

Understanding AI Mechanisms Supports Disciplinary Reasoning and Ethical Judgment in K–12 AI Literacy Education

Henriikka VARTIAINEN* [0000-0001-6005-907X], Juho Kahila [0000-0002-9913-0627],
Nicolas Pope [0000-0002-5694-4970], Matti Tedre [0009-0001-1956-4055]

University of Eastern Finland, Finland

e-mail: henriikka.vartiainen@uef.fi, juho.kahila@uef.fi, nicolas.pope@uef.fi, matti.tedre@uef.fi.

Abstract. While the urgency of AI literacy has been broadly acknowledged, little empirical research has examined how AI education supports the internalization of AI concepts. Whereas much AI education research reports short-term learning gains, this study investigates how students use AI concepts to mediate collaborative reflection and ethical reasoning after participation in a three-year co-design based cumulative AI education program. Drawing on sociocultural and cultural-historical activity theory, we qualitatively analyzed group interviews with 40 K–12 students in grades 6 and 9. The results show that students fluently used AI concepts as mediating tools. Instead of commonsense explanations and folk theories, students relied on scientific explanations and used AI concepts to explain classifier behavior, social media mechanisms, user engagement, and filter bubbles. These concepts and theoretical orientation to everyday data-driven systems extended to ethical reasoning and critical analysis of practices, structures, and problems in everyday life both at personal and societal levels. The study highlights the potential of cumulative AI education to foster transformative learning and agency, extending AI literacy beyond functional understanding toward volition and abilities to question existing practices, envision alternatives, and take informed action.

Keywords: AI literacy; AI education; K–12; Data agency; Transformative agency; Sociocultural theory; AI ethics; Computing education.

1. Introduction

Despite the infusion of algorithm- and data-driven technologies in our everyday lives, many people are unaware of how AI systems mediate and engineer our decision-making processes, knowledge creative activities, and the social practices we are involved in (Coeckelbergh, 2023; Pangrazio & Selwyn, 2023). Studies have shown that in their everyday interactions with AI tools and social media platforms, people develop various kinds of “folk theories” to explain how these systems and mechanisms work, which typically differ

* Corresponding author

significantly from scientific, disciplinary explanations of AI (DeVito et al., 2018; Büchi et al., 2023). If people are unaware of the limitations and mechanisms of AI, they may act against their best interest (Coeckelbergh, 2023) and uncritically adopt and adapt to socio-technical practices that exacerbate various kinds of ethical challenges and democratic concerns, such as total surveillance, hybrid influencing, behavior engineering, algorithmic biases, and social inequities, as well as diminishing trust in media, public authorities, and science (Bommasani et al., 2021; Solaiman, 2023; Zuboff, 2015).

Recent studies also indicate that the formation of more scientific conceptions of AI and related mechanisms is unlikely to happen through everyday experiences with AI among children and young people, even though they are active users of various kinds of social media services and AI-driven tools (Kahila et al., 2026; Mertala et al., 2022). In parallel, research has also revealed substantial disparities in the support children and youth receive from their parents and teachers concerning the data economy and digital infrastructure that run people's everyday interactions (Stoilova et al., 2020). These disparities exacerbate existing social inequalities and digital divides by making some children more vulnerable to manipulation, exploitation, and exclusion (Stoilova et al., 2020).

In the advent of AI-driven transformations and equity concerns, the urgency of AI education initiatives has been broadly acknowledged by, for example, UNESCO (Miao & Shiohira, 2024), EU AI Act (Article 4), and OECD (2025). At this point, approaches in AI education and educational research have typically focused on the use of AI tools in learning and teaching ("AI in education" or "learning with AI") or, more recently, proposed general frameworks, content listings and early findings from studies aimed at learning about AI (understanding how AI works) (Rizvi et al., 2023; Grover, 2024; Miao & Shiohira, 2024; OECD, 2025; Gu & Ericson, 2025). While these conceptual inventories and early initiatives of AI education can provide valuable insights, many of them remain untheoretical (Vartiainen & Tedre, 2024), are short-term and largely exploratory (Weidlich et al., 2025), and follow traditional interventionist research that ignores the context-bound nature of human learning and agency (Engeström, 2011). Such de-contextualization also means that teachers and learners are left to adapt to top-down approaches that ignore their voices and everyday realities (Bal et al., 2019), making it unlikely to achieve a transformative and sustainable educational impact in the long run (Engeström, 2011).

This study is part of longitudinal research and development work that engaged researchers from different disciplines and schoolteachers from 12 schools to co-design and implement a cumulative AI learning path for roughly 200 students. At the beginning of the three-year AI education program, participating students were in grades 4 and 7. Unlike approaches that see AI education as adaptation to existing beliefs, practices and technological transformations, this study's aim was to systematically support transformative learning and agency—understood as the volition and skills to question, shape, make and transform AI and data-driven practices. In particular, the aim was to cultivate children's conceptual AI understanding and data-driven thinking so that they could become mental tools for critically analyzing how AI and data-driven systems work, how they shape our everyday life and decision-making, and what kind of impacts and ethical concerns are involved.

Furthermore, the approach of this study differs from reproductive top-down interventions and traditional interventionist research, where researchers define how they want to change educational practices and then evaluate whether or not the desired outcomes are actually achieved (Engeström, 2011). Instead of providing teachers with a check-list of AI concepts to cover, our approach was based on contextualization, participatory design, and sustained collaboration with those affected by educational change efforts. It also differed from many existing initiatives by systematically building cumulative learning paths rather than one-off experiments covering disconnected AI concepts (Vartiainen & Tedre, 2025a, 2025b).

The previous studies related to this research program have focused on the theoretical and pedagogical foundations of AI education (Vartiainen & Tedre, 2025a, 2025b) as well as the immediate effects of the interventions (Vartiainen et al., 2025a; Vartiainen et al., 2025b). Pre-post assessments have also demonstrated significant and sustained improvements in individual students' conceptual understanding (Vartiainen et al., 2025a; Vartiainen et al., 2025b). The current study extends this work both theoretically and empirically. Theoretically, it contributes to understanding how the internalization of AI concepts as mediating tools can support transformative learning and shape how people perceive and engage with the data-driven realities of everyday life. Empirically, it shifts the focus from measuring individual learning outcomes to collaborative meaning-making and post-program group interviews. In order to do that, we organized 14 group interviews to examine how grade 6 and 9 students (N=40), after completing the three-year learning path, use AI concepts as mediating tools for thinking. While a recent conference paper drawing on the same group-interview dataset examined specifically how transformative agency in relation to social media was manifested in three theoretically selected, information-rich episodes (Vartiainen et al., 2026), the present study has a different analytical focus on conceptual understanding and ethical reflection, examining more broadly how students use AI concepts as mediating tools in situated, collaborative meaning-making across classifier reasoning, algorithmic bias, social-media mechanisms, and in reasoning about their ethical and societal impacts. In particular, this study sought to answer the following research question:

How do students use AI concepts in collaborative reflection and ethical reasoning after an AI education program?

We begin by grounding our approach to cultural-historical theories of learning and cognition, extended with the perspectives of critical pedagogy and transformative agency. Then, we elaborate how these theoretical ideas were transformed into a pedagogical model for guiding classroom practices that engaged 12 classes of students to explore AI concepts, mechanisms, and impacts through progressive inquiries. After presenting the context of the study, data collection, and analysis methods, we present theory-driven analysis of 14 group interviews organized at the end of the cumulative AI learning path. We conclude with a discussion of the contributions and limitations of the study, and present directions for future research.

2. Theoretical foundations

The theoretical foundations for this research and development work draw on classical sociocultural and cultural-historical theories of learning and cognition, rooted in the ideas of mediated action by Vygotsky (1978) and his followers. The point of departure has been that human interaction with the world occurs indirectly and is always mediated by social interactions and various physical or psychological tools that we use for creating knowledge and understanding about the world (Wertsch, 1993). This insertion of tools and cultural artifacts into human learning was revolutionary as individual actions could no longer be understood without the cultural means that do not simply mediate or facilitate human functioning, but influence and alter the structure of mental functions of the human agent (Edwards, 2007).

Vygotsky (2012) considered natural language as the primary psychological tool in human development. He viewed formal schooling as the key context for the development of scientific concepts, which he distinguished from spontaneous concepts that emerge through generalization of concrete everyday experiences (Vygotsky, 2012). Spontaneous concepts are often unsystematic and may include misconceptions, whereas scientific concepts are systematic, culturally formulated, and hierarchical knowledge, extending far beyond immediate observations or concrete experiences (Karpov, 2003; Vygotsky, 2012). Consequently, the development of scientific thinking requires systematic instruction. The extent to which individuals develop such higher mental functions depends on the opportunities to participate in activities in which this mode of thinking is scaffolded (Wells, 1994).

For scaffolding, Vygotsky (1978) introduces the notion of the *zone of proximal development* (ZPD), defined as “the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers” (p. 84). In other words, children are not left alone to develop scientific thinking but supported by a teacher or some other more experienced person to form meanings, solve problems, or complete tasks that he or she does not yet have the developmental capacity to do independently (Vygotsky, 1978). From this perspective, learning scientific concepts involves gradual internalization of culturally created knowledge and meanings into individuals’ own mental tools and expressive means (Vygotsky, 2012).

Internalization is best supported when scientific concepts are systematically related to students’ everyday experiences and explored through inquiry-based and knowledge creative activities (Chaiklin & Hedegaard, 2013). By that way, students are supported to symbolize the knowledge they construct through their own inquiries leading to increasingly explicit linguistic formulations of concepts and their relations (Chaiklin & Hedegaard, 2013). In other words, internalization of scientific concepts is expressed through language that is an important mediator of this development. Once scientific concepts are acquired, they begin to mediate thinking and problem-solving, enabling children to reason more independently of immediate observations (Karpov, 2003). Increasing complexity in the relationship between everyday experiences and scientific concepts can be observed when children orient

themselves more theoretically toward the surrounding world and use scientific concepts and ideas to critically analyze everyday experiences and reflect on problems and issues relevant to personal or societal life (Chaiklin & Hedegaard, 2013; Hedegaard, 2005).

While sociocultural theories have been criticized for emphasizing cultural transmission and socialization into existing practices (Wells, 1994), later expansions and developments in activity theory (Engeström & Sannino, 2010; Stetsenko, 2019) and in critical pedagogy (Freire, 1974) have offered important insights on the transformative nature of learning. Learning and comprehension are prerequisites for “reading the world” and for understanding how social, political, and economic structures shape our everyday experiences and produce oppressive practices (Freire, 1974). From this perspective, learning scientific knowledge is not limited to reproducing existing ways of thinking, but provides mental tools for perceiving the world differently and for recognizing underlying structures, practices, and problems that may otherwise remain invisible or taken for granted. Ability to analyze and question the status quo also provides the foundations of transformative agency needed for resisting oppressive practices, envisioning alternatives, and taking informed action (Stetsenko, 2019).

3. Pedagogical framework

For transforming the theoretical ideas into real-life practices, we have developed a pedagogical model for contextualized AI education. The CEDE model (Vartiainen & Tedre, 2025a) is based on knowledge creative learning (Hakkarainen et al., 2013) and inquiry- and design-based pedagogies (e.g. Bereiter & Scardamalia, 2006; Hakkarainen et al., 2013; Hakkarainen & Seitamaa-Hakkarainen, 2022) that, in line with sociocultural theories, emphasize the key role social interaction plays in a jointly constructed activity, and the role of tools and mediating artifacts in the evolving process of learning and inquiry. By engaging students in critically exploring, resisting, and transforming data driven practices of everyday life, our model is also grounded in critical pedagogy (Freire, 1974) and contemporary scholarship around transformative agency and expansive learning (Engeström et al., 2023; Stetsenko, 2019). While these theoretical and pedagogical ideas coming from the learning sciences provide solid groundings for developing pedagogical models for AI education, they have not been extensively explored in that context.

The CEDE model aims at advancing three interconnected dimensions of transformative data agency: comprehension (AI concepts and mechanism), critical thinking (impacts and ethics of AI) and data-driven design (envisioning and creating new AI-driven artefacts and solutions) (Vartiainen & Tedre, 2025a). Rather than assimilation to existing data-driven practices, it aims to support children to understand how AI systems work, cultivate their critical awareness needed to question and challenge the status quo along with data-driven design skills needed for taking practical actions by means of co-designing, and creating one’s own artefacts and inventions (Kahila et al., 2024). While learning projects following the model may vary in emphasis (e.g. focus more on data-driven design, the exploration of AI-related phenomena, or ethical questions), the model is structured around four main phases that can guide the activities in a variety of contexts (Vartiainen & Tedre, 2025a).

3.1 *Contextualization.*

Contextualizing learning activities and new concepts in students' lives has long been an important part of science instruction (Rivet & Krajcik, 2008), and also plays a central role within the CEDE model as well. Contextualization is promoted through real-life problems (Krajcik & Blumenfeld, 2006; Rivet & Krajcik, 2008) that are connected to core AI concepts and interrelationships between them. The problem or design challenge should be open-ended, personally meaningful, and ethically sound (Edelson & Reiser, 2006; Krajcik & Blumenfeld, 2006), as well as challenge students to work at the edge of their competence (Vygotsky, 1978), thereby creating a genuine need for collaboration. Such "need-to-know" situation (Rivet & Krajcik, 2008) is fostered through dialogical classroom discussions where AI concepts and mechanisms are linked to students' everyday experiences, such as digital technologies students already use, the socio-technical practices in which they participate, or AI-related societal issues they have encountered (e.g., automated decision-making). Making connections to students' personal lives and embedding the process of learning in a meaningful context also helps them to see the value of new concepts and why they are relevant to their life and future (Krajcik & Blumenfeld, 2006).

Contextualization and collective sense-making also serve as an instructional strategy for activating and externalizing students' prior knowledge, preconceptions, and beliefs, so that teachers as well as students themselves can notice the inconsistencies or insufficiencies in their current understandings (Aleknavičiūtė et al., 2023; Rivet & Krajcik, 2008). While prior knowledge developed in everyday observations and experiences can be incomplete, incorrect and resistant to change (Aleknavičiūtė et al., 2023), it also plays an important role in connecting scientific ideas in meaningful ways, so that new concepts do not remain isolated elements but become integrated with existing knowledge structures and may form meaningful units in long-term memory (Rivet & Krajcik, 2008). Diagnosis of the student's current level of understanding also serves as the basis for creating ZPD tailored to meet students' actual learning needs (Puntambekar & Hubscher, 2002).

3.2 *Exploration.*

In the second phase, students explore the problem space and the available tools and epistemic resources so that they can begin to formulate their own inquiry questions or design ideas. By creating their own questions and making plans for solving them, the students can further establish meaningful connections between their interests and experiences, inquiry activities, and scientific concepts under investigation. Yet, formulating worthwhile inquiry questions or design ideas can be demanding for novice learners due to the underdeveloped knowledge structures and the lack of experience (Rivet & Krajcik, 2008), and thus, their formulation is supported by the teachers, peers, tools, and curriculum materials including prompting questions.

3.3 Design and inquiry.

In the third phase, students are supported to participate in data-driven activities that aim to mirror those of experts when exploring complex phenomena (inquiry) or creating new artefacts (design). This phase is often tool- and technology-enhanced, so that students can work with their own data and then explain their observations using scientific concepts and ideas (Vartiainen & Tedre, 2025a). Yet, research has also shown that novice learners may have difficulties in engaging in inquiry activities, especially if they lack previous experience (Krajcik & Blumenfeld, 2006). Lack of previous knowledge may also focus their observations to surface features of the concepts under investigation (Chi et al., 1981). Thus, the collaborative learning activities are supported by educational technologies and worksheets that guide the children to follow the basic epistemic functions related to data-driven design and inquiry. They also use tools and technologies to create artefacts that address their own inquiry questions or design tasks and externalize the knowledge they have constructed (Krajcik & Blumenfeld, 2006). These can be, for example, inquiry reports, presentations, or physical or digital prototypes, such as AI-driven robots or apps.

3.4 Ethical and societal reflection.

In the fourth phase, students are guided to share their solutions and reflect on their learning and knowledge creation processes. Through their own projects, students are guided to reflect epistemological issues related to data-driven knowledge creation, including questions of trust and accountability. In addition, they are supported to apply data-driven reasoning to collaborative discussions and critical analyses of contextualized ethical and societal issues. By that way, they can also use their evolving understanding to reconsider existing data practices and power structures that are internalized into our consciousness through everyday interactions and through mass culture (see Shor, 1993). However, the students are not only made aware of existing practices and circumstances but are supported to critically question the status quo and take action, for example, by designing influence campaigns, position statements, political campaigns, or writing letters to decision-makers.

Altogether, CEDE provides a processual model for connecting conceptual, critical, and ethical perspectives to data-driven design or inquiry. Yet, internalization of scientific concepts and building increasingly higher-order conceptual networks is a complex and gradual process that takes time and effort (Vosniadou, 2013). Thus, there is also an evident need for building cumulative learning trajectories where students can first learn the basics before moving to more and more complex AI concepts and ethical considerations. The following sections describe how such a cumulative CEDE learning path was co-designed and actualized in a large-scale school project in Finland.

4. Methods

The three-year (2023-2025) process of co-designing cumulative AI learning paths involved researchers from computer science and education, teachers and school principals from 12 primary and secondary schools, as well as over 200 students (4th and 7th graders at the beginning of the project). Our approach for co-developing AI education was built on participatory design that shares similarities with the change laboratory methods (Sannino & Engeström, 2017). This included strategically and systematically involving teachers in co-creating new educational practices through sustained, iterative processes of analyzing and defining educational needs, co-designing solutions, implementing and testing them in practical school projects along with reflecting and evaluating the process and outcomes of it (Vartiainen & Tedre, 2025b).

In practice, our collaboration began with an analysis of the status quo, including our baseline studies and literature reviews that were shared and discussed with the participating teachers to stimulate ideas, involvement, and discussions. Breaking away from the status quo was actualized by collaboratively visioning what could be, visioning and negotiating what ought to be, and taking collective action for realizing that future (Haapasaari et al., 2016; Hopwood & Sannino, 2023). This unfolded in an iterative, three-year co-design process in which each round of interventions built on the insights, observations, and preliminary research findings of the previous one, altogether creating a pedagogical model and a set of mediational tools for the cumulative AI learning path. By that way, people from different backgrounds and with a variety of expertise engaged in collective action and future making for common good.

4.1 *Instructional interventions*

The first-year school intervention focused on data-driven design, particularly on concepts such as data, classifier, confidence, brittleness, and bias. These basic AI concepts were introduced through a hands-on co-design process in which students were making their own apps using their own data and the GenAI Teachable Machine (Pope et al., 2023). Rather than encountering new scientific concepts as abstract and decontextualized definitions, the students collaboratively explored them through the construction of externalized artefacts that were shared with others (for more information, see Pope, Vartiainen, et al., 2024). At the end of the learning project, the students were guided to explore the ethics of AI, especially the concept of algorithmic bias, through the use of generative AI. These first school projects included three workshops, each lasting approximately 1.5 to 2 hours.



Figure 1. Children co-designing and making their own apps with the support of the GenAI Teachable Machine (first-year intervention)

Building on the basic AI concepts and mechanisms introduced in the first intervention, the second year extended learning activities to social media mechanisms. Particularly, the conceptual focus was on data collection, profiling, and recommending that students explored with the support of profiling game (for more information, see Kahila, Vartiainen, Arkko, et al., 2024) and Someone social media simulator that affords students to use their own data to collaboratively explore the underlying mechanisms of social media in a real-time visual environment (for more information, see Pope et al., 2025; Pope et al., 2024). At the end of this project, students wrote letters to decision makers in which they reflected both the opportunities and harms of data collection, profiling, and recommending, and proposed ways in which social media mechanisms could better address the needs of children and young people as users. These second year school projects included two workshops, each lasting approximately 1.5 to 2 hours.

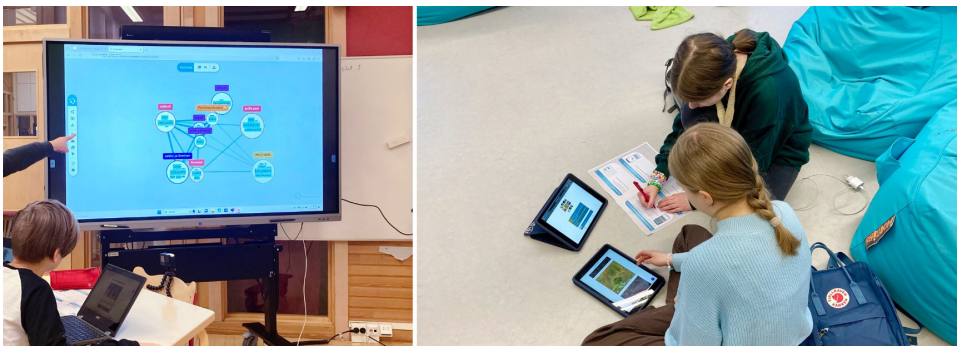


Figure 2. Children exploring the underlying mechanisms of social media with the support of Someone (second-year intervention)

In the third year of interventions, the focus was on ethics and the impacts of social media mechanisms, with particular attention to filter bubbles, manipulation, and intended and unintended consequences. These collaborative activities were supported by Someone, especially its heat map feature and customizable recommendation algorithm, which afforded students to explore and compare filter bubbles. At the end of the workshop, the

students were guided to reflect on how social media influences their own decision-making, behavior, and opinions, and to consider which aspects of their lives they do not want social media to influence. This intervention included one workshop (approximately 2 hours).

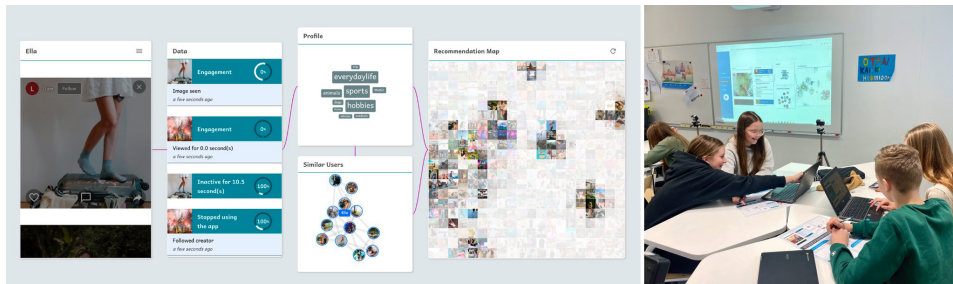


Figure 3. Left: Overview of the Someone workflow: Tracking users' data traces, engagement metrics, profiling, clustering similar users, and visualizing each user's space of possible recommendations (filter bubble). Right: Students using Someone in the classroom.

4.2 Participants and data collection

This study collected data from reflective interviews organized in spring 2025 right after the third year interventions. They involved students from 8 different schools and consisted of 14 group interviews (7 conducted in secondary schools, 7 in primary schools), with a total of 40 participants. Each interview group included two to four students, who were selected by their classroom teachers following Creswell's (2006, p. 127) principle of intensity sampling (information-rich informants who exemplify the phenomenon intensely but not extremely). Teachers were asked to identify students who had participated in the AI learning activities and were willing and able to reflect on their experiences.

The interviews were semi-structured and organized around thematic areas corresponding to the different phases of the cumulative AI learning path. The same interview protocol and semi-structured questions were used for both 6th- and 9th-grade students. They began with warm-up questions that invited students to recall tasks or activities they found particularly memorable or meaningful (e.g. *You have now participated in our school visits as part of this project for three years. Is there any task, activity, project, or lesson that has particularly stayed in your mind or that you found especially memorable? Why?*).

Based on the first-year intervention, which focused on data-driven design, students were shown a diagram of the GenAI Teachable Machine and asked to explain how they would teach fourth-grade students to build a classifier (*Let us imagine that your task is now to teach a fourth-grade student how to create an AI application. The student wants to create an application that can distinguish between a bottle and a can. Could you explain, step by step, what is involved in building a classifier? In other words, what is done in the six main stages, and why? What would be important to pay attention to in each stage?*). When needed, follow-up questions were asked to elicit explanations of core concepts and

their relationships, such as reasons for misclassification. As algorithmic biases were also explored during the first year activities, students were subsequently shown examples of images created by children using generative AI. They were then asked to explain what algorithmic bias is, what kind of impacts it can have, and how such biases might be mitigated (*Can you explain what algorithmic bias means? What causes it? What kinds of impacts can algorithmic biases have? How should they be corrected?*)

Based on the second-year intervention, which focused on social media mechanisms, students were asked what kinds of data are collected about them, what social media companies can infer from that data, for what purposes data are used by these companies, and why (e.g. *What kinds of data do social media platforms collect about you when you use them? What can social media companies infer about you from the data they collect? For what purposes do social media platforms use the data they collect and the information they infer about you?*). They were also asked to reflect on the benefits and risks of data collection, profiling, and recommending (*What benefits or positive aspects do you see in data collection, profiling, and recommendation on social media? What harms or risks do you see in social media platforms collecting and using your data?*).

Based on the third-year intervention, which focused on impacts and ethics of social media, students were asked to reflect how social media and especially recommendation systems influence their everyday life, their thinking, or opinions (e.g. *Have you noticed whether these social media mechanisms affect your everyday life in any way? [How?] How does the content you see on social media influence the way you think about your own life, yourself, or other people? In what ways do you think social media mechanisms can be used to deliberately influence you [e.g. opinions, decisions]?*). They were also asked to consider the implications of different users being exposed to different content (e.g. *What do you think are the impacts of different users seeing different content in their social media feeds?*).

Finally, the students were invited to propose improvements and vision alternatives. They were asked how social media could better support young users, what advice they would give to other children who are less aware of the data practices, and what actions different stakeholders (e.g. parents, schools, policymakers, tech companies, researchers, and students themselves) could take to make social media safer and more supportive of young people's well-being (reported elsewhere). These interviews resulted in a total of 8 h 39 min of audio data (average length of an interview was 37 minutes), transcribed verbatim. Real names are not reported.

Altogether, the interview protocol was structured around the key concepts, mechanisms, and ethical themes introduced during the three-year learning path. This enabled examination of how concepts originally encountered through classroom activities were subsequently used by students in collaborative reflection and reasoning, thereby linking the intervention design to the analysis of the internalization of AI concepts as mediating tools for thinking. Following a sociocultural and discursive perspective, students' reasoning was not approached as a decontextualized cognitive state, but as situated, mediated activity emerging through interaction with cultural tools and other participants (Schoultz et al.,

2001). Accordingly, the visual cues, interviewer prompts, and peer interaction are understood as specific linguistic and physical resources that served as points of departure for the reflective discussions and kept the children within a particular discursive framework. Internalization is thus interpreted here through students' situated use of scientific concepts as mediating tools in collaborative reasoning within these context-bound interactions following the three-year learning path.

4.3 *Data analysis*

To analyze how students used AI concepts in collaborative reflections, a theory-driven qualitative content analysis was conducted (Elo & Kyngäs, 2008), with the help of the qualitative analysis software Atlas.ti. In the first phase, all interview transcripts were read multiple times to gain familiarity with the data and to develop an overall understanding of the nature of students' collaborative discourses.

In the second phase, meaning units relevant to research questions were identified. Informed by sociocultural theory, these meaning units consisted of episodes in which students (subjects) used AI concepts (mediating tools) to explain specific objects, such as classifier behavior, social media mechanisms, or filter bubbles. Once identified, these units were examined in more detail to analyze the contextual connections within the subject-tool-object relationship (Vygotsky, 1978), like how students connected AI concepts and the object under discussion to their everyday lives, ethical considerations, or to experiences from the AI learning activities. Each meaning unit was given a descriptive name that captured its core content. In this way, the coding scheme was developed iteratively by expanding theory-driven analysis with data-driven insights. Sociocultural theory and the AI concepts embedded both in the design of the intervention and interview protocol provided initial analytical lenses, while repeated engagement with the interview data informed the contextualization, refinement, and further development of the coding scheme.

In the third phase, the coded meaning units were compared and contrasted to identify similarities, differences and relations between them. Through this process of abstraction, related codes were grouped together to form higher-order categories representing contextualized subject-tool-object relationships. In the final phase, these categories were reviewed and refined collaboratively among the researchers, resulting in analytical categories that are elaborated in the result section and summarized in discussion. To enhance the trustworthiness of the analysis, the emerging coding scheme and interpretations were discussed among the researchers throughout the process, and a detailed description of the interview context, along with data excerpts, is provided to allow readers to consider and evaluate these interpretations (Hammer & Berland, 2014).

5. Results

5.1 Data-driven design and probabilistic reasoning in the ML training pipeline

The first-year intervention focused on data-driven design, particularly on concepts such as data, classifier, confidence, brittleness, and bias. When presented with a diagram of the GenAI Teachable Machine as a visual prompt and asked to explain how they would guide a fourth grader in creating a classifier, all groups were able to recall the core workflow of machine-learning classification and demonstrated reasoning grounded in data-driven thinking. The following example from 9th graders illustrates how students appropriated scientific concepts related to classifiers as part of their situated explanations, two years after the original intervention:

Researcher: Could you now explain what should be done in the main steps so that fourth graders would know how to make a working AI application?

Hannes: In the first step, it classifies these things. There could be a bottle, and then you feed it pictures of bottles to it, so it can recognize the bottle. And then in the second step, there's a can, with pictures of cans, so the AI recognizes the can. And then it can tell them apart. And maybe also an empty one, so if there's nothing there, then...

Kaapo: And then in the second step, the one who teaches, so what pictures? Wasn't it like that? You take the data and the explanations there.

Hannes: Then in the third step, if you show it a new picture that it should recognize, it shows probabilities. And then which one it is. What it looks like to it.

Kaapo: And then in the fourth step, it shows those, if it recognizes, like here you see, scissors. Then it performs the action that was selected and sends it forward.

Hannes: And then at the end, there's just the answer.

Researcher: Well, what if things don't go so well in step three, if it doesn't recognize it? How would you advise what to do?

Hannes: Then you probably need to give it more information about what they look like. Like more data for the AI.

Researcher: At which stage would that happen?

Hannes: Adding more from the beginning of the process. Add more pictures there in step one. And then teach it again in step two.

To begin with, the students demonstrated an understanding that classifiers depend on training data (images) and on labeled classes such as “bottle,” “can,” and neither of them (“empty”). Importantly, their explanations did not describe classification as a rule-based, deterministic process. Instead, they described how a trained model generalizes from its training data to recognize new images using a probabilistic process. By explicitly referring to probabilistic reasoning, using their own words, the students showed they understood

a key epistemic feature of machine learning systems that distinguishes ML from classical programming: Data-driven AI systems do not follow a predetermined set of rules to produce deterministic outputs, but estimate the probability that an input belongs to certain categories based on regularities in the training data (Tedre et al., 2021).

Although they did not explicitly use the concepts of “confidence” or “brittleness,” they continued by reasoning that failures in recognition stem from the training data. This reasoning indicates how students were not only able to describe the basic ML training pipeline and articulate what the system is expected to do at each stage, but described appropriate troubleshooting strategies, such as returning to earlier steps to populate the classes with more data and retrain the model.

In similar vein, 6th graders engaged in data-driven reasoning and explanations that integrated conceptual and procedural understanding of how to build and refine a classifier:

Miska: First, you need to take lots of different bottles and take pictures of them, and the same for cans

Anna: From different positions and angles so it recognizes

Miska: Then when you click “teach the classifier,” then that thing pops up, whatever it’s called, and then...

Researcher: Do you mean the input? The texts help there, right?

Miska: Yeah, so when you put the can or the bottle there, then it shows whether it’s a can or a bottle.

Inka: There’s the percentage that shows how sure it is.

Researcher: Okay

Miska: But if it’s a different-colored can, like if you take one that’s a different color than the ones in the pictures, then it might not recognize it

Researcher: How could you fix that if it doesn’t recognize it?

Miska: Then you could take more pictures

Researcher: And what are those pictures, by the way?

Miska: Aren’t they the data?

The exchange begins with Miska identifying the first step of the process by explaining the need to collect different bottles and cans and take pictures of them, demonstrating an awareness of the importance of curating a varied and representative training dataset. Anna extends this reasoning by noting that the training dataset should contain images captured from different positions and angles, further highlighting an understanding of dataset diversity and quality. The students also display procedural understanding of the ML training pipeline as they continue to describe the training phase and how the classifier can subsequently be tested. They explicitly mention probabilistic confidence values, and they further show an awareness of brittleness by reasoning that the system may fail to recognize a dif-

ferently colored can if its kind was not present in the original training data. When asked how such a problem could be addressed, Miska proposes a solution familiar from the ML training pipeline, namely adding more images into the training data, and he then connects this action to the conceptual understanding that expanding the training dataset is a means of improving model robustness and generalization.

This group, like many others, also perceived this first-year intervention as the most memorable learning experience of the three-year AI education program. As the example below shows, students expressed a shared experience of freedom and autonomy in decision-making, indicating that the development of both conceptual and procedural understanding of basic AI concepts and mechanisms was deeply connected to feelings of agency and ownership:

Researcher: Was there a reason why it [the first year learning project] was especially fun compared to the others?

Miska: It wasn't as restricted as all the others

Anna: You could do it more freely

Inka: Yeah, and you could decide what to make it about.

Researcher: So you could decide for yourself. Was that an important thing?

Inka: Yeah

Taken together, the reflections in all interview groups suggest that students had internalized core machine learning concepts as mediating tools for thinking for situated reasoning. Within these context-bound interactions, they were not relying on intuitive explanations or folk theories, but were able to apply their conceptual understanding to explain and exemplify classifier behavior using discipline-specific language. Notably, they were able to retrieve and apply this integrated understanding even years after the original educational intervention.

5.2 Data traces, profiling, and reasoning about recommender systems

In the second year, the interventions focused on social media mechanisms, particularly on data collection, profiling, and recommending. In the interviews, many of the groups were able to name various types of data people provide explicitly, data traces that social media platforms collect from their users, data inferred from those, and how all those data are used for recommendations (van der Hof, 2017; Pangrazio & Selwyn, 2019). For example, a group of 6th graders explained:

Researcher: So... What's your understanding of what kinds of data are collected about you on social media?

Topi: All kinds

Researcher: Give as many examples as you can

Aapo: Well, who we search for

Topi: Yeah, what we watch

Topi: What we like

Researcher: Anything else?

Simo: Not really

Researcher: That's quite a lot. So, what can they figure out about you based on that data? What can social media companies infer from you?

Aapo: Age

Researcher: Based on the data

Aapo: Well... what's interesting or what you're interested in

Researcher: Mhm. Anything else?

Topi: Opinions

Researcher: Anything more?

Viljo: About how old you are

Topi: And where you live

Researcher: Yes. Anything else? That's already quite a lot. Why do they collect this data? They don't collect it for nothing. What do they use it for?

Topi: Advertising and recommendations

As the excerpt shows, the students identified a range of behavioral data traces, including what users search, watch, and “like,” indicating awareness of the overt and covert data traces generated through everyday use. When prompted to consider what platforms can infer from such data, the students named demographic attributes (such as age and place of residence), interests, and opinions, indicating that they recognized user profiling as an inferential process based on aggregated behavioral patterns (“observed data” or “data traces”; see World Economic Forum, 2011) rather than explicit self-disclosure (“volunteered data” or “data given”; see World Economic Forum, 2011). By connecting data collection and profiling with advertising and recommendations, the students demonstrated an awareness of the functional role of profiling within recommendation system pipelines, where predicted user attributes (“inferred data”) are operationalized to personalize content and enable behaviorally targeted advertising.

Many of the groups also recognized how social media platforms aim to maximize user engagement—a key mechanism and concept in the context of social media monetization and addiction. One group of 9th graders explained:

Peppi: Probably what content you watch and how much. Or like how long you watch it, and then what you like, follow, or share.

Researcher: Yes. Anything else come to mind?

Peppi: Well, then later it recommends... It recommends what you've paid the most attention to.

Researcher: Yes, quite a lot, right?

Peppi & Amalia: Yeah.

Researcher: So, when these data are collected, what can social media companies infer about you from them?

Amalia & Peppi: Your interests, where you live, hobbies, age, gender, maybe your name? And if you're going somewhere or visit certain places often.

Researcher: That's quite a lot!

Peppi: Yeah

Researcher: So now that they can figure out so much, what do they use this data for?

Jaana: To show content that suits the user

Peppi: Yeah, content that will get reactions later too

Amalia: To get the person to use social media more

In this exchange, Peppi and Amalia demonstrate understanding of the basic recommender system mechanisms by articulating how social media platforms log behavioral interaction data and use it to construct user models to provide personalized content. Peppi begins by specifying the types of engagement signals platforms track (viewing duration, likes, follows, and shares) and explains how recommendation processes prioritize those types of content the user has previously engaged with. When prompted to consider what platforms can infer from such data, the students identify a wide range of profile attributes, including interests, location, hobbies, age, gender, and patterns of movement, indicating an understanding of profiling as a continuous data-driven modeling process. In explaining the purposes of data collection, they refer to tailored recommendations and content as a means to optimize user engagement, recognizing that recommendation systems are designed to maximize continued interaction and prolong platform use, as Amalia reasons.

On the other hand, many student groups did not discuss how recommendations are also influenced by the actions and engagement histories of other users. One exception was group of 9th graders:

Tarja: Well, by what you do on social media, what you like, comment on, share, and follow. What we consume. And then also through other users. Similar users.

Jaana: And maybe also, like, if someone posts a lot of pictures from a certain neighborhood, then the platform can assume they live near there

Researcher: Yes. Now, when they can infer and collect all this information you've mentioned, what do social media companies do with that big amount of data they

have about you?

Tarja: Well, they use it to make social media as addictive as possible

Jaana: Yeah, so it recommends things you either really like or really hate, so you'll stay there longer. To trigger emotions.

Researcher: Anything else?

Tarja: Probably also for advertising, so they can, like I said, provide ads

As the excerpt shows, Tarja explicitly refers to “similar users,” indicating a more advanced conceptualization of recommendation mechanisms that extend beyond individual profiling to patterns across user populations. This suggests understanding of how recommender systems compare and combine user profiles and leverage similarity-based modeling to inform content selection. This understanding is critical because it gives learners a systemic view of recommendation: Individuals’ data do not shape only their own media environments but are aggregated into population-level models that influence what others see, thereby raising ethical questions about collective responsibility, power, and societal consequences of data-driven personalization. When the conversation turns to the purposes of these mechanisms, Tarja further states that the data are used to make social media “as addictive as possible,” while Jaana elaborates that platforms promote content likely to elicit strong affective responses—positive or negative—to increase engagement and sustain attention. Together these explanations show understanding that recommender systems are optimized to maximize engagement and time spent on platform, including emotionally stimulating content, which are also integrated with targeted advertising.

In their reflections, a few groups also mentioned the exploration of social media mechanisms as the most memorable experience in the three-year AI education program, as it expanded their existing understanding. For example, one group of 9th graders reflected:

Researcher: Was anything particularly memorable?

Kaisa: Probably just how social media collects data about us. That's probably what stuck the most

Researcher: Was that completely new? When we first covered social media properly last year, was it a new topic for you?

Kaisa: Well, I kind of knew something about it already, but this really expanded my perspective

Leila: Yeah, maybe it was surprising how much data are actually collected from really small things

Researcher: Right

Leila: And how much that then affects, for example, what gets recommended

Taken together, the reflections on social media mechanisms in the interviewed groups suggest that students applied their conceptual understanding in situated reasoning to ex-

plain and exemplify how platforms collect and infer information about users, personalize content, and shape user engagement. Most groups engaged in disciplinary reasoning about data collection, profiling, recommendation, and engagement optimization, while only one group articulated more advanced explanations involving similarity-based recommendation across user populations. Within these context-bound interactions, the students' use of scientific concepts also exemplifies how internalization is not just a matter of producing a "more advanced explanation" but involves an epistemological shift in how these children and young people understand their relationship with external reality. AI concepts were not retained as decontextualized definitions but were applied to critically analyze the data-driven realities of their everyday lives.

5.3 Reasoning about ethical and societal impacts of data-driven systems

At the end of the first-year intervention, students also explored the concept of algorithmic bias through activities involving text-to-image generative AI (they had experimented with the role of training data in image recognition earlier). When prompted to explain biases after two years, many groups explicitly referenced the role of training data and proposed solutions grounded in modifying or expanding those data. For example, one group of 9th graders explained:

Researcher: What causes it [the bias]? Any thoughts or memories?

Seppo: Maybe it's because those kinds of images are the most common on the Internet.

Researcher: That's right. If that's the case, as you said, how could you fix that algorithmic bias?

Tellervo: Get lots of different kinds of images of that topic online. Not just one type, all the same.

The excerpt above shows how students view algorithmic bias as a data-dependent phenomenon that originates from the training data rather than programming error or design flaw. Their mitigation strategy is to diversify the training data, which indicates that their data-driven reasoning transfers to the domain of ethics, too, making bias an outcome of the training pipeline that one can change, instead of just an abstract moral issue that is out of reach.

Beyond identifying the sources of bias and possible ways to mitigate it, the students also reflected on its broader societal impacts. One group of 9th graders engaged in critical reflection on how biases and stereotypes can be internalized and amplified if one further shares such images online:

Researcher: So could it [algorithmic bias] affect people's thinking in some way?

Kaisa: Well, maybe it distorts people's thinking to be more like the bias. Like, that nurses are all women...even though that's not true

Researcher: Yes, that could happen. Anything else comes to mind? Here's a follow-up question; if that happens, does it also affect your own actions when you see a certain kind of image of a nurse?

Tiina: Well, then you probably start thinking they look a certain way and so on. Or if you post it somewhere, like online, then it starts influencing others, and they also get a distorted picture of that.

Here students described algorithmic bias as a sociotechnical phenomenon, in which biased representations shape perceptions and actions, and, through sharing and interacting, feeds back into data-driven systems propagating and normalizing stereotypes at a collective level.

In the second year, the impacts of social media and data collection were briefly discussed, but in the third year, the interventions focused especially on ethics and the impacts of social media mechanisms, with particular attention to filter bubbles, manipulation, and their intended and unintended consequences. When reflecting on the impacts of social media mechanisms in their personal lives, many groups mentioned the addictive nature of data-driven systems. One group of 6th graders reflected:

Researcher: So, have you noticed how ...recommendation mechanisms, how it...does it affect your everyday life?

Miska: Brain rot

Anna: Well, it does affect you a bit... If you're on social media a lot, you can get so hooked ...that you might accidentally spend the whole day on your phone

Miska: If you want to keep up with trends, you areyou're constantly checking your phone

Inka: Yeah, and sometimes I forget what I was supposed to do, like clean my room, because of the phone. Then I check the time and suddenly I'm on Instagram

Miska: By accident

Inka: Yeah, by accident

Researcher: You just slip in there

Inka: Yeah

Researcher: Why does that happen?

Anna: Because it suggests pictures you like. Like what you've watched the most, videos etc

Inka: Yeah, it's collected so much data that it knows what I like. So then you get hooked... so you want to keep watching more and more

In this excerpt, the students reflect on how personalized recommendation systems shape their daily habits and attention, often leading to unintended overuse. When asked why

these impacts occur, they offer data-driven explanations by referring directly to how platforms tailor content based on past viewing patterns. They further elaborate that the system has collected so much data and that it can harness their preferences to keep them hooked and wanting to consume ever more content. The students explain that these mechanisms produce use patterns that undermine their agency and intentional choice, indicating emerging understanding of engagement optimization and attention engineering strategies as deliberate design choices in platform design. Such reasoning illustrates the students' ability to apply their conceptual understanding of data collection and recommendation to critically reflect and reconsider their everyday interactions with social media.

Negative impacts of social media mechanisms were also considered in terms of self-image and emotional well-being. For example, one group of 6th graders reflected:

Researcher: How do you think the content offered by social media, like what we just discussed, affects what you think about your own life, yourself, your friends, and other people?

Laura: Well, for me it really affected me when at one point there were tons of [fashion] models and all these super-pretty, thin, perfect women, so I immediately noticed I felt much more insecure

Pirkko: You start criticizing yourself

Laura: And then you're like, I don't like these clothes, I'm not good enough, why don't I have this or that

Pirkko: All, some of those people have probably edited all their videos and pictures. Like, if they're showing products, they might not even be real, even if they look exactly like that

Laura: Yeah, and for example in a hobby, if you see someone on IG like, wow, they're so good at that, so I also have to be good, otherwise I'll fall behind, then it creates really awful pressure and insecurity about whether I'm good enough

Here, the students reflect on how algorithmically personalized content feeds influence their self-perception and emotional well-being by systematically feeding them unrealistic representations of "perfect women". Laura's account shows how continued exposure to such content can intensify social comparison, triggering critical self-evaluation and diminished self-confidence. Pirkko extends this reflection by emphasizing the role of manipulated or edited content, noting that many images are filtered or altered, or depict products that may not be genuine, underscoring the constructed and inauthentic nature of online representations. Laura further connects these mechanisms to performance pressures, describing how seeing highly skilled individuals online generates feelings of inadequacy and anxiety about not being good enough. Together, these reflections demonstrate the students' recognition of how personalized content feeds and image manipulation can shape norms of beauty, success, and competence in ways that negatively affect their own confidence, self-worth, and emotional well-being.

The students also recognized the role of influencers in shaping opinions and swaying beliefs. For example, one group of 6th graders reflected:

Paavo: If someone says something bad about another person and says you can't trust them, then I probably wouldn't trust them after that

Maire: But why? If you don't have your own information about them, why would you trust what others say?

Paavo: I don't know

Maire: It depends, if it's your best friend who says it, then...

Paavo: Yeah, but if your best friend talks about someone and says something that might be true, then I might start thinking the same about that person

Maire: Yeah, but on social media too, people who have lots of followers, if they say something, then everyone believes it just because...

Paavo: But if someone with two followers says something, then everyone just wonders what they're doing

Maire: Yeah, like that's not true

In this exchange, the students demonstrate an awareness of how social dynamics and influencers on social media may shape their own opinions and beliefs. Paavo begins by describing a situation in which hearing negative information about someone would cause him to distrust that person, while Maire challenges this by questioning why one would accept such claims without personal knowledge. As the dialogue unfolds, the students connect these interpersonal dynamics to the workings of social media platforms, noting how visibility metrics and follower count function as proxy for credibility. This reasoning signifies an understanding of how common platform-mediated amplification and attention mechanisms privilege certain voices over others, influencing both personal and public trust, opinions, and attitudes.

Opinion swaying in social media was also reflected in terms of filter bubbles and politics, as group of 9th graders reflected:

Researcher: Have you noticed whether these social media mechanisms we've discussed affect your everyday life in any way?

Tarja: Well, yes, going back to the previous task where filter bubbles and things like that were mentioned, how things are seen only from one perspective, while other viewpoints remain hidden. That definitely affects thinking, even if you don't fully realize it. I don't know

Jaana: That was a good

Researcher: Do you have any examples of that?

Tarja: Well, politics comes to mind. During the U.S. elections, I only got content from one side, not much from the other

Jaana: That's good, I think others got that too, I don't even live in America, but I still got it. I got only about Kamala Harris. I can't even pronounce her name, but from her and nothing about Donald Trump

Researcher: So it was really skewed at that point?

Jaana: Yes, definitely. It might even influence who you vote for

Tarja: Yeah, even if we are not there, but in general

In this exchange, the students extend their reasoning about algorithmic recommendation and curation mechanisms to the political domain of another country, demonstrating awareness of how social media can shape public discourse and influence political attitudes. Tarja begins by explicitly referencing the concept of filter bubble, noting that recommendation systems can tailor users to see issues from only one perspective while obscuring alternative viewpoints, indicating recognition of the ideological consequences of algorithmically curated information. When asked for concrete examples, Tarja points to her experience during the U.S. elections, by reflecting observations of receiving content representing only one political side. Jaana affirms this by elaborating how she, despite not residing in the United States, was shown exclusively content about one candidate and not the other. This reflection illustrates how the students were able to apply the concepts of filter bubbles and polarization to critically interpret their personal experiences in ways connected to broader societal questions and ethical concerns.

In addition to politics, filter bubbles were also reflected in terms of friendship and concerns about eroding shared ground on which peer relationships are built, as one group of 9th graders reflected:

Researcher: Does it have any impact that different people see different content?

Siiri: Well, maybe the mildest example is when someone comes to school and asks, "Have you seen this video on TikTok?" And their friends are like, "Of course not. What video is that?" That's the mildest case, but it was the first thing that came to mind.

Researcher: Yes, that's a good example. Anything else?

Siiri: I always think of politics, like seeing only positive content about one person and negative about another, and then missing the fact that both have good and bad sides. Things aren't so black and white.

Elsa: Yeah, and it might also make you grow apart from a friend. If your friend sees a lot of content you really hate, or stuff like that. Then your interests might start to diverge a lot. And then friendships might not last anymore

Going beyond passive adaptation to existing data-driven practices and their harmful impacts on everyday life, one group of 9th graders described actively resisting these mechanisms by taking intentional, shared action:

Researcher: So, what could you do if you wanted to prevent it from affecting your

daily life, like the way you described, like TikTok constantly showing cat videos and you can't stop watching. What can be done?

Maija: Well, it's good to set restrictions in the settings and...

Liisa: If you could just avoid looking at them, then they probably wouldn't show up anywhere

Maija: And I've deleted some apps that I noticed take up so much time, like I keep opening them all the time, and you can't even study for a test without having the app open. So I've deleted some apps.

Liisa: Me too. And then we deleted them together as a group

Maija: It helps

Researcher: Really? That's a great joint decision

Maija: We all deleted Snapchat, almost all of us friends

Researcher: Okay

Maija: Because it takes so much time, it's so addictive

Liisa: And none of us have TikTok either

Maija: No, we don't

Liisa: That one's apparently super addictive too...

... if I deleted Snapchat alone, I might easily be left out. But if everyone deletes it, then no one has it. So no one is excluded

Rather than expressing apathy or portraying themselves as powerless in the face of persuasive technologies, the students describe concrete strategies for resisting platforms designed to be addictive, such as using restrictions and deleting applications. Importantly, the discussion foregrounds the collective dimension of transformative agency as they elaborate how deleting apps was a joint decision and coordinated action among friends, ensuring that no one would be socially excluded. Through this collective disengagement, the students resist not only platform addiction but also the social dependencies that bind users to these platforms. In doing so, they transformed activities they perceived as harmful by negotiating shared norms and taking coordinated action to reclaim control over their attention while simultaneously strengthening their friendships.

Based on their own experiences of participation in long-term AI education, students also articulated the need for schools to take action and educate about social media mechanisms and AI systems for all students. For example, one group of 6th graders emphasized the importance of raising understanding on a broad scale:

Researcher: What about the next level; parents, schools, and teachers? Can they do anything to make social media a better and safer place?

Meri: Maybe talk about these things. In schools, for others too, around the world and

in Finland, maybe that would help people understand a bit better.

Pirkko: I don't know, maybe in social or environmental studies there could be a whole unit just about social media. I'm not sure if that's social or environmental studies. I don't know, maybe there could even be a new subject, a "social media subject."

They highlighted strategies for institutional and collective responsibility as something that transcends the common "personal data strategies" and "data tactics" often presented in the literature on youth and social media (Pangrazio & Selwyn, 2023). By making a call for systematic education in schools, they propose that formal education is responsible for developing collective data literacy and agency—which is crucial as transformative change with regard to data-driven systems requires collective action rather than individual awareness (Stetsenko, 2019).

A similar vision was expressed by 9th graders, who reflected how others should also have the opportunities to develop such AI understanding that they now have:

Maija: I think everyone should know the basics of AI, what AI does, how it affects things, what data is collected, and on what basis things happen. Those basics would be good for everyone to know. Because we're in this... what's this project called again?

Researcher: Generation AI

Maija: Yeah, since we're part of this, we know. But probably not everyone, at least not the students in our school

Liisa: Definitely not

These reflections are significant because they not only demonstrate how students valued what they had learned, but also they recognized the asymmetry of knowledge between those who have participated in structured AI education and those who have not. They point out a crucial insight: Schools are one of the few institutions capable of reaching almost everyone in the age cohort, and, like vaccination programs, society-wide actions require a critical mass. Furthermore, their comments point toward a vision of critical AI literacy as a shared societal responsibility, where schools play a pivotal role in equipping all learners, not just those involved in specialized projects, with the conceptual tools needed to understand, question, and act upon the systems that shape their lives.

Taken together, the reflections between the interview groups illustrate how conceptual understanding was applied in contextualized ethical reflection involving personal and societal considerations of the impacts of data-driven systems. Students were not just recalling isolated definitions of algorithmic bias, profiling, recommendation, engagement optimization, and filter bubbles, but applied them in critical analysis on how these systems shape perceptions, opinions, behavior, emotions and social relationships in context-bound activities of everyday life. They further demonstrated critical consciousness of the intended and unintended consequences of these systems, including stereotyping, manipulation, addiction, social comparison, and polarization. While only one group described collective,

transformative action to resist platform power, the reflections across all groups suggest transformative learning in that students used scientific concepts to critically analyze and question existing data-driven practices shaping personal and societal lives.

6. Discussion

This study examined how students who participated in a cumulative, co-design-based AI learning path used AI concepts as mediating tools in situated collaborative reflection and ethical reasoning. Drawing on sociocultural and cultural-historical theories, we analyzed students' situated use of scientific concepts in reflective group interviews as indicators of their internalization as mediating tools for thinking. The following pages discuss how this mediation enabled students' participation in AI discourses, provided theoretical orientation to everyday data practices, and cultivated criticality and a transformative stance towards AI-driven sociotechnical practices.

6.1 *Participation in AI discourses.*

The findings show students' sustained and fluent use of AI concepts as part of situated discourses, where these concepts functioned as mediating psychological tools in explaining and reasoning about AI-related phenomena. From Vygotskian (1978) perspective, such persistence can be seen as a process where initially external scientific concepts (first encountered through specific tools and social interaction in classroom settings) had become internalized mental tools that could be used as part of language interactions beyond the original hands-on learning contexts. Notably, students used concepts and their relations even years after the original interventions and across discussion topics extending beyond the specific activities through which they were initially encountered. In addition, the result of the study indicated that contextualization of AI concepts in meaningful problem-solving (Rivet & Krajcik, 2008) and data-driven design processes cultivated feelings of ownership and agency, which further supported construction of conceptual knowledge and volition to learn scientific concepts that initially were unfamiliar to the students. From the perspective of the research question, AI concepts that were originally introduced in co-design and inquiry projects in a specific educational context became to mediate students' explanations, reasoning, and joint language in a broad range of discussion topics.

6.2 *Theoretical orientation to data-practices of everyday life.*

In the reflective interviews the students were also able to engage in collaborative reasoning that was no longer tied to immediate experiences and everyday conceptions (Vygotsky, 1978). Instead, their discussions about social media mechanisms indicated a more generalized, model-based reasoning about what happens under the hood of applications they are using in their everyday life. Although most student groups did not consider how the activities of other users influence their recommendations, they were still able to reason how data collection and profiling lead to tailored content and advertisements, and how such content

is optimized to sustain attention and maximize user engagement. From a sociocultural perspective, this situated use of scientific concepts can be interpreted as internalization involving an increasingly theory-informed orientation toward everyday life (Chaiklin & Hedegaard, 2013) as conceptual understanding was applied in language interactions to collaboratively explain underlying computational mechanisms and economic logics of social media services. From the perspective of the research question, this shows an orientation different from commonsense “folk theories” (DeVito et al., 2018). It shows a theoretical orientation, where internalized AI concepts and understanding of technological processes were used to interpret their everyday interactions on online platforms in the broader context of data-driven systems and revenue models of social media platforms.

6.3 Criticality towards AI driven socio-technical practices.

The results further revealed how conceptual understanding provided groundings for critical discourses of existing practices at multiple levels. At the personal level, students critically reflected how recommendation feedback loops can produce intended addictive behaviors, such as compulsive checking, explicitly linking these effects to data-driven mechanisms that infer preferences from their past behaviors. They also identified how repeated exposure to unrealistic images of “perfect women” negatively impacts on their self-esteem and how their opinions and beliefs may be shaped by influencers, whose visibility is amplified by social media mechanisms. At the interpersonal level, students reflected the concept of filter bubbles through eroding social ground in which their relations are built on. They also recognized how political influencing was part of their own feeds and extended their critical considerations to a more societal level of concerns of how filter bubbles influence people’s voting behaviors. At societal level, students also considered the harmful impacts of algorithmic biases but also proposed data-driven mitigation strategies. From the perspective of the research question, the students’ social critique shows how AI concepts provided epistemic resources for ethical reasoning, enabling students to connect their personal experiences (e.g., addiction and insecurity) to algorithmic and data-driven processes and their societal consequences, instead of purely personal problems for them.

6.4 Transformative data agency.

Addressing the remaining component of the research question, collective agency, the results of the study indicate how conceptual understanding provided important groundings for the development of critical consciousness, including the Freirean (1974) ability to re-read the world and question existing practices. Here, the students were critically analyzing how existing practices shape their personal behavior, opinions and identities and they also anticipated indirect effects of social media mechanisms on social relationships and public discourse. This illustrates a difference between surface-level descriptions of what they see or experience and deeper level considerations of how and why these data-driven systems encountered in everyday life mediate and engineer what people see or not, and with what potential consequences.

Taken together, these reflections between the interview groups suggest an epistemological shift associated with transformative learning and agency, involving changes in the ways students perceive and understand external reality. From this perspective, learning scientific AI concepts is not merely a matter of reproducing existing knowledge but of the gradual development and internalization of theoretical tools for thinking that can be applied to analyze a variety of context-bound activities, structures, and problems of everyday life at both personal and societal levels. For one group of students, this critical reading of data-driven practices had led to shared action, illustrating what can be understood as a form of radical transformative data agency (Stetsenko, 2019). After identifying problems in the existing practices, they decided to remove certain apps and acted together to resist the harmful effects of addictive systems in ways that prevented social exclusion. Here internalized AI concepts and understanding of data-driven processes mediated reflection but also action, enabling students to resist, protest, and disrupt socio-technical systems (e.g., app deletion together, collective disengagement) (for similar tactics in other contexts see Pangrazio & Selwyn, 2023, ch. 7) and to find consensus over norms to mitigate social risks such as exclusion.

While these findings are significant in showing how youth had taken an “activist stance” (Pangrazio & Selwyn, 2023), the question remains how to cultivate transformative agency in and through AI education more broadly. That is, how to support children and young people to develop their stances and voices, exert influence, and resist oppressive practices and imbalances of power in an AI-driven world (Stetsenko, 2019). This calls for AI education that is not just adaptation to existing beliefs, practices and technological transformations, but supports students in becoming change-makers who have the volition, understanding, and skills to question, shape, make, and transform AI and data-driven practices for the common good. Yet, education alone cannot transform the social, economic, technological, and political structures that undermine agency and fuel inequity. Instead of placing responsibility on individual actors, cultivating children’s data agency calls for multi-actor support and coordinated action across educational, technological, institutional, and policy domains.

7. Key Contributions

The results show that participation in a cumulative, three-year, co-design based AI learning path supported the internalization of AI concepts as mediating psychological tools in situated collaborative reasoning. Their use of these concepts reflected a theoretical orientation toward everyday online practices and provided epistemic resources for critical and ethical reflection on AI-driven sociotechnical practices. In one group, students’ critical understanding was also manifested in collective transformative action to resist and protest these practices.

Students fluently applied machine learning concepts, such as training data, profiling, engagement optimization, and probabilistic systems, one or two years after the original educational interventions. Whereas much AI education research reports short-term learning

gains, in this study the persistence of discipline-specific reasoning across multiple years suggests that AI concepts had become mediating tools for students' reasoning beyond the original learning contexts. What is more, learners consistently grounded their ethical reasoning in ML mechanisms instead of abstract moral judgments: bias as data distribution, addiction as engagement optimization, and distorted self-image as algorithmic amplification.

Whereas many AI literacy initiatives work at the individual level of “my data”, some students in this study demonstrated population-level reasoning about data-driven systems. This indicates an emerging understanding of how individual data contribute to collective outcomes. Approaches to AI education that focus on the individual level tend to foreground personal responsibility and individual coping strategies (Pangrazio & Selwyn, 2019), but students' reflections themselves also illustrate how data agency is relationally constituted through interactions among individuals, communities, cultural tools, and the broader social structures in which activity is situated. Some students also viewed the development of AI competences as a shared societal responsibility—AI education as a public good—which views AI education as a right.

Finally, the results of the study highlight the importance of applying ideas derived from educational theories and longstanding educational research when developing new educational practices, learning environments and scaffolding structures for AI education. Through contextualized AI inquiry, students were positioned as knowledge creators and active agents (Hakkarainen et al., 2013), who explored core concepts through meaningful problems (Edelson & Reiser, 2006; Rivet & Krajcik, 2008) and with the help of tools and curriculum materials designed to scaffold the construction of more integrated mental models of how these concepts interact. In addition, the students were scaffolded to make conceptual connections to their everyday life and societal issues throughout the progressive inquiries (Kahila et al., 2024). Yet, the formation of richer conceptual networks and the internalization of scientific concepts is a gradual process that takes time and requires revisiting core concepts across a range of activities and contexts (National Research Council, 2007; Vosniadou, 2013). While prior research has highlighted that multiple features of learning influence the transfer of knowledge, the findings of this study suggest that exploring AI concepts in multiple contexts supports the development of a more flexible representation of knowledge (National Research Council, 2000). Students' use of AI concepts was not task or tool specific to single contexts of application but flexible across personal, group, and collective contexts.

8. Implications for Policy and Practice

Instead of demonstrating causal effects of particular pedagogical practices, these theoretical and pedagogical insights, illustrated through the empirical analysis, point to promising design principles for AI education. These include providing sustained opportunities to revisit AI concepts across contexts, grounding conceptual learning in meaningful inquiry and co-design, and creating opportunities for students to use scientific concepts in collab-

orative reflection on the ethical and societal implications of data-driven systems. Several features of the co-design approach and CEDE model appear particularly relevant for cultivating transformative learning and agency in and through AI education.

First, our findings suggest that the design and implementation of transformative AI education depend on sustained cross-boundary collaboration between researchers and teachers. Involving teachers in all phases of the co-design process is necessary for the development of contextualized learning activities that are responsive to students' needs and supports the long-term cultivation of transformative learning and agency.

Secondly, learning tasks and AI concepts should be systematically contextualized to learners' everyday experiences, technologies, and socio-technical practices. Such contextualization provides opportunities to consider the relationship between students' prior knowledge, experiences, and folk theories and the new scientific concepts being taught. At the same time, connecting AI concepts to issues that are meaningful and relevant to learners' everyday lives can spark and sustain interest and motivation to work at the edge of their competences.

Thirdly, the process of learning should position children as agentic knowledge creators and critical future-makers. This means providing distributed scaffolding and mediated activities that enable children to experience agency and take ownership of their learning and participation. It involves making abstract AI concepts and ethical challenges as objects of collaborative learning and hands-on inquiry through learners' own designs, data, and inquiry questions.

Finally, the findings highlight the importance of cumulative learning paths that revisit and deepen core concepts over time. The development of richer conceptual networks takes time and is challenging, as it often requires the reorganization of internal mental models to adopt new ways of thinking (Vosniadou, 2013). Fundamentally, developing a conceptual and ethical understanding of data-driven mechanisms is not about knowledge acquisition nor adaptation to existing digital conditions and sociotechnical practices, but rather a gradual process of epistemic and transformative learning through which children are supported to develop mental tools and agency that enable them to examine, question, and shape those conditions, and to become responsible future makers.

9. Limitations and future directions

While this study is grounded on sociocultural insights on context-bound and situative nature of human learning and agency, such context-specific design can also be considered as limitation of empirical research. The interventions behind student reflections were conducted within a particular co-design process and in a specific educational setting in Finland, involving emergent interactions of particular students, teachers, educational technologies, curriculum materials, and larger educational structures (e.g. assessment practices, national curriculum, educational administration). Finnish teachers are highly educated professionals who enjoy a high level of trust and autonomy in their work, and there are no standard-

ized testing, external audits, or outside teaching supervision (Niemi & Lavonen, 2020). The importance of research-based approaches for developing educational practices is also acknowledged at all levels of the educational system, from the national strategy and policy levels to their implementation in teacher education (Niemi & Lavonen, 2020). While these contextual factors limit the transferability of the findings to educational systems operating under different structural conditions, they are significant in that they provide theoretical, pedagogical and empirical groundings for contextualized AI education that can support the development of scientific and ethical AI understanding.

In addition, the interview sample was relatively small and based on teacher-led intensity sampling and voluntary participation. Moreover, the study focused on students' collaborative reflections and reasoning constructed in specific social settings. Consistent with the sociocultural framework guiding the study, students' use of AI concepts as mediating tools for reasoning was understood as emerging within context-bound activity following the three-year learning path in which they had participated. It also means that the analysis does not capture individual levels of understanding and how it might be expressed in isolation or in different settings. However, we have earlier studied the development and persistence of data-driven reasoning and ethical understanding (Kahila et al., 2024; Vartiainen et al., 2025a; Vartiainen et al., 2025b) in earlier phases of this program and thus, the analytical focus here was placed on collaborative meaning making and reasoning. Yet, the qualitative analysis of students' reflections involved interpretation. For ensuring the quality and rigor of the analysis, we followed Shenton's (2004) strategies for increasing trust in qualitative research. We have provided transparent accounts of our theoretical and pedagogical commitments and interpretative framework, as well as detailed descriptions of the research context, data collection, and analytic procedures. All interpretations were discussed and critically examined among researchers who all were involved in the three-year project. We have also presented multiple data excerpts for allowing readers to assess the plausibility and trustworthiness of the interpretations. Although this study did not aim for generalizability, detailed descriptions of the theoretical foundation, pedagogical approach and co-design, research context, data, and methods enable readers to assess the transferability of the findings (Shenton, 2004).

Yet, a great deal of future work remains to be done. From a theoretical perspective, future work should continue building a robust theoretical framework for transformative AI education. This theoretical work should be grounded on empirical exploration of contextualized activity systems along with emerging contradictions within them (Engeström & Sannino, 2010). Transformative AI education should also move from reproductive and technodeterministic approaches toward formative interventions that involve teachers and learners from diverse cultural backgrounds, educational settings, and with different educational needs. Building solid theoretical foundations and research-based pedagogical models therefore depend on cross-boundary collaboration and long-term research conducted in and across multiple contexts and levels of education.

Acknowledgments

This study received funding from the Strategic Research Council (SRC) established within the Research Council of Finland, grants #352859 and #352876.

References

- Aleknavičiūtė, V., Lehtinen, E., & Södervik, I. (2023). Thirty years of conceptual change research in biology – A review and meta-analysis of intervention studies. *Educational Research Review*, 41, 100556. <https://doi.org/10.1016/j.edurev.2023.100556>
- Bal, A., Afacan, K., & Cakir, H. I. (2019). Transforming Schools from the Ground-Up with Local Stakeholders: Implementing Learning Lab for Inclusion and Systemic Transformation at a Middle School. *Interchange*, 50(3), 359–387. <https://doi.org/10.1007/s10780-019-09353-5>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021). On the Opportunities and Risks of Foundation Models. <https://doi.org/10.48550/arxiv.2108.07258>
- Büchi, M., Fosch-Villaronga, E., Lutz, C., Tamò-Larrieux, A., & Velidi, S. (2023). Making sense of algorithmic profiling: user perceptions on Facebook. *Information, Communication & Society*, 26(4), 809–825. <https://doi.org/10.1080/1369118X.2021.1989011>
- Chaiklin, S., & Hedegaard, M. (2013). Cultural-historical theory and educational practice: Some radical-local considerations. *NUANCES: Estudos Sobre Educação*, 25(1), 30–44.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science*, 5(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2
- Coeckelbergh, M. (2023). Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence. *AI and Ethics*, 3(4), 1341–1350. <https://doi.org/10.1007/s43681-022-00239-4>
- Creswell, J. W. (2006). *Qualitative Inquiry & Research Design: Choosing Among Five Approaches*. SAGE Publications: Thousand Oaks, CA, USA.
- DeVito, M. A., Birnholtz, J., Hancock, J. T., French, M., & Liu, S. (2018). How People Form Folk Theories of Social Media Feeds and What it Means for How We Study Self-Presentation. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3173574.3173694>
- Edelson, D. C., & Reiser, B. J. (2006). Making Authentic Practices Accessible to Learners: Design Challenges and Strategies. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 335–354). Cambridge University Press.
- Edwards, A. (2007). An interesting resemblance: Vygotsky, Mead, and American pragmatism. In *The Cambridge Companion to Vygotsky*. <https://doi.org/10.1017/CCOL0521831040.004>
- Elo, S., & Kyngäs, H. (2008). The Qualitative Content Analysis Process. *Journal of Advanced Nursing*, 62(1), 107–115.
- Engeström, Y. (2011). From design experiments to formative interventions. *Theory & Psychology*, 21(5), 598–628. <https://doi.org/10.1177/0959354311419252>
- Engeström, Y., Rantavuori, P., Ruutu, P., & Tapola-Haapala, M. (2023). From Future Orientation to Future-Making : Towards Adolescents' Transformative Agency. In N. Hopwood & A. Sannino (Eds.), *Agency and Transformation* (pp. 107–138). Cambridge University Press.

- Engeström, Y., & Sannino, A. (2010). Studies of expansive learning: Foundations, findings and future challenges. *Educational Research Review*, 5(1), 1–24. <https://doi.org/https://doi.org/10.1016/j.edurev.2009.12.002>
- Freire, P. (1974). *Pedagogy of the oppressed*. Seabury Press.
- Grover, S. (2024). Teaching AI to K-12 learners: Lessons, issues, and guidance. In *Proceedings of ACM Computer Science Education (SIGCSE) 2024 Conference*, pages 1–7, Portland, OR, USA. ACM.
- Gu, X. and Ericson, B. J. (2025). AI literacy in K-12 and higher education in the wake of generative AI: An integrative review. In *Proceedings of the 2025 ACM Conference on International Computing Education Research V.1, ICER '25*, pages 125–140, New York, NY, USA. ACM.
- Haapasaari, A., Engeström, Y., & Kerosuo, H. (2016). The emergence of learners' transformative agency in a Change Laboratory intervention. *Journal of Education and Work*, 29(2), 232–262. <https://doi.org/10.1080/13639080.2014.900168>
- Hakkarainen, K., Paavola, S., Kangas, K., & Seitamaa-Hakkarainen, P. (2013). Socio-cultural perspectives on collaborative learning: Towards collaborative knowledge creation. In C. Hmelo-Silver, C. Chinn, C. Chan, & A. O'Donnell (Eds.), *The International Handbook of Collaborative Learning* (pp. 57–73).
- Hakkarainen, K., & Seitamaa-Hakkarainen, P. (2022). Learning by inventing : Theoretical foundations. In T. Korhonen, K. Kangas, & L. Salo (Eds.), *Invention pedagogy : The Finnish approach to maker education* (pp. 15–27). Routledge.
- Hammer, D., & Berland, L. K. (2014). Confusing Claims for Data: A Critique of Common Practices for Presenting Qualitative Research on Learning. *Journal of the Learning Sciences*, 23(1), 37–46.
- Hedegaard, M. (2005). *Radical-local teaching and learning : a cultural-historical approach* (Seth Chaiklin, Ed.) [Book]. Aarhus University Press.
- Hopwood, N., & Sannino, A. (2023). Motives, Mediation and Motion: Towards an Inherently Learning- and Development-Orientated Perspective on Agency. In N. Hopwood & A. Sannino (Eds.), *Agency and Transformation* (pp. 1–34). Cambridge University Press.
- Kahila, J., Vartiainen, H., López-Pernas, S., & Tedre, M. (2026). Children's Data Privacy Concerns on Social Media. *Journal of Media Literacy Education*.
- Kahila, J., Vartiainen, H., Arkko, E., Lin, A., Pope, N., & Tedre, M. (2024). Enhancing Understanding of Data Traces and Profiling Among K–9 Students Through an Interactive Classroom Game. *Proceedings of the 19th WiPSCE Conference on Primary and Secondary Computing Education Research, WiPSCE '24*. <https://doi.org/10.1145/3677619.3677635>
- Kahila, J., Vartiainen, H., Tedre, M., Arkko, E., Lin, A., Pope, N., Jormanainen, I., & Valtonen, T. (2024). Pedagogical framework for cultivating children's data agency and creative abilities in the age of AI. *Informatics in Education*. <https://doi.org/10.15388/infedu.2024.15>
- Karpov, Y. V. (2003). Vygotsky's Doctrine of Scientific Concepts: Its Role for Contemporary Education. In A. Kozulin, B. Gindis, V. S. Ageyev, & S. M. Miller (Eds.), *Vygotsky's Educational Theory in Cultural Context* (pp. 65–82). Cambridge University Press.
- Krajcik, J. S., & Blumenfeld, P. C. (2006). Project-based learning. In *The Cambridge handbook of the learning sciences* (pp. 317–333). Cambridge University Press.
- Mertala, P., Fagerlund, J., & Calderon, O. (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, 100095. <https://doi.org/10.1016/j.caeai.2022.100095>
- Miao, F. and Shiohira, K. (2024). *AI Competency Framework for Students*. UNESCO, Paris, France.

- National Research Council. (2000). *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. The National Academies Press.
- National Research Council. (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8*. Committee on Science Learning, Kindergarten Through Eighth Grade. In *Teaching Science*.
- Niemi, H., & Lavonen, J. (2020). Teacher Education in Finland: Persistent Efforts for High-Quality Teachers. In L. Lefty & J. W. Fraser (Eds.), *Teaching the World's Teachers* (pp. 153–178). Johns Hopkins University Press.
- OECD (2025). *Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education (Review Draft)*. OECD, Paris, France.
- Pangrazio, L. and Selwyn, N. (2019). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*, 21(2):419–437.
- Pangrazio, L., & Selwyn, N. (2023). *Critical Data Literacies: Rethinking Data and Everyday Life*. The MIT Press.
- Pope, N., Kahila, J., Laru, J., Vartiainen, H., Roos, T., & Tedre, M. (2024). An Educational Tool for Learning about Social Media Tracking, Profiling, and Recommendation. 2024 IEEE International Conference on Advanced Learning Technologies (ICALT), 110–112. <https://doi.org/10.1109/ICALT61570.2024.00038>
- Pope, N., Kahila, J., Vartiainen, H., Saqr, M., López-Pernas, S., Roos, T., Laru, J., & Tedre, M. (2025). An XAI Social Media Platform for Teaching K-12 Students AI-Driven Profiling, Clustering, and Engagement-Based Recommending. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(28), 29203–29211. <https://doi.org/10.1609/aaai.v39i28.35194>
- Pope, N., Kahila, J., Vartiainen, H., & Tedre, M. (2023). Children's AI Design Platform for Making and Deploying ML-Driven Apps. <https://doi.org/10.36227/techrxiv.24320794.v1>
- Pope, N., Vartiainen, H., Kahila, J., Laru, J., & Tedre, M. (2024). A No-Code AI Education Tool for Learning AI in K-12 by Making Machine Learning-Driven Apps. 2024 IEEE International Conference on Advanced Learning Technologies (ICALT), 105–109. <https://doi.org/10.1109/ICALT61570.2024.00037>
- Puntambekar, S., & Hubscher, R. (2002). Scaffolding in complex learning environments: What we have gained and what we have missed. *Educational Psychologist*, 40, 1–12.
- Rivet, A. E., & Krajcik, J. S. (2008). Contextualizing instruction: Leveraging students' prior knowledge and experiences to foster understanding of middle school science. *Journal of Research in Science Teaching*, 45(1), 79–100. <https://doi.org/https://doi.org/10.1002/tea.20203>
- Rizvi, S., Waite, J., and Sentance, S. (2023). Artificial intelligence teaching and learning in K-12 from 2019 to 2022: A systematic literature review. *Computers and Education: Artificial Intelligence*, 4:100145.
- Sannino, A., & Engeström, Y. (2017). Co-generation of societally impactful knowledge in Change Laboratories. *Management Learning*, 48(1), 80–96. <https://doi.org/10.1177/1350507616671285>
- Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 97–118).
- Schoultz, J., Säljö, R., & Wyndhamn, J. (2001). Heavenly talk: Discourse, artifacts, and children's understanding of elementary astronomy. *Human Development*, 44(2–3), 103–118.
- Shenton AK. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22, 63–75. <https://pdfs.semanticscholar.org/452e/3393e3ecc34f913e8c49d8faf19b9f89b75d.pdf>

- Shor, I. (1993). Education is politics: Paulo Freire's critical pedagogy. In P. McLaren & P. Leonard (Eds.), *Paulo Freire: A critical encounter* (pp. 25–35). Routledge.
- Solaiman, I. et al. (2023). Evaluating the social impact of generative AI systems in systems and society. *ArXiv Preprint ArXiv:2306.05949*.
- Stetsenko, A. (2019). Radical-Transformative Agency: Continuities and Contrasts With Relational Agency and Implications for Education. *Frontiers in Education*, Volume 4-2019. <https://doi.org/10.3389/educ.2019.00148>
- Stoilova, M., Livingstone, S., & Nandagiri, R. (2020). Digital by default: Children's capacity to understand and manage online data and privacy. *Media and Communication*, 8(4), 197–207. <https://doi.org/10.17645/mac.v8i4.3407>
- Tedre, M., Denning, P. J., and Toivonen, T. (2021). CT 2.0. In *Proceedings of the 21st Koli Calling International Conference on Computing Education Research, Koli Calling '21*, pages 1–8, New York, NY, USA. ACM.
- van der Hof, S. (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):409–445.
- Vartiainen, H., Kahila, J., Tedre, M., López-Pernas, S., & Pope, N. (2025a). Enhancing children's understanding of algorithmic biases in and with text-to-image generative AI. *New Media & Society*, 9(27), 14614448241252820. <https://doi.org/10.1177/14614448241252820>
- Vartiainen, H., Kahila, J., Pope, N., López-Pernas, S., Valtonen, T., & Tedre, M. (2025b). Long-term Impacts of K-12 AI Education Interventions on Social Media Mechanisms. *Proceedings of FIE 2025*.
- Vartiainen, H., Kahila, J., Tedre, M., & Pope, N. (2026). Understanding Data-Driven Mechanisms of Social Media Can Facilitate Collective, Transformative Agency. *Proceedings of EdMedia 2026*, 1600–1608.
- Vartiainen, H., & Tedre, M. (2024). How Text-to-Image Generative AI Is Transforming Mediated Action. *IEEE Computer Graphics and Applications*, 44(2), 12–22. <https://doi.org/10.1109/MCG.2024.3355808>
- Vartiainen, H., & Tedre, M. (2025a). The CEDE Model: A Learning-Sciences Based Framework for Critical and Transformative K–12 AI Education. *25th Koli Calling International Conference on Computing Education Research, Koli Calling '25*.
- Vartiainen, H., & Tedre, M. (2025b). Towards transformative AI education through cross-boundary co-design. *Proceedings of International Conference on Smart Learning Environments (ICSLE 2025)*.
- Vosniadou, S. (2013). Conceptual change in learning and instruction: The framework theory approach. In *International Handbook of Research on Conceptual Change*. <https://doi.org/10.4324/9780203154472>
- Vygotsky, L. (1978). *Mind in Society*. Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Vygotsky, L. S. (2012). *Thought and language*. The MIT Press.
- Weidlich, J., Gašević, D., Drachsler, H., & Kirschner, P. (2025). ChatGPT in Education: An Effect in Search of a Cause. *Journal of Computer Assisted Learning*, 41(5), e70105. <https://doi.org/10.1111/jcal.70105>
- Wells, G. (1994). Learning and teaching “scientific concepts”: Vygotsky's ideas revisited. <https://gpc-maths.org/data/documents/wells-scientificconcepts.pdf>
- Wertsch, J. V. (1993). *Voices of the Mind*. <https://doi.org/10.4159/9780674045101>

- World Economic Forum. (2011). *Personal Data: The Emergence of a New Asset Class*. World Economic Forum, Cologny, Switzerland.
- Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75–89. <https://doi.org/10.1057/jit.2015.5>