provide better recommendations.

In the post-interviews, several pupils mentioned the KNN Exploration Tool as the most challenging activity of the intervention which was also reflected in the videos from the groups documenting their resulting models. One pupil from group 8 correctly explained the process as; *"[Choosing] what you want to figure out, and what you need to figure it out"*, however, this pupil went on to explain that she did not understand what would happen from there. Each group documented their model in a template, where they wrote down their chosen data types and rationales for choosing them. A common error between them was that pupils thought of the options in the interface (two input variables and one output) as three variables used to describe a new data entry. However, all groups seemed to have a basic understanding, after the Embodied KNN exercise, about how patterns can emerge from data, which they applied when designing their models. This showed in the pupils' models, and their presentation of these, where their arguments were based on patterns in the data. Few pupils did, however, explicitly reference the Embodied KNN activity in their explanations of their models. This and pupils' challenges with understanding the tool might indicate that connections between the two activities could be strengthened even more.

6.3 Emotionally engaging with, and reflecting on the process and content of learning

In the The Data Compass exercises, emotions played an important role in engaging pupils in reflecting on implications of sharing personal data. While few group discussions started with arguments on how sharing some data could be unfavourable or against their own interest, many discussions were prompted by a group member sharing how they *felt* about sharing the data. If they experienced sharing a specific type of data as creepy, scary, or uncomfortable, they were hesitant to do so. This was especially clear with 'location' data, which many groups were unwilling to share. A pupil from group 1 expressed this as: *"It's scary if they know where you are all the time. It just feels wrong and is uncomfortable"*, after having been pressured from other group members to explain why, the group member elaborated: *"They can stalk you. They can earn money on stalking you."*. A pupil from group 2 argued *"people can kidnap me!"*. Group 5 simply moved the location label to the far left of the Data Compass indicating that they were not willing to share it. In general, pupils also found it uncomfortable if data became too personal, e.g., if it could be used to learn something about their personality. One pupil from group 2 would not share 'how snobbish you are' because *"They should not know how I am behind the screen"*.

The Embodied KNN activity also spurred emotional reactions from the pupils in discussions afterwards. The class discussed a case, where pupils' height and hours of sedentary activity per week were used to predict their favourite sports activity. As it turned out, taller pupils seemed to like e-sport more. In their discussions, one male pupil reasoned; *"It's boys that are of that height"*. At this point, a teacher revealed to the pupils, that this was intentional and that we had used height as a *proxy* for determining the pupils' gender. This generalisation observably provoked especially female pupils' leading to further discussions of using proxies in data collection. Several pupils mentioned this specific episode in the post interviews, and they used it to explain how data can be used as a proxy, e.g., a pupil from group 7 explained: *"I have also learned, how there can be some hidden data behind some data, e.g., if you just give your shoe size or your height, then you think, that it is not too bad, but then they can actually know which gender you are[..]"* .

In classroom discussions, teachers were quick to pick up on pupils' emotions and use them for discussing broader implications of the technologies and questioning if there were legitimate reason for these feelings.

7 THREE PRINCIPLES FOR ENRICHING PUPILS' RELATIONSHIP TO DATA

In this paper we have investigated how educational activities and tools can be designed to enrich k-9 pupils' relationship to data in order to support a critical data literacy. We argue, that designing such educational activities should not be a choice (neither a spectrum) between diving into computational concepts or critical implications, instead the two should be viewed as crucial instruments for making the other more concrete.

Building on the analyses and theory presented above, we present three design principles aimed at educators and designers of tools and activities for supporting children's data literacy. The principles do not correspond directly to the modes of activity identified in section 3 but are an operationalisation of our work with and findings from addressing them.

7.1 Make data personal and close to support emotional engagement

With the first principle, we recommend to emotionally engage k-9 pupils in data and the implications associated to the aggregation of personal data to support them in diverse ways of thinking [22]. To engage pupils emotionally, we recommend that activities and tools link data and their implications to pupils' experienced world. This is the world they emotionally engage with and create meaning from in their everyday lives, why learning activities and tools must emphasise how, and why, data play an important role in this world in order to make discussions and reflections about data personally relevant to pupils. When pupils start to make these connections between data and their own lives, they can start engaging emotionally in topics concerning these data.

This principle is visible throughout the intervention design: Pupils' personal data are collected in the Pop Quiz and used throughout the intervention to make choices in exercises and discussions centered around pupils themselves and their peers. Through the Data Compass, pupils discuss how willing they are to share their personal data. In the Embodied KNN exercise, pupils experience how their personal data is used in a computational model and compared with their classmates' data. Finally, with the KNN Exploration Tool, pupils explore how the class' data and their choices on how to use the data will affect whether they and their friends would be able to attend the School for Gifted Children.

In the deployment, we saw how group discussions were prompted by pupils sharing how they felt about sharing some data or experienced certain uses of their personal data as creepy or uncomfortable. We also saw how pupils were provoked when they discovered how their height were used to determine their gender in the Embodied KNN activity and used this as an example to describe the implications with this kind of practice. Finally, teachers actively used these personal experiences to engage pupils in classroom discussions about the broader implications of the technologies and practices.

Opposite to Picard et al. [22], who mainly focus on engaging children through positive emotions about, and relationships to, the content of learning, the designs in the intervention mainly focus on spurring negative emotions about, and critical reflections on, data-driven technologies. This may seem as a counter-productive stance towards technology, and we want to emphasise that this principle should not be interpreted as designs should strive to make data-driven technologies scary or present them as something that should be avoided. It is important that these emotional reactions are accompanied by reflections on why situations, curated by this new technology, are experienced as creepy, scary or uncomfortable and questioning if there is a legitimate reason for feeling this way. We acknowledge the quality of related studies which have successfully engaged teenagers in more positive relationships to data-driven technologies and practices by supporting them in utilising data to engage in their close communities [13, 14, 35]. We do, however, argue that in order to critically engage pupils in how their personal data are aggregated and analysed, it is crucial to let them experience how data can be used in ways that are also undesirable for them and their peers, and situations where they today are powerless in the hands of computational systems.

When working with this principle, it is key to be attentive to the balance between pupils feeling empowered and pupils feeling powerless. Letting pupils experience how they can use data for the benefit of themselves or others, is important for engaging them in actively taking advantage of, and agency in, the development of data-driven technologies. On the other hand, exposing to them how they in their daily lives, often do not have the option to take advantage of these technologies and how the technologies instead take advantage of them, is important for engaging them in taking a critical stance towards how their personal data are used, and how these systems are designed. Here, designers and teachers must be accountable for not burying pupils in negative emotions and anxiety, but guide and scaffold them in facing these power inequalities.

7.2 Provide rich, contextualised representations of data to support reflections

With the second principle, we recommend showing k-9 pupils rich representations of data, which tell the story of how and why data are collected and put it into relationship with the surrounding world. Aligning with Turkle and

Papert's critic of overemphasising abstraction in CS education [31], we argue that teaching data as an abstract concept and discussing the implications of the aggregation of personal data from abstract perspectives of *Justice* and *How You Should Morally Act* risk not addressing, or relating to, the real and concrete moral problems pupils experience in their everyday lives. Furthermore, formal representations of data risk to only provide pupils with an abstract relationship to data which is *"given solely by definition, and ... only by simple rules"* [31]. In today's data economy, data are much more than a set of values or observations. It is a value-asset, which has a history and a purpose and should be treated and taught as such. Pupils should be scaffolded in asking questions such as *"what is the purpose of this data?"*, *"how was it collected?"*, *"who owns it?"*, and *"why is it useful?"*.

This principle is visible throughout the intervention design: The personal data from the Pop Quiz provide the activities with rich contexts where pupils can reflect on conflicting interests and responsibilities in their everyday life and form concrete arguments which engage with their everyday thoughts and choices. Through the Data Compass, pupils explicitly reflect on who may benefit from data and how they can be valuable. In the Embodied KNN activity, pupils experience how they themselves can be perceived as data points. And In the KNN Exploration Tool, pupils take on the role of a stakeholder, for whom their personal data are valuable.

In the deployment, we saw how pupils, through the Data compass exercise, discussed how they shared data in their everyday life, and became creative in imagining how their data could be used by others. We also saw how these discussions led to realisations about the multitude of ways data can be perceived and used, and how valuable it can be. Finally, we saw how pupils saw the contradictions between their rationales for why and when you should, or should not, share certain data, and how they share personal data in their everyday lives.

In this manner, the intervention enriched pupils relationship to data by providing contexts for the use of data, expose the ambiguity of data, and engage them in the dilemmas which arise in practices around personal data. These concrete reflections on data supported pupils in forming personal opinions and being reflective regarding how they share their data. On the other hand, the intervention did not provide any concrete directions on how and where they could promote change in their everyday life. Thus, discussions lost their touch with pupils' everyday thoughts and choices when turning towards concrete actions pupils could do. To limit which personal data they share, pupils would either have to limit their use of social media which means opting-out of parts of their social lives, or demand changes of the practices in the computational systems they consume which seem unmanageable for an eighth-grade pupil. The intervention did, however, not address these conflicting interests and perspectives.

We encourage future use of this principle to explore how to provide pupils with concrete examples on where and how they, individually or collectively, can change their behaviour without paying too steep a price, and how they can take the first steps to call for change. These examples could be used to actively engage pupils in concrete reflections (and actions!) on how they are willing to change their behaviour, and if, and why, they are accepting the current status quo.

7.3 Connect different ways of engagement with computational modelling to support sense-making

With the third principle, we recommend engaging k-9 pupils in multiple types of activities with computational processing of data which each provide different entries to, and emphasise different perspectives on, computational models, and to scaffold pupils in connecting these for sense-making. Computational models are difficult to understand, and it can be difficult for pupils to connect them with their experienced world. The possibilities for making these connections are, however, manifold as these models analyse (through data) and operate on this world. Clements defines good educational tools (he calls them manipulatives) as *"those that aid pupils in building, strengthening, and connecting various* *representations of mathematical ideas"* [8]. Building on top of this definition, we add that good educational tools and activities for teaching computational models should also aid pupils in building, strengthening, and connecting the mathematical ideas behind (an inherent in) computational models to their experienced world and the critical questions related to these models. In his manner, they can support pupils in making sense of the possibilities and limitations of the models, and decode [9, 10] design agendas behind them.

This principle is visible throughout the intervention design: We worked with different ways of engaging with specifically KNN-models, to scaffold pupils in making sense of how they operate, and what they are used for. The data from the pop-quiz was used in all activities to connect the different representations. In the Embodied KNN exercise pupils embody how computational models aggregate and analyse their personal data, at least in a general sense. They can follow the steps in the algorithm, or they can just look around and see who is standing close to them to get a sense of the model. The interface of the Exploration Tool builds directly on top of pupils' experiences from the Embodied KNN exercise to support them in connecting the two and provide them with the opportunity to build and explore other models than the ones they experienced in the Embodied KNN exercise.

In the deployment, we saw how pupils were attentive to the position of each other in the Embodied KNN exercise and perceived themselves as being part of the discussed KNN models. They used this rich representation to engage with and make sense of the KNN models drawn on the whiteboard afterwards. Furthermore, we saw how pupils used the Embodied KNN to make sense of how to look for patterns in data with the Exploration Tool.

We did, however, also see that even though pupils were able to describe accurately what they did in the Embodied KNN activity, they struggled with connecting these experiences to the models they build with the Exploration tool. This illustrates how such embodied (or tangible) activities may only, on their own, support pupils in creating relationships that are sensory-concrete [8]. Without the support of the sensory experiences from being in the Embodied KNN exercise they could not make sense of the exact steps in the KNN algorithm. This hindered pupils in reflecting deeper on the designs of their own KNN models for The School for Gifted Children.

Despite pupils not forming concrete relationships to the exact steps of the KNN algorithm, the intervention did illustrate how embodied activities, construction/exploratory activities, and classroom discussions can be be interconnected through the design of activities and tools. How digital tools can be designed to extend embodied activities by supporting pupils in recalling experiences from these, and how embodied activities can engage pupils in computational models through social interactions and link them to their personal lives.

This specific set of activities, the connections between them, and choice to focus on the KNN model are just our current approach on how this principle can be applied. We see ample room for exploration of other approaches on how different ways of engagements with computational models can be designed to supplement each other.

8 CONCLUSION

In this paper, we have presented three design principles for tools and activities aiming to support the development of critical data literacy in children in K-9 education. We argue that educators and designers of such tools should aim to *make data personal and close to support emotional engagement*, *provide rich, contextualised representations of data to support reflections* and *connect different ways of engagement with computational modelling to support sense-making*. We arrive at these principles by bringing together educational tools and activities from a co-design process, empirical data from an intervention in an eighth-grade classroom, and didactic theories from maths education on how to enrich pupils relationship to concepts/objects. We hope that these principles will help promote and further critical data literacy, so as to empower future generations in our increasingly data-reliant societies.

9 SELECTION AND PARTICIPATION OF CHILDREN

The children were volunteered by their teacher, whom also participated in the co-design process. The children and their parents/legal guardians were asked for their written consent which was collected through a form. The form contained all details about the collection of data through notes, photographs and videos as well as the storage, analysis and sharing in aggregated and pseudonymised form of these data. The parents were also consented separately from this on whether pictures of their children could be used in publication about the intervention, and this publication only feature photos of children which parents consented to this.

REFERENCES

[1] Dor Abrahamson, Mitchell J. Nathan, Caro Williams-Pierce, Candace Walkington, Erin R. Ottmar, Hortensia Soto, and Martha W. Alibali. 2020. The Future of Embodied Design for Mathematics Teaching and Learning. *Frontiers in Education* 5 (2020), 147. https://doi.org/10.3389/feduc.2020.00147

[2] A Agassi, I Y Wald, H Erel, and O 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

[3] Stefan Baack. 2015. Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society* 2, 2 (2015), 2053951715594634. https://doi.org/10.1177/2053951715594634 arXiv:https://doi.org/10.1177/2053951715594634

[4] Tim Bell, Frances Rosamond, and Nancy Casey. 2012. Computer science unplugged and related projects in math and computer science popularization. , 398–456 pages. https://doi.org/10.1007/978-3-642-30891-8{_}18

[5] Rahul Bhargava, Erica Deahl, Emmanuel Letouzé, Amanda Noonan, David Sangokoya, and Natalie Shoup. 2015. Beyond data literacy: Reinventing community engagement and empowerment in the age of data. (2015).

[6] 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

[7] Leanne Bowler, Amelia Acker, and Yu Chi. 2019. Perspectives on Youth Data Literacy at the Public Library. (2019).

[8] Douglas H. Clements. 2000. 'Concrete' Manipulatives, Concrete Ideas. *Contemporary Issues in Early Childhood* 1, 1 (2000), 45–60. https://doi.org/10.2304/ciec.2000.1.1.7 arXiv:https://doi.org/10.2304/ciec.2000.1.1.7

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

[10] Daniella DiPaola, Blakeley H. Payne, and Cynthia Breazeal. 2020. Decoding Design Agendas: An Ethical Design Activity for Middle School Students. In *Proceedings of the Interaction Design and Children Conference* (London, United Kingdom) *(IDC '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3392063.3394396

[11] Paul Dourish. 2001. *Where the action is*. MIT press Cambridge.

[12] Liz Dowthwaite, Helen Creswick, Virginia Portillo, Jun Zhao, Menisha Patel, Elvira Perez Vallejos, Ansgar Koene, and Marina Jirotka. 2020. "It's your private information. it's your life.": young people's views of personal data use by online technologies. In *Proceedings of the Interaction Design and Children Conference*. ACM, 121–134. https://doi.org/10.1145/3392063.3394410

[13] Engida H. Gebre. 2018. Young Adults' Understanding and Use of Data: Insights for Fostering Secondary School Students' Data Literacy. *Canadian Journal of Science, Mathematics and Technology Education* 18, 4 (12 2018), 330–341. https://doi.org/10.1007/s42330-018-0034-z

[14] Samantha Hautea, Sayamindu Dasgupta, and Benjamin Mako Hill. 2017. *Youth Perspectives on Critical Data Literacies*. Association for Computing Machinery, New York, NY, USA, 919–930. https://doi.org/10.1145/3025453.3025823

[15] 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

[16] Magnus Høholt Kaspersen, Karl-Emil Kjær Bilstrup, and Marianne Graves Petersen. 2021. *The Machine Learning Machine: A Tangible User Interface for Teaching Machine Learning*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3430524.3440638

[17] 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.

[18] Paul Marshall, Sara Price, and Yvonne Rogers. 2003. Conceptualising Tangibles to Support Learning. In *Proceedings of the 2003 Conference on Interaction Design and Children* (Preston, England) *(IDC '03)*. Association for Computing Machinery, New York, NY, USA, 101–109. https://doi.org/10.1145/953536.953551

[19] 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

[20] Seymour Papert. 1980. *Mindstorms : children, computers, and powerful ideas.* Basic Books. 230 pages. https://dl.acm.org/citation.cfm?id=1095592

[21] Frank Pasquale. 2015. *The black box society.* Harvard University Press.

[22] Rosalind W Picard, Seymour Papert, Walter Bender, Bruce Blumberg, Cynthia Breazeal, David Cavallo, Tod Machover, Mitchel Resnick, Deb Roy, and Carol Strohecker. 2004. Affective learning—a manifesto. *BT technology journal* 22, 4 (2004), 253–269.

[23] Marc Uri Porat and Michael Rogers Rubin. 1977. *The information economy.* Number 77. Department of Commerce, Office of Telecommunications.

[24] Mitchel Resnick, Fred Martin, Robert Berg, Rick Borovoy, Vanessa Colella, Kwin Kramer, and Brian Silverman. 1998. Digital manipulatives: new toys to think with. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 281–287.

[25] Mitchel Resnick and Brian Silverman. 2005. Some Reflections on Designing Construction Kits for Kids. In *Proceedings of the 2005 Conference on Interaction Design and Children* (Boulder, Colorado) *(IDC '05).* Association for Computing Machinery, New York, NY, USA, 117–122. https://doi.org/10.1145/1109540.1109556

[26] Gabe Scelta, Hamid Rashid, Hoi Wai Jackie Cheng, Marcelo LaFleur, Mariangela Parra-Lancourt, Alex Julca, Nicole Hunt, S. Islam, and Hiroshi Kawamura. 2019. Frontier Technology Quarterly January 2019: Data Economy - radical transformation or dystopia? (01 2019).

[27] Marie-Monique Schaper, Rachel Charlotte Smith, Mariana Aki Tamashiro, Maarten Van Mechelen, Mille Skovhus Lunding, Karl-Emil Kjær Bilstrup, Magnus Høholt Kaspersen, Kasper Løvborg Jensen, Marianne Graves Petersen, and Ole Sejer Iversen. Under review. Principles for Teenagers' Learning About Emerging Technologies and Their Societal Impact: Machine Learning and Augmented Reality in K-12 Education. *Available at SSRN 4013380* (Under review).

[28] Sonja Špiranec, Denis Kos, and Michael George. 2019. Searching for critical dimensions in data literacy. (2019).

[29] Tom Symons and Theo Bass. 2017. Me, my data and I: The future of the personal data economy. (2017).

[30] Tiffany Tseng, Yumiko Murai, Natalie Freed, Deanna Gelosi, Tung D. Ta, and Yoshihiro Kawahara. 2021. PlushPal: Storytelling with Interactive Plush Toys and Machine Learning. In *Interaction Design and Children* (Athens, Greece) *(IDC '21).* Association for Computing Machinery, New York, NY, USA, 236–245. https://doi.org/10.1145/3459990.3460694

[31] Sherry Turkle and Seymour Papert. 1990. Epistemological pluralism: Styles and voices within the computer culture. *Signs: Journal of women in culture and society* 16, 1 (1990), 128–157.

[32] Philip Vahey, Ken Rafanan, Charles Patton, Karen Swan, Mark van 't Hooft, Annette Kratcoski, and Tina Stanford. 2012. A cross-disciplinary approach to teaching data literacy and proportionality. *Educational Studies in Mathematics* 81, 2 (9 2012), 179–205. https://doi.org/10.1007/s10649-012-9392-z

[33] H. Vartiainen, M. Tedre, and T. 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

[34] Uri Wilensky. 1991. *Abstract meditations on the concrete and concrete implications for mathematics education.* Epistemology and Learning Group, MIT Media Laboratory Cambridge, MA.

[35] Sarah Williams, Erica Deahl, Laurie Rubel, and Vivian Lim. 2014. City Digits: Local Lotto: Developing Youth Data Literacy by Investigating the Lottery. *Journal of Digital Media Literacy* (2014).

[36] Christine Ee Ling Yap and Jung-Joo Lee. 2020. "Phone apps know a lot about you!": educating early adolescents about informational privacy through a phygital interactive book. In *Proceedings of the Interaction Design and Children Conference.* ACM, 49–62. https://doi.org/10.1145/3392063.3394420

[37] A. Zimmermann-Niefield, S. Polson, C. Moreno, and R.B. 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

[38] 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