Enhancing Understanding of Data Traces and Profiling Among K–9 Students Through an Interactive Classroom Game

Juho Kahila
Henriikka Vartiainen
juho.kahila@uef.fi
henriikka.vartiainen@uef.fi
University of Eastern Finland
Joensuu, Finland

Eetu Arkko
Anssi Lin
eetu.arkko@uef.fi
anssi.lin@uef.fi
University of Eastern Finland
Joensuu, Finland

Nicolas Pope
Matti Tedre
nicolas.pope@uef.fi
matti.tedre@uef.fi
University of Eastern Finland
Joensuu, Finland

ABSTRACT

With the increasing popularity of social media among ever younger children, there is a growing need to educate young learners about the key concepts and mechanisms related to data collection, profiling, and content recommendation on online platforms. This paper presents a gamified K–9 classroom activity where learners analyze data traces to construct and iteratively refine detailed profiles of a mystery online person, thereby learning to recognize the digital traces typical of online interactions. Using data from 163 fifth- and eighth-grade students collected over 11 game sessions, the results demonstrate that the majority of learners were able to analyze and integrate various data traces to assemble coherent profiles, showing an emerging competence in recognizing data collection and profiling in online interactions, including their ubiquity, multimodality, and their consequences. The results also reveal that the learners' ability to critically reflect on profiling is still developing. This paper contributes to the fields of AI education and social media literacy by demonstrating the feasibility of early education on social media mechanisms.

CCS CONCEPTS

• **Social and professional topics → K-12 education**; *Computing literacy*; *Computational thinking*; **Computing education**; Children.

KEYWORDS

Artificial intelligence, Datafication, K-9, K-12, School, AI education, AI literacy, Social media literacy, Profiling, Tracking, Gamification

ACM Reference Format:
Juho Kahila, Henriikka Vartiainen, Eetu Arkko, Anssi Lin, Nicolas Pope, and Matti Tedre. 2024. Enhancing Understanding of Data Traces and Profiling Among K–9 Students Through an Interactive Classroom Game. In *The 19th WiPSCE Conference on Primary and Secondary Computing Education Research (WiPSCE '24), September 16–18, 2024, Munich, Germany*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3677619.3677635

WiPSCE '24, September 16–18, 2024, Munich, Germany

ACM ISBN 979-8-4007-1005-6/24/09
https://doi.org/10.1145/3677619.3677635

1 INTRODUCTION

As social media becomes increasingly integrated in people's daily lives, their user base is becoming ever younger, and tracking users starts at an ever younger age [1]. For children social media platforms are communication tools, entertainment platforms, information devices, and much more [41]. But with the benefits of social media platforms come data collection practices that—often against the EU's data protection legislation (GDPR)—track children, profile them, and use their data for manipulative practices, such as behavior engineering and targeted advertising [1, 17, 46]. The risks range from privacy invasions to cyberbullying, addictive behavior, and exposure to inappropriate content [2].

In response, media literacy educators have turned towards research on how to best educate children about social media, from particular risks and safe behaviors to media education broadly conceived [4, 29]. Some basic concepts related to social media platforms, tracking, and profiling can be introduced relatively early in schooling [43]. It is particularly important to educate those novice users, who are active in social media but may not be aware of the implications and outcomes of their activities on social media platforms, such as those concerning mis- and disinformation, privacy, and security [24, 37, 44, 46]. The earlier the platform mechanisms are demystified, the less room there is for developing their own "folk theories" of how social media systems work and why—theories that may be simple and intuitive, yet false [9, 12, 25].

In the field of computing education research, a quickly growing body of work addresses educational questions related to the mechanisms of social media and datafication [5, 13, 15, 30, 33, 35, 36, 43, 45, 51]. Approaches in computing education research include, for example, data literacy [37], data awareness [18–20], data agency [55], AI literacy [30], AI competency [5], and many more. Common to those is that understanding how to use AI is not enough, nor is superficial folk-theory-based "control" over the systems (cf. [4]). Young users of social media and online platforms should understand the mechanisms and key concepts of their data-driven practices, but curricula rarely cover those [32, 54]. Understanding the mechanisms, not just uses, of social media is crucial for developing critical media literacy for the social media era [4], including responsible online behavior, strengthened data agency, and taking control of one's digital self [51]. However, a recent systematic survey of AI education tools in K–12 showed a lack of collaborative classroom tools with low barrier of entry [59].

To address the need for learning innovations in the K–9 AI and social media education space, this paper presents a classroom activity, tailored for young learners, that aims to teach a number of basic concepts and principles central to how social media works. Those concepts are data collection, user profiling, and content recommendation. The classroom activity was designed around a tailor-made and gamified system that introduces children to the concepts above. The study posed two research questions:

RQ1: To what extent can children recognize the variety of data traces associated with online activities?
RQ2: How well do children understand data-driven profiling and its constraints?

2 METHODOLOGY

2.1 Context and study design

This study is part of a series of design-based research studies focused on integrating AI topics into education [38–40, 55, 57]. It belongs to a large scale research program that aims to foster children's data agency, that is, people's volition and capacity for informed actions that make a difference in their digital world [49]. In the program, researchers, software developers, and schoolteachers co-design, in hands-on workshops, educational technologies, pedagogical models, and scaffolding strategies for teaching AI themes to schoolchildren. The first round of the school interventions took place in spring 2023 [22]. This paper reports on a second round of interventions, designed to demonstrate the basic mechanisms of social media to children and conducted in spring 2024.

The design of the classroom activities was based on collaborative learning and inquiry that engage children in interest-driven and knowledge creative activities (e.g., [14, 42]). Furthermore, the design of the activities and educational technology incorporated elements of game-based learning with a narrative story embedded in the game [28]. By that way, the aim was to make the learning activities more interesting, fun, and engaging [47], and thus, sustain students' interest in completing the task designed to enhance their understanding of the key concepts.

2.2 Participants

The data for this study were collected from workshops in 11 Finnish schools, one class in each school. This included six fifth-grade classes and five eighth-grade classes. The same participants had attended workshops the previous year to learn basic AI concepts [22, 55]. Nearly all participants used social media at least weekly, with WhatsApp, YouTube, Snapchat, Spotify, and TikTok being the most popular platforms among them [23]. The workshop was integrated into regular school activities and each class had children who were involved in the data collection and some who were not involved. A total of 163 students (100 fifth graders and 63 eighth graders) participated in this part of the study. The study followed the guidelines set by the National Board on Research Integrity [50]. A research permit was secured from local educational authorities, and guardians provided informed consent. Children were explained the same verbally in the classroom. The participant information sheet summarized the study's goals, voluntary participation, withdrawal rights, data collection and processing, and result reporting. It also asked permission to publish and present in academic settings the digital artifacts created in the workshops.

Figure 1: Children engaged in the classroom activity. Teacher's screen shows a hint and students' devices show a range of profile elements to adjust.

2.3 The Classroom Activity Setup

The classroom game introduces a mystery person χ, and presents hints that gradually reveal more and more about the person. (The person's profile was created by teachers and researchers, and is not a real person.) The children's job is to describe the person as closely as possible: Age, family members, interests, hobbies, home address, and daily routines. The classroom game uses two apps: One for a classroom view, projected on the classroom screen and controlled by the teacher, and one for a children's "profiling" app that are all peer-to-peer connected to the teacher's computer (Fig. 1).

Firstly, the activity introduces children to different types of data and the insight that everything one does on any online platform is a potential source of data. Examples include typical actions on social media platforms, such as liking a post or commenting on a photo, stopping image feed for just a while, and watching a video. Each of these actions generates data points that, when collected and analyzed, can reveal insights about a user's preferences, interests, and online behavior—and even about one's moods and daily non-online routines. More subtle sources of data include the time spent on different tasks online and location data that maps out the physical movements of users. The activity introduces children to data collection in terms of multi-modality (all interactions with online platforms generate data about the user) and tracking (pervasive and systematic monitoring of user's online behavior [61]).

Secondly, the activity introduces children to the idea that data from different sources can be combined to create increasingly detailed profiles of users. It presents profiling as a process that gathers a practically limitless, comprehensive collection of user's data, as well as online interactions of all kinds, aimed to faithfully describe the user through the user's behaviors, preferences, and past activities. It presents the idea of inferring, from user profiles, information about users that cannot be directly collected. Profiling is connected to recommending, as one use case for user profiles created, with the insight that profiles are not limited to viewing a user's past: They can be used to predict user's future actions, as well as their interests, behaviors, secrets, and even moods and daily routines. Children are also introduced to the differences between human-made profiling relying on human intuition, and automated profiling as done by online platforms.

Three distinct classes of data are involved: volunteered data ("data given"), observed data ("data traces"), and inferred data [53, 60]. *Volunteered data* comprises information that users willingly and consciously share with the platform provider. These data points range from basic profile details, like name, email, and phone number, to dynamic content such as posts and videos users share on social media, and they are often a requirement for social interaction with other users, or for using the platform at all [37, 53, 60].

Unlike volunteered data, *observed data* are collected subtly from user interactions with the platform, often without the user's explicit awareness, although in most cases under the guise of consent via end-user license agreements [37, 53, 60]. This category encompasses a broad spectrum of data, including online behaviors, interaction data, and device information [21]. A sufficiently large quantity of those data allows for the creation of detailed behavioral and demographic profiles, including, for example, demographic variables (age, place of residence, income bracket, religion), psychometric variables (lifestyles, cultural and political preferences), and behavior profiles (consumption habits, eating habits, sleep patterns).

The third category, *inferred data*, extends beyond direct collection of user data, using collected data to statistically or otherwise create assumptions about an individual [60]: Not just the missing bits of data but also preferences, habits, and potential future actions [7, 26]. Inferred data enables useful functions for many modern applications, from recommending friends on social media to suggesting products based on purchase history—but it also enables advertisement, behavior engineering, manipulation, and exploitation [8, 16, 27].

2.4 Workshop Organization and Data Collection

The workshop lasted roughly two 45-minute class periods with a short break in between. Researchers and teachers collaborated in teaching. The workshop was led by a researcher who had subject expertise on social media's mechanisms, and who is also an experienced classroom teacher (hereafter referred to as the teacher). Schoolteachers' knowledge of their students played a pivotal role; they supported the students' collaboration and regulated their activities during the workshop.

The workshop was divided into three stages: 1) Introducing basic concepts related to data collection and profiling and contextualizing them to the participants' daily lives. 2) Playing the profiling game. 3) Reflecting on the game lessons and relating them to learners' daily lives.

1. Introduction and contextualization. The workshop began with the researchers briefly introducing themselves and recalling what was done in previous year's workshops [22, 55]. After this, participants were introduced to this year's topic: the basic mechanisms of social media. They were also introduced to the key concepts of the workshop: data, profile, and recommender. Additionally, there was a discussion about how this year's topic relates to what was learned in the previous year's workshops. This was followed by a discussion about what kinds of data online digital services collect from users, mentioning various concrete examples, such as what you have watched, what you have liked, whom do you follow on social media, and what you have purchased or searched for online.

After discussing data collection, the discussion moved to how the data collected from users can be used to profile users and offer them personalized content. This was illustrated with two examples presented to the participants. The first example involved a person visiting an online store to see gaming headphones, a console, and a video game. The teacher explained that the store now has data on who viewed these items. The discussion then focused on what the data revealed about this person and how confident can one be about what they inferred about that person. The teacher explained

Table 1: Data Source Examples Involved in the Activity

Data source	Activity
Listening to music	The person χ listened to a music track in Spotify.
Tagging and engaging	χ was tagged in a picture on Instagram and χ engaged with the picture.
Creating content	χ added a picture to a Snapchat story.
Watching a video	χ watched a TikTok video and engaged with it.
Following	χ follows four people in social media.
Social recommendations	Five friends recommend χ a video on Whatsapp.
Engagement	χ stopped scrolling to see an ad on Instagram feed.
Search	χ searched for information on Google.
Reactions	χ reacted to a local news with an emoji.
Location tracking	χ regularly visits four places at different hours of day/week.

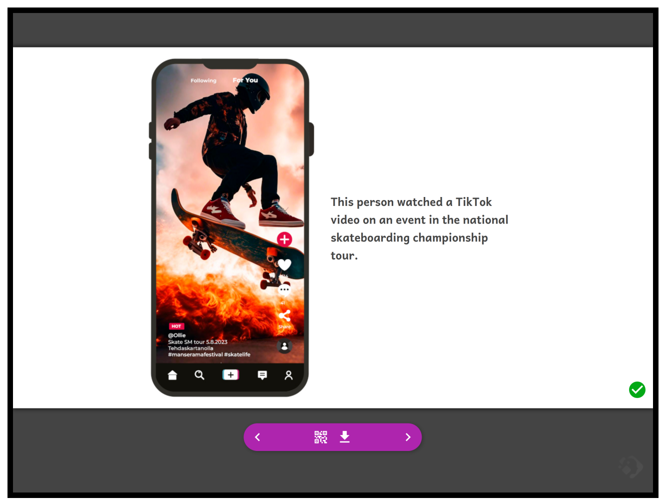

Figure 2: Example view of visual and verbal hints shown to children on the classroom projector.

that companies can use data about visitors to create user profiles of them.

The teacher also introduced another example, which demonstrated that data can also be used to predict a person's next actions. In this example, a person purchased flip-flops, sunscreen, and a swimsuit in an online store. Children then discussed what alternative next actions this purchase history would suggest (Buying a beach towel? Heading to the beach?). They discussed also how those predictions can turn out to be completely wrong (Were the purchases a gift to someone else?). The teacher emphasized that conclusions drawn from data are educated guesses based on the data—and adding data can improve the guess.

Next, the teacher explained that ads we see online are often based on our past activities. The students then suggested what could be advertised to the individuals in the previous examples. Finally, the teacher explained that social media data can even be used to predict a person's mood. As an example, the teacher showed three pictures that a person had posted: cakes, a happy dog meme, and a beach selfie, and then asked the students to suggest the person's likely mood.

This introduction to the topic was given as an interactive lecture with slides. Topics were discussed using concrete examples from the children's daily lives, aiming to increase the theme's relevance to them by connecting it with their everyday experiences [3].

2. The profiling game. The introduction was followed by playing a profiling game. The teacher explained the game's goal: to use hints from a person χ's data traces to build a detailed profile of that person. Students worked in pairs formed by the teachers. Each pair used one device and joined the game using a QR code. The game featured ten data sources (Table 1) about person χ, shown sequentially on a large screen by the teacher (Fig. 2). New questions (Table 2) appeared on the students' devices throughout the game (Fig. 3). Participants could refine earlier responses at any time, and they were reminded that new data might clarify previous questions. After all data sources were displayed, the game showed all ten data sources together on one screen, and students had a moment to further refine the profiles they had made.

After children had completed the profiling (**Task 1**), new questions appeared on their screens. The questions asked were:

Task 2: "What profile elements were you certain of? Justify your answer."
Task 3: "Which data traces were the most useful for profiling? Justify your answer."
Task 4: "What additional data would you need to profile the person better? Justify your answer."

At the end of the game session, the profiles created by the children were projected on a large screen, using the game's virtual bulletin board feature (Fig. 4). In addition, the teacher revealed a complete "correct" profile of χ, allowing the children to evaluate the profiles they had created. Teachers actively observed the students' activities and provided on-demand support. For example, they assisted the students by rephrasing and explaining questions offered by the game or giving hints on where the children should focus.

Table 2: The profile questions answered by the children, added incrementally at the different data trace slides.

Trace	Question
1	How old is the person?
2	What can you infer about the person's family?
3	What can you infer about where the person lives?
4	What can you infer about the person's interests?
9	What can you infer from the person's reaction?
10	What can you infer from the person's daily routine?

3. Reflecting and relating the lessons to one's daily lives. After the game session, students were introduced to two basic mechanisms of social media: They discussed what content could engage χ to spend more time on that platform, and what products or services could be advertised to χ. The teacher then explained the differences between human-made and AI-made profiling: Humans rely on their personal experiences, intuition, and ability to psychologically identify as another person—whereas the algorithms on social media platforms lack all those but can track many more data points, millions of them, and compare them with the data and behaviors of millions of other users. The teacher emphasized that social media constantly

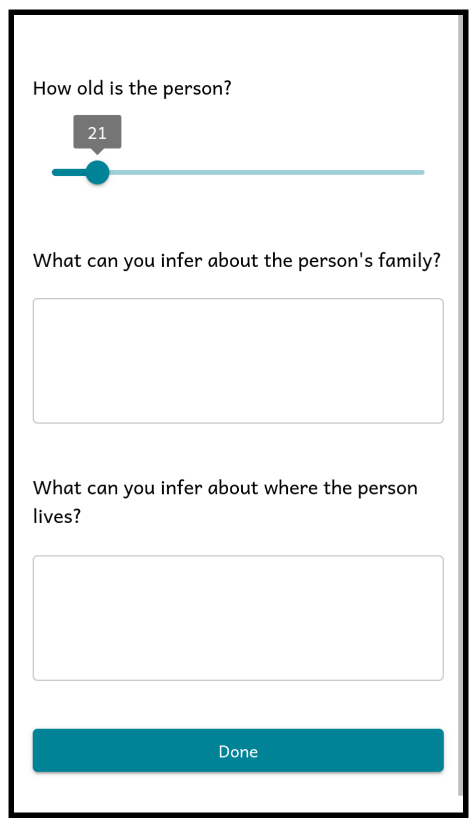

Figure 3: Screen shot of children's app, on which they gradually develop a profile of the mystery person χ on a smartphone or school laptops.

profile users automatically. Everything the children, their friends, and family do on these platforms generates data about them. The teacher then instructed the students to discuss in pairs and list all the data that can be collected from their social media use, what can be inferred from those data (like they inferred in the game), and for what purposes do they think these inferred data are used. In this final task, where participants discussed what they had learned in the workshop and reflect on that in the context of their everyday lives, the goal was to contextualize the content to the children's interests, experiences, and everyday lives, thus making the lessons more meaningful to them [3].

2.5 Analysis of Data

An inductive approach to qualitative content analysis [11] was used to analyze the extent to which children can recognize the variety of data traces associated with their online activities and to understand data-driven profiling and its constraints. Coding categories were developed and finalized between three authors for each task from Task 1 to Task 4. The entire data set was then coded by two other authors, with agreement ranging from 76.1% to 84.5%. Cohen's kappa [6] was calculated for the inter-rater reliability of coding. Interrater reliability was substantial in two out of four tasks ($\kappa_2 = 0.737, SE_2 = 0.067$ and $\kappa_4 = 0.647, SE_4 = 0.080$) and moderate in two ($\kappa_1 = 0.576, SE_1 = 0.090$ and $\kappa_3 = 0.572, SE_3 = 0.084$). Disagreements were discussed and minor changes to categorization were made.

2.6 Technical Description of the System

The classroom system consists of two parts: A teacher's view run on the teacher's computer and shown on the classroom projector, and children's view run on children's school laptops or mobile devices. The teacher's view starts with a single observed data point: a music track the mystery person χ listened to on Spotify (Table 1). Children's app starts with a task to write down what they believe an age range for person χ to be. Children can discuss in their group and mark their view on their app. The teacher then proceeds to the next observed data points: a picture on Instagram where person χ was tagged, and χ engaged with that picture.

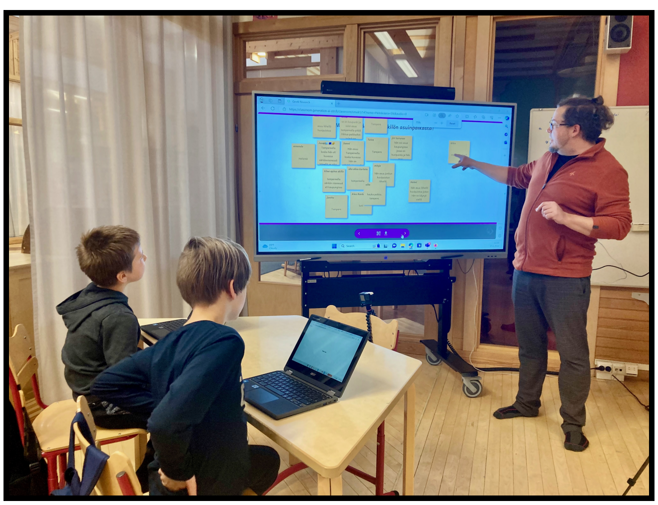

Figure 4: Profiles created by the children, shown on the classroom screen.

The two applications use React to create a responsive browser based app that operates on a wide range of devices that are typically found in a classroom. WebRTC is used to enable peer to peer communication between devices within the classroom to ensure the data are end-to-end encrypted and kept within the local WiFi network if possible. The teacher machine generates a QR or text code that, using signaling servers, allows each device to connect with that teacher machine directly. Question and slide data are loaded from JSON data files that describe the images, text, and question boxes to be presented on each screen, allowing easier customisation of the contents. As children improve their profile with each new data trace, a snapshot of the current status of their profile is saved, yielding ten snapshots over the course of the game. The app was also used for collecting data about children's reflecting and relating the lessons to their daily lives. All edits along with the final answers from each of the children's devices are logged to the teacher's machine for download at the end of the session if necessary. No data are sent to a server. In the event of a problem the children can reconnect to their game by downloading the log data for their session from the teacher machine and resuming the activity (something which happens if the browser page is accidentally closed). The app is freely available for download and further development at https://github.com/knicos/genai-profiling.

3 RESULTS

Over the game session, students profiled mystery person χ by using ten data traces and supporting questions provided by the teacher (task 1). Their answers were categorized into three categories based on how well the children were able to describe the person χ. Figure 5 shows the results for task 1.

Out of the children's responses to task 1, 4.8% lacked an adequate description of data involved in the exercise (category 0). Most of those were completely empty, or contained very short phrases or comments. One third (33.3%) of responses fell into category 1. Responses in that category were mostly made at a very general level and expressed as statements that identified some profile elements from the example data traces but lacked detail; For instance, one learner filled in: "*he is a brother in a family*," "*He lives in Tampere*," "*skateboarding, ice hockey, gaming and mopeds*," "*He walks the dog, skates and goes to school*," and "*he was angry*." (Quotes are translations from Finnish language.)

The majority of children's answers (61.9%) fell into category 2: They were able to extract many kinds of information from the example data traces given by the teacher. Compared to categories 0 and 1, responses in category 2 had significantly more depth and detail. One group from grade eight was able to determine the hometown, neighborhood and even the street address of person χ along with diverse answers to other questions as well:

Age *15*
Family *The person's younger sister has a birthday, and they are celebrating it with a nice cake. They live near trees. Their family is wealthy.*

Hometown	*He lives in Tampere by Lake Näsijärvi at Marsipaanitie 7. Särkänniemi is in Tampere.*
Interests	*He enjoys skateboarding and follows competitions. He is into sports and particularly enjoys ice hockey, following both the sport and its players enthusiastically. He watches YouTube videos, especially gaming content. His friend is called 'Pätkis', and they enjoy gaming together while watching gaming videos. He plans to get a moped license and is searching online for information about it. He also listens to music.*
Activities	*The person attends school five times a week, goes skateboarding three times a week, and visits the dog park with their dog twice a week.*
Reaction	*He enjoys skateboarding and dislikes the city's decision to close the skate ramp. He is angry, sad, and frustrated about it.*

Most children were able to extract coherent information from many kinds of data traces with varying levels of detail, and were able to profile the mystery person well. Most responses presented the results without expressing uncertainty, while others included phrases like "most likely" and "we think that" to express varying degrees of confidence.

3.1 Reflections on the profiling task

The children were asked to reflect on the profiling they had performed and think of ways they could improve the profile (Tasks 2, 3, and 4). Furthermore, they were asked to justify their responses. The responses were categorized into four categories, described in Figure 6. The responses to tasks 2 through 4 somewhat depended on the profile students had created, as those pairs who had not created an adequate profile typically had difficulty trying to reflect and improve on it.

Task 2. Confidence. Task 2 asked, "Which [profile elements] are you certain of, and why? Which ones are you not? Justify your answer." In their answers the children described the extent to which they were more or less sure about their profile elements, and justified their answers. Three in five (59.5%) children described their confidence in their evaluation in a relatively shallow way without justifying their answers. In category 1 children, for example, pointed out some basic details about χ which they were sure of, but did not specify why: One pair of fifth graders wrote that χ likes skateboarding, lives in Tampere, and follows ice hockey "*because those things were self-evident.*"

Almost one in five children (18.5%) were able to describe their confidence in their evaluation of profile elements in a rich manner that includes multiple perspectives. In category 2 answers, children pointed out, for instance, that they are more sure about multiple things about mystery person χ, which they were sure of: "*We can be sure that the person uses social media, listens to music, and lives in Finland because we have data that proves these facts. However, we cannot be certain about the person's age, gender, or all of their interests because we don't have enough data to confirm these*" (eighth grade students).

In category 3 answers, children described their confidence in their profile elements in a rich, justified way that included multiple perspectives. Those were found in 11.3% of children's responses.

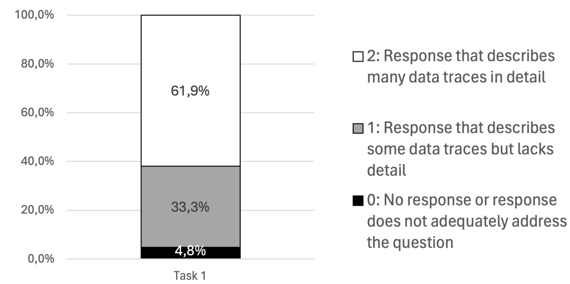

Figure 5: Children's responses to task 1, categorized in three categories (N=82 pairs of children).

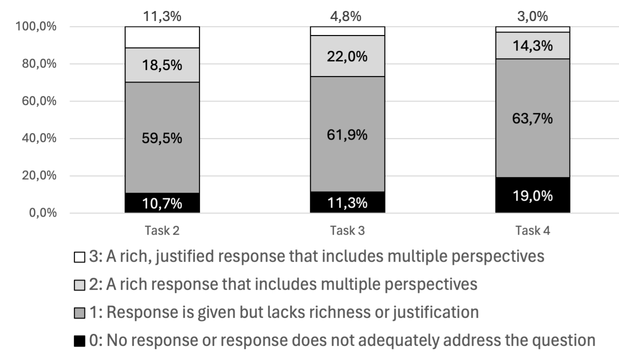

Figure 6: Children's responses to tasks 2 through 4, categorized in four categories (N=82 pairs of children).

For instance, a pair of eighth grade students wrote, "*Sure about home town, because Särkänniemi* [a popular amusement park] *is there. Sure about Spotify use. Sure about him/her liking skateboarding, because there's a lot of skating-related data. Also owns Whatsapp, Snapchat, and Instagram. Also map service. Unsure about age, because it's hard to estimate that from those* [data traces]. *Unsure about family members; there's little data* [about them]."

Task 3. Relative usefulness of different data traces. Task 3 asked, "Which data traces were the most useful for profiling? Justify your answer." Out of children's responses to task 3, 11.3% fell into category 0 (Fig. 6). Most responses (61.9%) listed a few data traces without explaining why they were the most useful for profiling (category 1)—for example, "*The last picture, skateboarding, and moped license thingy because we got most information from them.*" (Fifth graders).

Category 2 contained 22.0% of responses to task 3. Those responses typically weighed on the importance of different data traces or presented a rich variety of data traces. For example, fifth graders wrote, "*The best pictures were the skateboarding picture and the picture showing what he does during the week. It was very informative* [to know] *when it was when he moves where.*" Category 3 (4.8% of responses) typically explained what information bits for the profile could be extracted from each data trace. For example, eighth graders wrote that the most useful data traces were "*A picture with* *the amusement park ride, as it shows the location. The schedule picture, as you can see from it that he is of school age. The skateboarding picture, as from that you can see what his interests are.*"

Task 4. Improving the profile. In task 4 the children answered question "What additional data would you need to profile the person χ better? Justify your answer." Task 4 had a substantial amount of answers in category 0 as 19.0% of the responses did not provide a adequate response. Most of the category 0 responses were left completely blank.

Category 1 contained 63.7% of the responses, where the children did present a relevant answer but the answers were typically simple and did not contain deeper reflection. Many answers in category 1 were given at a very general level and did not further elaborate what kind of data they would need specifically: "*More pictures of the person and more information about his life*" (Fifth grade).

Responses in category 2, including 14.3% of responses, were more specific and presented more data sources than responses in category 1, but they often lacked justification or explanation: "*age, picture of the person's face, hobbies, interests and family, because you can get a lot of information from them*" (Fifth grade). Responses in category 3 (3.0% of all responses) showed deep reflection on what additional data would be needed to improve the profile of person χ. In those responses children were able to provide multiple data sources that could be used to improve the profile, and children justified their answers well: "[The profile would benefit from data about] *other things he has bought online and things he has liked in social media and pictures he has uploaded. What the person looks like, gender for certainty. Because the style and spending behavior of the person can be deduced from what the person bought online. You can get more information about interests from likes on social media. Uploaded pictures can for example tell where the person has been*" (Fifth graders).

The responses suggested data that would be useful for human-made profiling—such as photos of the person—but also included data useful for automatic profiling—such as online purchase history. However, as the differences between profiling performed by humans and by computers were only discussed after the profiling game, the children did not raise this distinction in their description of data traces.

The reflection task on profiling (tasks 2-4) proved to be more difficult than the profiling task itself was (task 1). As the abstraction level of the tasks increased, the number of answers in category 3 decreased and the answers in category 0 increased. However, the number of answers in category 1 remained high throughout tasks 2 to 4, indicating an emerging understanding of data-driven profiling as a phenomenon. The study did not, however, measure children's learning of the difference between profiling made by humans and profiling made by computers; that topic was presented after the intervention and was not commensurably measured.

4 DISCUSSION

This study set out to develop, test, and evaluate an interactive, gamified activity for helping fifth and eighth -grade children to understand how data collection, profiling, and recommending work on online platforms. The study introduced a classroom game to make these concepts more tangible. The classroom activities showed students how everyday actions on social media, like liking a post or watching a video, generate data that can reveal a lot about their personas, preferences, moods, and behaviors. The activity also taught children how these data can be combined to create detailed profiles of them, and how those profiles can be used to predict interests and future behaviors. The activity also introduced the differences between human-made and automated profiling.

Related to RQ1 ("To what extent can children recognize the variety of data traces associated with online activities?") the breakdown of responses to task 1 shows that children had little to no difficulty recognizing a variety of data traces associated with online activities, or combining them to build a relatively good profile of person χ. The majority of responses (61.9%) contained more or less coherent descriptions of how data traces come together to form a profile of a person.

Related to RQ2 ("How well do children understand data-driven profiling and its constraints?"), children's responses to task 2 show nascent understanding of how information inferred from data traces depends on the variety, quality, and quantity of data given and data observed: 29.8% of responses belonged to category 2 or 3 (rich responses that include multiple perspectives). Similar to task 2, also children's weighing of the relative usefulness of different data traces for profile-building (task 3) were promising: 26.8% of responses to task 3 belonged to category 2 or 3 (rich responses that include multiple perspectives).

Children's ideas for what data would be useful for improving the profile (task 4) were mostly a few mentions of additional data traces (63.7%) but still 17.3% of responses gave a rich description that contained a variety of perspectives.

5 RECOMMENDATIONS FOR FUTURE DESIGNS

The study gave rise to a series of design recommendations, derived from our findings and observations, and supported by learning theory. The recommendations are aimed at enhancing the educational experience in ways that stimulate children's critical thinking and engagement. Through these recommendations we aim to equip the reader with actionable ideas for classroom activities that engage children in learning about data traces, profiling, and recommending.

1. Connect disciplinary concepts (e.g., data collection, profiling, recommending) with students' interests and everyday experiences. Consider designing interventions that connect learning activities to the interests and lived experiences of the children. In practice, promoting and sustaining children's curiosity and interests can be fostered by providing opportunities to explore real-life data that relate to the world in which the children live, learn, and play [28]. Connections with students' everyday experiences and the tools they already use also provide a meaningful context for exploring new computational concepts and practices, to see why concepts need to be learned, and for relating those concepts to critical awareness [34]. This may involve incorporating relatable examples to investigate, such as real-life online services, data-traces and storylines with familiar characters (e.g., other children) and their daily activities [28].

2. Promote playful peer-to-peer learning of disciplinary concepts (e.g., data collection, profiling, recommending). Consider providing ways for children to engage in collaborative reasoning by making and sharing their own observations and explanations in relation to data-traces of everyday life. To foster social interaction and collaborative learning, children can be guided to work in pairs and to share their thoughts and explanations with the whole class so that the inquiry becomes a collaborative endeavor. In this process, the tools and technologies can be utilized to sustain interest and motivation by gradually unveiling new data-traces and by presenting question prompts that ask students to externalize, articulate, and reflect their lines of thinking. In these tool-mediated activities, the learners can also be specifically prompted to collaboratively construct and reconstruct their evolving understanding by refining and evaluating the profiles which they are building.

3. Provide on-demand, tailored, and distributed scaffolding for learning disciplinary concepts (e.g., data collection, profiling, recommending). Acknowledge that teachers play a crucial role in helping students complete problem-solving tasks in game-based learning, with the overall aim of developing their conceptual understanding [47]. Even though new concepts and lines of reasoning can be explicitly mediated and scaffolded by educational technology and related prompting questions, the teachers' support is needed for mindful and productive engagement with the tasks, tools, and peers [48]. In particular, teachers' support is needed for providing an explanatory rationale for why understanding of the concepts is relevant to students' life and future [3], which can be fostered by using real life examples in teachers presentations as well as in dialogic discussions that bridges the disciplinary concepts with learners' interest, intuition, and evolving understanding. When completing the game activities, the array of scaffolding means provided by a teacher may also include, for example, listing student interests, asking probing questions, providing hints or feedback, highlighting the critical features of the task, and offering tailored, on-demand guidance [3, 52].

After the game activities, the teachers also play an important role in supporting critical reflections and discussions that connect children's own observations and explanations with the societal and ethical implications of AI and data-driven practices. As noted by Duncan-Andrade and Morrell [10], consciousness and comprehension are also an important prerequisite for critique, since it is impossible to critique existing relations without being aware of or understanding them. This does not necessarily require adopting a negative stance or telling children "do's and don'ts", but providing children with intellectual tools, discourses, and tasks that can help them to analyze, understand, and critique the practices and systems that shape their everyday lives [28, 58]. Building a thematic continuity—using learning activities, curriculum materials, tools, peer discussions, and teacher's facilitation [48], for example—may also facilitate the creation of a rich web of memorable associations between new concepts and everyday experiences. This kind of critical re-reading of the world may generate opportunities to build new ways of knowing and acting upon the world [31], pave the way for envisioning and pursuing informed actions towards alternative possibilities [58], and uncover new avenues for developing one's data agency [56].

ACKNOWLEDGMENTS

Funded by the Strategic Research Council (SRC) established within the Research Council of Finland (Grants #352859 and #352876). The authors thank the January Collective. We extend our heartfelt thanks to the teachers and children who actively participated in developing and testing this material. We thank Teemu Roos for inspiring ideas on AI education, plugged and unplugged.

REFERENCES

[1] Veronica Barassi. 2020. *Child data citizen: How tech companies are profiling us from before birth.* The MIT Press, Cambridge, MA, USA.

[2] Elena Bozzola, Giulia Spina, Rino Agostiniani, Sarah Barni, Rocco Russo, Elena Scarpato, Antonio Di Mauro, Antonella Vita Di Stefano, Cinthia Caruso, Giovanni Corsello, and Annamaria Staiano. 2022. The Use of Social Media in Children and Adolescents: Scoping Review on the Potential Risks. *International Journal of Environmental Research and Public Health* 19, 16 (2022). https://doi.org/10.3390/ijerph19169960

[3] ChanMin Kim Brian R. Belland and Michael J. Hannafin. 2013. A Framework for Designing Scaffolds That Improve Motivation and Cognition. *Educational Psychologist* 48, 4 (2013), 243–270. https://doi.org/10.1080/00461520.2013.838920

[4] David Buckingham. 2019. *The Media Education Manifesto.* Polity Press, Cambridge, UK.

[5] Thomas K.F. Chiu, Zubair Ahmad, Murod Ismailov, and Ismaila Temitayo Sanusi. 2024. What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open* 6 (2024), 100171. https://doi.org/10.1016/j.caeo.2024.100171

[6] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates, Hillsdale, NJ, USA.

[7] Michael D. Conover, Bruno Goncalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011. Predicting the Political Alignment of Twitter Users. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing.* 192–199. https://doi.org/10.1109/PASSAT/SocialCom.2011.34

[8] Kate Crawford. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* Yale University Press, New Haven, CT, USA.

[9] Michael A. DeVito, Jeremy Birnholtz, Jeffery T. Hancock, Megan French, and Sunny Liu. 2018. How People Form Folk Theories of Social Media Feeds and What it Means for How We Study Self-Presentation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, Canada) *(CHI '18).* ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173694

[10] Jeffrey M. Duncan-Andrade and Ernest Morrell. 2008. *The Art of Critical Pedagogy: Possibilities for Moving from Theory to Practice in Urban Schools.* Number 285 in Studies in the Postmodern Theory of Education. Peter Lang, New York, NY, USA.

[11] Satu Elo and Helvi Kyngäs. 2008. The qualitative content analysis process. *Journal of advanced nursing* 62, 1 (2008), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

[12] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I "like" it, then I hide it: Folk Theories of Social Feeds. In *Proceedings of the 2016 CHI conference on human factors in computing systems.* San Jose, CA, USA, 2371–2382.

[13] Shuchi Grover. 2024. Teaching AI to K-12 Learners: Lessons, Issues, and Guidance. In *Proceedings of ACM Computer Science Education (SIGCSE) 2024 Conference.* ACM, Portland, OR, USA, 1–7.

[14] Kai Hakkarainen, Sami Paavola, Kaiju Kangas, and Pirita Seitamaa-Hakkarainen. 2013. Sociocultural Perspectives on Collaborative Learning: Toward Collaborative Knowledge Creation. In *The International Handbook of Collaborative Learning*, Cindy E. Hmelo-Silver, Clark A. Chinn, Carol K. K. Chan, and Angela M. O'Donnell (Eds.). Routledge, London, UK, 57–73.

[15] Fredrik Heintz and Teemu Roos. 2021. Elements Of AI - Teaching the Basics of AI to Everyone in Sweden. In *Proceedings of the 13th International Conference on Education and New Learning Technologies (EDULEARN21).* IATED, Online, 2568–2572. https://doi.org/10.21125/edulearn.2021.0559

[16] Vincent F. Hendricks and Mads Vestergaard. 2019. *Reality Lost: Markets of Attention, Misinformation and Manipulation.* Springer Open, Cham, Switzerland.

[17] Arne Hintz, Karin Wahl-Jorgensen, and Lina Dencik. 2018. *Digital Citizenship in a Datafied Society.* Polity Press, Cambridge, UK.

[18] Lukas Höper and Carsten Schulte. 2023. The data awareness framework as part of data literacies in K-12 education. *Information and Learning Sciences* (2023). https://doi.org/10.1108/ILS-06-2023-0075

[19] Lukas Höper and Carsten Schulte. 2024. Empowering Students for the Data-Driven World: A Qualitative Study of the Relevance of Learning about Data-Driven Technologies. *Informatics in Education* (2024). https://doi.org/10.15388/infedu.2024.19

[20] Lukas Höper, Carsten Schulte, and Andreas Mühling. 2024. Learning an Explanatory Model of Data-Driven Technologies can Lead to Empowered Behavior: A Mixed-Methods Study in K-12 Computing Education. In *Proceedings of the 2024 ACM Conference on International Computing Education Research (ICER '24)*. ACM, New York, NY, USA. https://doi.org/10.1145/3632620.3671118

[21] Jakob Jünger. 2018. Mapping the Field of Automated Data Collection on the Web: Collection Approaches, Data Types, and Research Logic. In *Computational Social Science in the Age of Big Data: Concepts, Methodologies, Tools, and Applications*, Cathleen M. Stuetzer, Martin Welker, and Marc Egger (Eds.). Herbert von Halem Verlag, Köln, Germany, 104–130.

[22] Juho Kahila, Henriikka Vartiainen, Matti Tedre, Eetu Arkko, Anssi Lin, Nicolas Pope, Ilkka Jormanainen, and Teemu Valtonen. 2024. Pedagogical framework for cultivating children's data agency and creative abilities in the age of AI. *Informatics in Education* (2024). https://doi.org/10.15388/infedu.2024.15

[23] Juho Kahila, Henriikka Vartiainen, Matti Tedre, and Sonsoles López-Pernas. 2024. Children's Data Privacy Fears on Social Media. *Submitted, in review* (2024).

[24] Caroline Keen. 2022. Apathy, convenience or irrelevance? Identifying conceptual barriers to safeguarding children's data privacy. *New Media & Society* 24, 1 (2022), 50–69. https://doi.org/10.1177/1461444820960068

[25] Daniel Klug, Yiluo Qin, Morgan Evans, and Geoff Kaufman. 2021. Trick and Please. A Mixed-Method Study On User Assumptions About the TikTok Algorithm. In *Proceedings of the 13th ACM Web Science Conference 2021* (Virtual Event, United Kingdom) *(WebSci '21)*. ACM, New York, NY, USA, 84–92. https://doi.org/10.1145/3447535.3462512

[26] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110, 15 (2013), 5802–5805. https://doi.org/10.1073/pnas.1218772110

[27] Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111, 24 (2014), 8788–8790. https://doi.org/10.1073/pnas.1320040111 arXiv:https://www.pnas.org/content/111/24/8788.full.pdf

[28] Priya Kumar, Jessica Vitak, Marshini Chetty, Tamara L. Clegg, Jonathan Yang, Brenna McNally, and Elizabeth Bonsignore. 2018. Co-designing online privacy-related games and stories with children. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (Trondheim, Norway) *(IDC '18)*. ACM, New York, NY, USA, 67–79. https://doi.org/10.1145/3202185.3202735

[29] Sonia Livingstone. 2014. Developing social media literacy: How children learn to interpret risky opportunities on social network sites. *Communications* 39, 3 (2014), 283–303. https://doi.org/doi:10.1515/commun-2014-0113

[30] 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)*. ACM, New York, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376727

[31] Allan Luke. 2014. Defining critical literacy. In *Moving critical literacies forward: A new look at praxis across contexts*, Jessica Pandya and JuliAnna Ávila (Eds.). Routledge, New York, NY, USA.

[32] Ramon Mayor Martins and Christiane Gresse Von Wangenheim. 2022. Findings on Teaching Machine Learning in High School: A Ten - Year Systematic Literature Review. *Informatics in Education* (2022). https://doi.org/10.15388/infedu.2023.18

[33] Pekka Mertala, Janne Fagerlund, and Oscar Calderon. 2022. Finnish 5th and 6th grade students' pre-instructional conceptions of artificial intelligence (AI) and their implications for AI literacy education. *Computers and Education: Artificial Intelligence* 3 (2022), 100095. https://doi.org/10.1016/j.caeai.2022.100095

[34] Luis Morales-Navarro and Yasmin B. Kafai. 2023. Conceptualizing Approaches to Critical Computing Education: Inquiry, Design, and Reimagination. In *Past, Present and Future of Computing Education Research : A Global Perspective*, Mikko Apiola, Sonsoles López-Pernas, and Mohammed Saqr (Eds.). Springer International Publishing, Cham, 521–538. https://doi.org/10.1007/978-3-031-25336-2_21

[35] Luis Morales-Navarro, Yasmin B. Kafai, Ha Nguyen, Kayla DesPortes, Ralph Vacca, Camillia Matuk, Megan Silander, Anna Amato, Peter Woods, Francisco Castro, Mia Shaw, Selin Akgun, Christine Greenhow, and Antero Garcia. 2024. Learning about Data, Algorithms, and Algorithmic Justice on TikTok in Personally Meaningful Ways. In *Proceedings of the 18th International Conference of the Learning Sciences (ICLS)*. International Society of the Learning Sciences.

[36] Viktoriya Olari and Ralf Romeike. 2024. Data-related concepts for artificial intelligence education in K–12. *Computers and Education Open* 7 (2024), 100196.

[37] Luci Pangrazio and Neil Selwyn. 2019. 'Personal Data Literacies': A Critical Literacies Approach to Enhancing Understandings of Personal Digital Data. *New Media & Society* 21, 2 (2019), 419–437. https://doi.org/10.1177/1461444818799523

[38] Nicolas Pope, Juho Kahila, Jari Laru, Henriikka Vartiainen, Teemu Roos, and Matti Tedre. 2024. An Educational Tool for Learning about Social Media Tracking, Profiling, and Recommendation. https://doi.org/10.48550/arXiv.2402.01813 arXiv:2402.01813 [cs.CY]

[39] Nicolas Pope, Juho Kahila, Henriikka Vartiainen, and Matti Tedre. 2024. Children's AI Design Platform for Making and Deploying ML-Driven Apps. https://doi.org/10.36227/techrxiv.24320794.v1

[40] Nicolas Pope, Henriikka Vartiainen, Juho Kahila, Jari Laru, and Matti Tedre. 2024. A No-Code AI Education Tool for Learning AI in K-12 by Making Machine Learning-Driven Apps. In *2024 IEEE 24th International Conference on Advanced Learning Technologies (ICALT)*.

[41] Jenny S. Radesky, Heidi M. Weeks, Rosa Ball, Alexandria Schaller, Samantha Yeo, Joke Durnez, Matthew Tamayo-Rios, Mollie Epstein, Heather Kirkorian, Sarah Coyne, and Rachel Barr. 2020. Young Children's Use of Smartphones and Tablets. *Pediatrics* 146, 1 (2020), e20193518. https://doi.org/10.1542/peds.2019-3518

[42] Mitchel Resnick. 2017. *Lifelong Kindergarten: Cultivating Creativity through Projects, Passion, Peers, and Play*. The MIT Press, Cambridge, MA, USA.

[43] Saman Rizvi, Jane Waite, and Sue Sentance. 2023. Artificial Intelligence teaching and learning in K-12 from 2019 to 2022: A systematic literature review. *Computers and Education: Artificial Intelligence* 4 (2023), 100145. https://doi.org/10.1016/j.caeai.2023.100145

[44] Neil Selwyn and Luci Pangrazio. 2018. Doing data differently? Developing personal data tactics and strategies amongst young mobile media users. *Big Data & Society* 5, 1 (2018), 1–12. https://doi.org/10.1177/2053951718765021

[45] R. Benjamin Shapiro and Rebecca Fiebrink. 2019. Introduction to the Special Section: Launching an Agenda for Research on Learning Machine Learning. *ACM Transactions on Computing Education* 19, 4 (2019), 30:1–30:6.

[46] Mariya Stoilova, Sonia Livingstone, and Rishita Nandagiri. 2020. Digital by default: Children's capacity to understand and manage online data and privacy. *Media and Communication* 8, 4 (2020), 197–207.

[47] Liping Sun, Marjaana Kangas, Heli Ruokamo, and Signe Siklander. 2023. A systematic literature review of teacher scaffolding in game-based learning in primary education. *Educational Research Review* 40 (2023), 100546. https://doi.org/10.1016/j.edurev.2023.100546

[48] Iris Tabak. 2004. Synergy: A Complement to Emerging Patterns of Distributed Scaffolding. *Journal of the Learning Sciences* 13, 3 (2004), 305–335. https://doi.org/10.1207/s15327809jls1303{_}3

[49] Matti Tedre, Henriikka Vartiainen, Juho Kahila, Tapani Toivonen, Ilkka Jormanainen, and Teemu Valtonen. 2020. Machine Learning Introduces New Perspectives to Data Agency in K–12 Computing Education. In *2020 IEEE Frontiers in Education Conference (FIE)*. 1–8. https://doi.org/10.1109/FIE44824.2020.9274138

[50] TENK. 2023. *The Finnish Code of Conduct for Research Integrity and Procedures for Handling Alleged Violations of Research Integrity in Finland.* Number 4/2023 in Publications of the Finnish National Board on Research Integrity. TENK. https://tenk.fi/sites/default/files/2023-05/RI_Guidelines_2023.pdf

[51] Teemu Valtonen, Matti Tedre, Kati Mäkitalo, and Henriikka Vartiainen. 2019. Media Literacy Education in the Age of Machine Learning. *Journal of Media Literacy Education* 11, 1 (2019), 20–36.

[52] Janneke van de Pol, Monique Volman, and Jos Beishuizen. 2010. Scaffolding in Teacher–Student Interaction: A Decade of Research. *Educational Psychology Review* 22, 3 (2010), 271–296. https://doi.org/10.1007/s10648-010-9127-6

[53] Simone van der Hof. 2017. I agree, or do I? A rights-based analysis of the law on children's consent in the digital world. *Wisconsin International Law Journal* 34, 2 (2017), 409–445.

[54] Maarten Van Mechelen, Rachel Charlotte Smith, Marie-Monique Schaper, Mariana Tamashiro, Karl-Emil Bilstrup, Mille Lunding, Marianne Graves Petersen, and Ole Sejer Iversen. 2023. Emerging Technologies in K–12 Education: A Future HCI Research Agenda. *ACM Transactions on Computer-Human Interaction* 30, 3, Article 47 (2023). https://doi.org/10.1145/3569897

[55] Henriikka Vartiainen, Juho Kahila, Matti Tedre, Sonsoles López-Pernas, and Nicolas Pope. 2024. Enhancing children's understanding of algorithmic biases in and with text-to-image generative AI. *New Media & Society* (2024), 14614448241252820. https://doi.org/10.1177/14614448241252820

[56] Henriikka Vartiainen, Lotta Pellas, Juho Kahila, Teemu Valtonen, and Matti Tedre. 2022. Pre-Service Teachers' Insights on Data Agency. *New Media & Society* (2022), 1–20. https://doi.org/10.1177/14614448221079626

[57] Henriikka Vartiainen, Tapani Toivonen, Ilkka Jormanainen, Juho Kahila, Matti Tedre, and Teemu Valtonen. 2020. Machine learning for middle schoolers: Children as designers of Machine Learning Apps. In *Proceedings of the 2020 FIE Conference*. Uppsala, Sweden.

[58] Vivian Maria Vasquez, Hilary Janks, and Barbara Comber. 2019. Critical Literacy as a Way of Being and Doing. *Language Arts* 96, 5 (2019), 300–311.

[59] Christiane Gresse von Wangenheim, Jean C.R. Hauck, Fernando S. Pacheco, and Matheus F. Bertonceli Bueno. 2021. Visual tools for teaching machine learning in K-12: A ten-year systematic mapping. *Education and Information Technologies* 26 (9 2021), 5733–5778. Issue 5. https://doi.org/10.1007/s10639-021-10570-8

[60] World Economic Forum. 2011. *Personal Data: The Emergence of a New Asset Class.* World Economic Forum, Cologny, Switzerland. https://www.weforum.org/publications/personal-data-emergence-new-asset-class/

[61] Shoshana Zuboff. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* Profile Books, New York, NY, USA.