Supporting Teachers in Teaching Critical AI Literacy from Co-design to Classroom Implementation

Yu-Yu Liu
Aarhus University, Denmark
yuyu.liu@cc.au.dk

Inger Malene Surrow
Aarhus University, Denmark
imsurrow@cc.au.dk

Ekta Shokeen
Aarhus University, Denmark
ekta.shokeen@cc.au.dk

Luke Connelly
Aarhus University, Denmark
luke@cs.au.dk

Ane Vielandt Jensen
Aarhus University, Denmark
av@cc.au.dk

Mikkel Hjorth
VIA University College, Denmark
misi@via.dk

Bjarke Lindsø Andersen
Professionshøjskolen Absalon,
Slagelse, Denmark
bjan@pha.dk

Christian Dindler
Aarhus University, Denmark
dindler@cavi.au.dk

Marianne Graves Petersen
Aarhus University, Denmark
mgraves@cc.au.dk

Ole Sejer Iversen
Aarhus University, Denmark
oiversen@cc.au.dk

As Artificial Intelligence (AI) becomes increasingly ubiquitous in children's everyday lives, it is crucial to develop children's critical AI literacy. Yet, teaching critical AI literacy remains challenging in practice, with classrooms defaulting to technical skill acquisition while marginalising critical reflections on AI's implications. Recent reviews call for supporting teachers in teaching critical AI literacy. In response, this paper explores how teachers develop strategies for teaching critical AI literacy through co-design and what happens when they implement them in the classroom. Following a design-based research approach, we trace one teacher's journey in detail across co-design workshops, pre- and post-implementation interviews, and observation in a 7th-grade classroom. Using dialogic formative assessment as an analytical lens, we identify opportunities and challenges across four phases: co-design workshops, independent lesson planning, classroom implementation, and post-implementation reflection. We propose research directions for HCI research for designing support for facilitating critical AI literacy in formal education: (1) supporting teachers in developing dialogic facilitation repertoires; and (2) designing tools as dialogic partners that scaffold critical reflection on AI's societal implications.

Critical AI Literacy. Computing Education. Design-based Research.

1. INTRODUCTION

As Artificial Intelligence (AI) increasingly shapes children's everyday experiences, researchers in Human-Computer Interaction (HCI) and Child-Computer Interaction (CCI) have highlighted the importance of critical AI literacy in K-12 education (e.g. Charisi et al. 2020; Shrivastava et al. 2024; Veldhuis et al. 2025; Schaper, Malinverni, and Valero 2020; Yim 2024). Building on AI literacy framework (Long and Magerko 2020), critical AI literacy extends beyond technical understanding to include critical reflection on AI's societal implications (Veldhuis et al. 2025).

However, critical AI literacy in K-12 still gravitates toward technical skill development, with limited focus on fostering children's agency, critical reflections on AI's implications (Bosch et al. 2025; Yim 2024). Therefore, recent reviews have called for research supporting teachers' pedagogical strategies for critical AI literacy (Velander, Otero, and Milrad 2024; Veldhuis et al. 2025; Yim 2024). In response, our study explores two research questions: **(1) How can we support teachers in developing teaching strategies for critical AI literacy through co-design? and (2) How are**

© Liu et al. Published by BCS Learning and Development Ltd.

these teaching strategies operationalised within classroom practice?

We adopted a Design-based Research (DBR) approach (Anderson and Shattuck 2012), conducting co-design workshops with teachers around critical AI literacy and introducing the learning tool, Engine Room. We then traced one teacher from workshop participation to classroom implementation and reflection interviews. To analyse the study, we combined dialogic learning and formative assessment as an analytical lens (Wegerif 2013; Bennett 2011), examineing how interations among tools, students, and teachers support or constrain critical reflection.

We find that co-design between teachers and researchers enabled the development of high-level teaching strategies for critical AI literacy but did not sufficiently prepare teachers to facilitate critical reflection in the classroom. We further find that the tool's affordances prioritised technical understanding over personal relevance and meaningful collaboration, limiting its ability to support dialogic critical reflection. Based on these findings and informed by critical AI literacy and dialogic learning theory, we contribute two directions for future HCI research: (1) supporting teachers' dialogic facilitation repertoires, and (2) designing critical AI literacy tools as voices in dialogic space that scaffold critical reflection.

2. RELATED WORK

2.1 Critical AI Literacy: From Theoretical Frameworks to Implementation in K-12 Education

Researchers have increasingly called for AI literacy in K-12 contexts, broadly defined as learners' ability to use, understand, and critically engage with AI systems and their impacts (Long and Magerko 2020; Ng et al. 2021; Touretzky et al. 2019). Beyond technical understanding, recent curriculum frameworks have integrated critical reflection on AI's implications as a core dimension of AI literacy. These efforts include the RAD framework (Recognize, Analyse, and Deliberate) (Solyst et al. 2025), the DORIT model (Construction, Purpose, Use, Argumentation, Value, and Consequences) (Dindler et al. 2023), and the project-based RAICA (Responsible AI for Computational Action) curriculum (Bosch et al. 2025). This trend aligns with the current agendas in CCI research, e.g., computational empowerment (Iversen, Smith, and Dindler 2018; Antle et al. 2021), and critical data literacy (Dangol and Dasgupta 2023; Hautea, Dasgupta, and Hill 2017; Liu, Dybboe, et al. 2026), intelligence-based AI literacy (Yim 2024), which emphasises students' agency to critically question and reconstruct technologies.

Synthesising this body of work through critical literacy learning objectives, Veldhuis et al. (2025) established a critical AI literacy framework. Positioning learners as both users and designers of AI, this framework advocates for fostering students' critical reflection on AI's societal and ethical implications. Their establishment represents the field's growing consensus that preparing children to thrive in AI-mediated environments requires not only technical understanding but also critical reflection on AI systems' implications.

In this paper, we use "critical reflection" to refer to learners' engagement with the purposes, values, and consequences of AI systems (Dindler et al. 2023), questioning for whom AI systems are designed, what values they embed, and what societal implications they carry (Velander, Otero, and Milrad 2024; Veldhuis et al. 2025).

Despite the growing consensus around critical AI literacy in research and curriculum design, a persistent gap remains between frameworks and classroom practice. Research on K-12 critical AI literacy often emphasises technical skills over critical reflection, agency, and assessment (Casal-Otero et al. 2023; Yim 2024; Druga, Otero, and Ko 2022). As a result, opportunities for students to interrogate AI systems in relation to bias, values, and social consequences remain sidelined (Kafai and Proctor 2022; Ng et al. 2021; Yim 2024). Recent reviews (Velander, Otero, and Milrad 2024; Veldhuis et al. 2025; Yim 2024) therefore call for more research efforts to support teachers in teaching critical AI literacy by developing their own knowledge and pedagogical strategies. Our study directly responds to this urgent need in the field.

2.2 Teachers' Role in Teaching Critical AI Literacy

Recent HCI research has developed a range of tools and resources aimed at supporting critical AI literacy in K-12 contexts. These interventions include robotic platforms aimed at supporting children's creativity (Han and Han 2025; Ali et al. 2024; Williams et al. 2024) and workshops engaging students with ethical scenarios (Morales-Navarro, Noh, and Kafai 2025; Bilstrup, Kaspersen, and Petersen 2020; Salac et al. 2023; Morales-Navarro et al. 2024), and authentic civic datasets (Kaspersen et al. 2021; Kaspersen et al. 2022). Other approaches deploy design fiction scenarios set in familiar contexts for children (Collyer-Hoar, Rubegni, and Tomczyk 2026), speculative design cards and worksheets (Kenny et al. 2025), and web-based image generation platforms that spark reflection on algorithmic bias (Collyer-Hoar et al. 2025). These efforts resonate with the culturally responsive computing framework, which positions students' lived experiences, cultural identities, and community knowledge as learning assets from

which critical reflection on technology can emerge (Scott, Sheridan, and Clark 2015).

However, facilitating critical reflection on AI implications in teacher-led classrooms remains challenging (Veldhuis et al. 2025; Antle et al. 2021). As Dangol et al. (2025) note, static or rigid AI education would struggle to cultivate an accurate understanding of AI's capabilities given the unprecedentedly rapid AI evolution. Bosch et al. (2025) found that teachers were more likely to modify or replace activities related to critical AI literacy than those focused on technical AI concepts, suggesting that teachers' challenges in teaching critical AI literacy are not primarily curricular but pedagogical. Further, prior work reports that teachers often feel underprepared to facilitate discussions around bias, fairness, and accountability in AI systems (Ng et al. 2021). It requires teachers to continuously reflect on how their facilitation is connected to students' identities, lived experiences, and sociocultural contexts, and cannot be achieved solely through professional development focused on subject content (Scott, Sheridan, and Clark 2015).

Therefore, supporting teachers requires pedagogical and assessment approaches that embed critical reflection within learning, rather than treating it as an optional add-on activity. Prior reviews note that assessment practices for critical AI literacy remain underdeveloped (Liu et al. 2025; Shokeen et al. 2025; Casal-Otero et al. 2023). At the same time, co-design workshops with teachers and AI experts show promise for supporting teachers' theoretical understanding of critical AI literacy while engaging them as active co-designers of pedagogical strategies (Casal-Otero et al. 2023).

2.3 Overview of Our Lens of Dialogic Formative Assessment

To examine teachers' pedagogical strategies, we combine dialogic learning and formative assessment as an analytical lens to inform our co-design workshops and data analysis.

Dialogic learning views knowledge as co-constructed through dialogue and interaction around critical problems (Alexander 2008; Shor and Freire 1987; Major et al. 2018; Linden and Renshaw 2004) By treating dialogue as a means to transform social relations in the classroom and raise awareness of broader societal relations, dialogic approaches foreground student voices and shift learners from passive recipients to critical thinkers (Shor and Freire 1987; Luke, 2014; Manalo 2020). Following Wegerif (2013), we extend the definition of dialogue beyond teacher-student interaction to include interaction with learning tools. From this perspective, technology act as an active participant by opening and expanding dialogic space, while making learner's engagement and participation visible. Similarly, Veldhuis et al. (2025) advocate combining dialogic approaches to foster critical reflection on AI's implications throughout design and making activities. However, much critical reflection remains internal (Vygotsky 1980) and therefore remains difficult for teachers to recognise and respond to in the classroom.

Formative assessment addresses this challenge by making learning visible through observable learning evidence embedded in activities (Shokeen et al. 2025; Biggs 1996; Wiliam 2018). Formative assessment is the process through which teachers and students use feedback to adjust teaching and learning in order to improve intended learning outcomes (Bennett 2011). Through dialogue, teachers can interpret students' verbal responses, interactions, and engagement with learning tools as evidence of learning, supporting metacognition (Flavell 1979). This perspective is particularly relevant for critical AI literacy, where recent work emphasises process-oriented assessment over evaluating static outcomes (Liu et al. 2025; Veldhuis et al. 2025; Liu, Jensen, et al. 2026).

In our study, we operationalised these two frameworks at two levels. During the co-design workshops, formative assessment provided a practical vocabulary for identifying "signs of learning" related to technical and critical objectives. During analysis, we combined dialogic learning and formative assessment to examine how interactions among teacher, student, and tool supported or failed to facilitate critical reflection.

3. METHODOLOGY

We adopted a qualitative Design-Based Research (DBR) approach for its affordances of building theoretical understanding in a practical context, focusing on designing and testing a significant intervention, and involving collaboration between researchers and practitioners (Anderson and Shattuck 2012). Our primary objective was to advance a theoretical agenda regarding critical AI literacy in interaction design and to provide a practical impact on classroom practice. While DBR typically involves multiple iterations, Anderson and Shattuck (2012) acknowledge that iteration of DBR is often challenging given the limited resources and time available to researchers and teachers. They characterise the evolution of DBR as a process of "research through mistakes" (p.17), noting that interventions are rarely implemented perfectly.

Our study represents the initial cycle of this DBR process. Our approach to collaboration was that researchers provided the conceptual frameworks and technical tools, while teachers retained final autonomy over lesson planning and implementation. Following the workshops, researchers remained available on demand to provide consultation as

teachers finalised their lesson plans. However, researchers adopted a non-interventionist stance during the actual classroom implementation. This choice was a deliberate DBR strategy to evaluate how our principles of critical AI literacy were enacted and facilitated independently by a practitioner in real classrooms. The outcomes of this initial cycle provide the essential real-world insights required to refine our theoretical understanding and evolve our design principles for future work.

The study consisted of two sequential phases: co-design workshops and a subsequent classroom case study tracking their implementation using the learning tool, Engine Room.

3.1 Learning Tool: The Engine Room

The Engine Room is an AI literacy tool providing digital and unplugged activities centred on simple natural language processing (NLP) algorithms, including Markov Chains, n-grams, and bag-of-words techniques. Within the activities, learners can choose and customise the datasets to introduce subject-specific content, as demonstrated in (Connelly, Bilstrup, and Petersen 2025). These activities allow teachers and students to explore how text becomes data for classification or generation. We use three activities shown in Figure 1: (a) the unplugged bag of words; (b) text-generating word wheels; and (c) N-gram Markov Chain text generation. We chose the Engine Room because: (1) the platform and its learning activities support Danish (the native language of teachers and students); (2) the tool was developed through a two-year participatory design process with teachers explicitly aimed at AI literacy education and Computational Empowerment (Connelly, Bilstrup, and Petersen 2025; Iversen, Smith, and Dindler 2018), to empower children through constructive, analytical and critical engagement with AI systems.



(a) An unplugged bag of words, where students examine and attempt to guess which text has been cut into words. Using the bag, they can reflect on tokenisation practices and what is gained/lost by representing text in this way.

(b) Text-generating word wheels, where students can spin the wheels to create a sentence. From this, students can reflect on how the system could be improved, as well as flaws in the existing system.

(c) N-gram Markov Chain text generation, where students can work with larger, custom datasets which directly result in a text generation model, as well as seeing the model "weights" for each word visualised.

Figure 1: *The three activities from the Engine Room used in this paper.*

3.2 Curricular Alignment

To situate our study in a formal educational context, our research design and interventions were explicitly aligned with the Danish national curriculum Technology Comprehension (the Danish Ministry of Education 2026). Theoretically grounded in the Computational Empowerment framework (Iversen, Smith, and Dindler 2018), the TC curriculum promotes students' capacity to constructively explore and critically analyse how technologies are embedded with human values and intentions, how they shape our lives individually and as society, and to criticall engage with their ethical implications (Andersen et al. 2024; Liu, Jensen, et al. 2026). By anchoring our work in this curriculum, we operationalise the academic construct of critical reflection outlined in critical AI literacy literature (Veldhuis et al. 2025) into formal classrooms.

3.3 Data Collection and Study Design

3.3.1 Phase 1: Co-design Workshops:
Adopting (Casal-Otero et al. 2023)'s approach, we conducted two co-design workshops including teachers and our research team from AI literacy, teacher training, learning sciences, and CCI to collaboratively develop didactic proposals. The workshop activities are summarised in Appendix A.1. The participants were eight K-9 teachers from five public schools in two municipalities, recruited based on two criteria: (1) prior experience in K-9 technology education, and (2) currently teaching grades 7–9 (lower secondary school).

Workshop 1 (Critical AI Literacy and Computational Empowerment): We introduced teachers to the frameworks of critical AI literacy and CE (Iversen, Smith, and Dindler 2018) and the DORIT model (Dindler et al. 2023). We introduced concepts, such as bag of words, tokenisation, and N-grams, and linked them to modern Large Language Models (LLMs). Teachers developed knowledge and brainstormed pedagogical strategies for connecting technical understanding to broader societal implications of AI.

Workshop 2 (Learning Evidence for Formative Assessment): To mitigate the challenge that K-12 critical AI literacy often emphasises technical skills over critical reflection, agency, and assessment (Casal-Otero et al. 2023; Yim 2024; Druga, Otero, and Ko 2022), this workshop focused on designing formative assessment evidence. Teachers articulated learning goals, designed activities, and identified "signs of learning" (evidence of students' understanding and critical reflection) (Biggs 1996; Wiliam 2018). We encouraged teachers to set learning goals based on the DORIT model (Dindler et al. 2023) to intentionally scaffold the transition from technical AI understanding and execution (technology use and construction) toward critical reflection (such as argumentation, purposes, value, and consequences).

3.3.2 Phase 2: Classroom observation and Case Study

Following the workshops, we tracked multiple participating teachers to evaluate their respective implementations. Yet, in this paper, we intentionally report an in-depth exemplar case study of one teacher (Male, 11 years of STEM teaching experience, located in Aarhus). This teacher was selected because he produced rich workshop outputs and demonstrated strong engagement in the project.

Pre-implementation Interview: Prior to the classroom session, we conducted a semi-structured interview with the teacher. The goal was to review the lesson plan finalised by the teacher and discuss the teacher's pedagogical strategies. See Appendix A.2. for the pre-implementation semi-structured interview guideline.

Classroom Observation: We observed the teacher teaching a 90-minute lesson in a 7th-grade classroom. The class consisted of 16 students (10 male, 6 female). Students were organised by the teacher into pairs to facilitate collaboration. Two GoPro cameras recorded two randomly selected pairs. Recordings captured peer dialogue and interactions with the tool. Observation focused on teacher instruction, student interaction with the tool, classroom dialogue.

Post-implementation Interview: Immediately following the lesson, we conducted a semi-structured interview. We guided the teacher to reflect on the implementation, students' learning outcomes, and the effectiveness of the formative assessment strategies. See Appendix A.3. for the post-implementation semi-structured interview guideline.

3.4 Data Analysis

Data Preparation and Transcription: All video recordings and interview audio were imported into Condens (Condens Insights GmbH 2025) for management and transcription. Video data was reviewed alongside transcripts to capture non-verbal interactions and screen-based activities. The data was transcribed and translated from Danish to English. To ensure accuracy, a researcher who is a native speaker of Danish proofread all translated transcripts against the original audio.

Coding Process: Data analysis used a hybrid thematic analysis approach, combining deductive and inductive coding (Braun and Clarke 2006; Azungah 2018). Two researchers began with a deductive codebook (See Appendix A.4) informed by the theoretical frameworks of dialogic learning, formative assessment, and critical AI literacy. After familiarising themselves with the dataset, the researchers coded the data synchronously, allowing for emergent inductive codes (See Appendix A.5).

Reflexivity and Researcher Positionality: Because analysing the qualitative classroom vignettes is highly interpretive, we actively managed our positionality during analysis (Holmes 2020). Our interdisciplinary research team brings together complementary expertise in teacher training, learning sciences, CCI, and computer science focused on AI literacy. The entire team collaborated closely on the study design, workshop facilitation, and empirical data collection, establishing an engaged relationship with the participant context.

As a research team, we recognised that it is inherently more straightforward to teach technical AI mechanisms than it is to facilitate relevant critical reflections (Ng et al. 2021). This awareness informed the design of and our division of labour for workshops in Phase 1. Researchers from computer science and CCI backgrounds led Workshop 1 to develop teachers' conceptual knowledge of AI mechanisms and computational empowerment, while researchers from learning sciences and teacher training led Workshop 2 to help teachers design learning activities based on formative assessment strategies (Biggs 1996; Wiliam 2018) aimed at scaffolding classroom dialogue and critical reflection.

To balance these perspectives during analysis, the team members with extensive K-12 teaching experience led the initial analysis of the classroom recordings, bringing an insider pedagogical lens to examine the teacher's challenges. To mitigate individual bias and ensure relevance to interaction design, researchers specialising in CCI and learning sciences then cross-examined these interpretations. Rather than relying on rigid statistical intercoder reliability, we leveraged our varied disciplinary perspectives to interrogate disagreements and arrive at a negotiated agreement on the final themes (Saldaña 2021; Clarke et al. 2023).

4. FINDINGS

In this section, we present our thematic analysis of the focal teacher's workshop outputs, interview transcripts, and classroom observation. The findings are organised as four phases of the teacher's journey, moving chronologically from co-design workshops through classroom implementation. In each phase, we describe the support, gains, and challenges the teacher perceived in teaching critical AI literacy and highlight specific moments where implementation did not actualise the pedagogical strategies. We align the empirical findings with the phases of the teacher journey and outline the gains and challenges, as illustrated in Appendix A.6.

4.1 Phase 1: Developing Pedagogical Strategies for Critical AI Literacy in Co-design Workshops

Through two co-design workshops, the teacher developed knowledge about critical AI literacy and high-level pedagogical strategies for teaching it in

his classroom. The workshop outputs produced by the focal teacher, including sticky notes and lesson plan templates, reveal that the teacher developed a genuine understanding of AI and showed confidence in planning for both technical and critical goals and formative assessment evidence within the supported co-design workshop environment.

In the first workshop, teachers were invited to brainstorm lesson ideas using inspirational cards that paired AI technical concepts, such as N-grams and word clouds, with critical reflection questions (e.g., the purposes embedded in machine learning technologies) (See Appendix A.7). For instance, teachers were provided with questions for inspiration, such as "What functions do tools like SchoolGPT serve and for whom?" The focal teacher engaged actively with fellow teachers and researchers, generating ideas for connecting these AI concepts through the Engine Room tool to classroom contexts in ways that would be relevant to his students. At this stage, these initial ideas were exploratory rather than complete lesson plans.

In the second workshop, the teacher developed a more detailed lesson plan using a structured template that aligned learning objectives, teaching activities, and formative assessment strategies (See Appendix A.8). The teacher articulated learning objectives that addressed, respectively, technical understanding, critical reflection, and hands-on construction. Moreover, he identified specific signs of learning he would look for to assess these learning objectives. He articulated the reliability and bias of AI as explicit learning objectives for critical reflection. Further, he planned to find learning evidence based on students' reflection on how the database determines the GenAI output and consequences through written outputs and dialogic discussion.

In this phase, the teacher demonstrated interest and confidence in planning for the dual goal of technical and critical AI literacy. The collaborative and high-level nature of the co-design workshops, which focused on conceptual goal setting and brainstorming rather than detailed activity design, made learning goals of critical AI literacy achievable and manageable. Co-design with teachers when providing a hands-on AI literacy learning tool and structured lesson plan templates successfully supported the teacher in developing pedagogical knowledge of critical AI literacy and articulating both technical and critical objectives as equally valued learning goals. Crucially, the teacher's plans also reflected contextual knowledge of his students and classroom that the research team could not generate independently, reinforcing the value of co-design as a method for developing pedagogically situated strategies rather than generic curricula. However, as the following phase reveals, this confidence would be tested when the teacher moved from collaborative workshop brainstorming to independently finalising his lesson plan for classroom implementation.

4.2 Phase 2: From Pedagogical Strategies into Classroom Implementation

When the focal teacher moved from the collaborative workshop environment to independently finalising his lesson plan for a specific class and timeframe, the confidence he had shown during co-design workshops began to wane. The pre-observation interview reveals two interrelated challenges: an instructional tension between dual learning goals and a recognition that the learning tool placed the responsibility of critical AI literacy facilitation primarily on the teacher.

Tension Between Technical and Critical Goals: During the pre-observation interview, the teacher identified a persistent instructional tension between his two primary learning objectives: technical AI concepts (e.g., N-grams) and critical AI literacy goals (e.g., fairness and agency). While he understood the importance of critical reflection on AI implications and valued critical AI literacy as a framework, he expressed a pragmatic preference for teaching technical concepts, noting that the learning activities facilitated by the tool were mostly related to technical aspects. He described his planning process as *"*messy*"*, explaining that he felt like *"pulling in two directions: that the formative assessment should be about critical AI literacy, and that most of the AI activities are obviously related to the technical aspects. We have to see if those two work together."* This tension reveals the gap that, while co-design successfully supported the teacher in articulating both goals as equally valued, the concrete activity design exposed that technical goals were structurally embedded in the tool, while critical reflection goals remained dependent on teacher-led facilitation.

Reflection on the learning tool's focus on technical concepts: A further challenge emerged in the teacher's reflection on the learning tool itself. Despite the tool's stated aim of supporting critical AI literacy, the teacher observed that it primarily made the technical mechanics of NLP and LLMs transparent, noting: "*Critical AI literacy is not the most obvious aspect in the learning tool. It's something derived from it."* His recognition showed that the teacher understood that the learning tool would not independently provoke students' critical reflection on AI's implications; it required him to actively derive critical reflection dialogue from learning activities, compounding the instructional pressure he was already navigating.

Together, these challenges demonstrate that by the time the teacher entered the classroom, his strategies for teaching critical AI literacy were strained. The following phase examines what

happened when these challenges encountered the reality of classroom implementation.

4.3 Phase 3: Implementing Critical AI Literacy in Classroom

The classroom observation revealed that the triadic dialogue among teacher, student, and tool successfully mediated technical understanding but encountered challenges when transitioning to critical reflection. This section first presents the contexts in which dialogic learning emerged, then discusses two specific moments where facilitation of critical reflection became difficult.

4.3.1 Technical Understanding Through Tool Interaction and Teacher-led Inquiry Scaffolding

The teacher's plan to teach the probabilistic nature of LLMs through analogy was successfully realised through students' hands-on engagement with the "word wheels" (*Figure 1b)* from the Engine Room tool. As students spun the wheels to generate sentences, the interaction with the tool functioned as a form of private dialogue, which is an iterative cycle of action and reflection facilitated by the analogy itself. Students were observed spinning the wheel, reading the generated results, and writing down the sentences, such as *"*I love Michael*"*. Repeated interaction with the analogy prompted students to externalise their thinking, i.e., *"*'I love Michael'. We had that sentence before.*"* This captured the learning evidence that a student realised that the "weighted model" behaviour they were witnessing was effectively "probability", a concept they recognised from their prior mathematics education. Here, the tool successfully mediated technical understanding without requiring explicit teacher facilitation, functioning as an active participant in the dialogic learning process.

After this tool-mediated private dialogue, the teacher transitioned students' internal reasoning into a shared space through inquiry scaffolding. By using inquiry, the teacher prompted students to reflect not only on the AI's output, but also on their own metacognitive strategies during the learning activities. This was particularly evident during the "bag of words" exercise, where the teacher not only checked for correct answers but also assessed the students' underlying reasoning. As the teacher later reflected in the post-interview: *"I had a dialogue with some different students about what I mean by 'the bag of word'. For example, I asked them what kind of strategies they used (to find out the text in the bag of words)."* By inviting students to externalise their strategies, the teacher gathered formative evidence of technical AI literacy, gauging whether students could articulate the logic of data processing rather than simply complete the task. These moments show that the combination of tool-mediated private dialogue and teacher-led inquiry scaffolding effectively supported students in developing and articulating technical understanding.

4.3.2 Moments When Facilitation of Critical Reflection Became Difficult

Contrary to the teacher's planned goals, the classroom observation revealed a scarcity of critical reflection dialogue. While students were physically grouped in pairs, their interaction was often parallel rather than collaborative, and little evidence of critical reflection on AI's implications was observed. Below, we identified two moments in the observation, showing how the structural conditions make critical reflection difficult.

Moment 1: Retreat to Technical Explanation. When the teacher's initial attempts to prompt critical reflection failed to elicit student engagement, the teacher retreated to direct technical explanation. During a discussion on chatbot limitations, the teacher started with an inquiry prompt, *"What could you use a chatbot for?"* When a student provided a minimal response *"ask questions"*. At this moment, instead of probing the student's understanding to spark a critical discussion on AI ethics or agency, the teacher's pedagogical strategy shifted from inquiry to direct explanation:

> **Teacher:** "Yes... But there are some things that you shouldn't ask a chatbot. There is a guard... a stop that says if you have certain thoughts, you should seek help. What happens then?"
> **Student:** "There are some rules behind chatbots?"
> **Teacher:** "Yeah... there is a system prompt and some guidelines... such stop blocks to tell chatbots that 'OK, this must not be answered.'"

This teacher-student dialogue illustrates the facilitation gap identified in Phase 2: the teacher lacked a concrete repertoire of dialogic facilitation for critical reflection. The student's question *"There are some rules behind chatbots?"* could be an opening for critical inquiry into AI's bias and agency, but the teacher redirected it toward technical mechanics. As the teacher focused on delivering technical content, the discussion intended for critical reflection was sidelined, and the lesson defaulted to the technical comfort zone he had predicted but hoped to avoid. This challenge shows that the teacher needed more resources and support in facilitation for critical AI literacy in the classroom.

Moment 2: Self-sufficient Activities Undermined Collaboration and Dialogue Opportunities. We found that the self-sufficient learning activities undermined opportunities for collaboration and dialogue. The initial learning tasks, such as the "wheel analogy" (Figure 1b) and text generation (Figure 1c), were sufficiently simple for students to complete individually. This requires no meaningful collaboration or negotiation to progress. We observed that most pairs worked in parallel without

engaging in shared reasoning. Most groups worked together in ways where they took turns to spin the wheel and noted down the sentence outcome, or where one student navigated the learning tool and the others watched without sparking much engaged discussion. This indicates that when a learning tool is simple and does not require distinct contributions from each student in the group, it may inadvertently facilitate individual learning rather than collaborative learning, hence undermining dialogue opportunities for critical reflection.

4.4 Phase 4: Post-implementation Reflection on the Gaps

In this phase, we outline how the teacher reflected on the gap between his planning and implementation. Below, we identified four points based on the teacher's account in the post-implementation interview.

4.4.1 Lack of Capacity for Facilitating Critical Reflection

In the post-implementation interview, the teacher recognised the dynamic of self-sufficient learning activities described in Moment 2 above and said he was "frustrated" about the lack of critical reflection. Therefore, he highlighted the importance of having enough time for dwelling and assessing group dialogues. He noted *"To conduct (formative) assessment during conversations requires time to dwell in those conversations...So it is important that activities are self-propelled and that (the teacher) can not only move around, but also have the time and capacity for it (formative assessment)."* His insight advocated for learning activities mediated by tools that are 'self-propelled', which means that students can continue and are willing to progress in learning activities with minimal requirement of teachers' intervention. Reducing the students' dependence on the teacher to move forward in learning activities allows the teacher to dedicate time to formative assessment and facilitation, actively participating in dialogues, and scaffolding critical reflection essential for critical AI literacy. This creates the necessary space for the teacher to not only provide technical explanations but also facilitate critical reflection.

4.4.2 Neutral Datasets Inhibited Critical Reflection

The teacher also reflected that the lack of personal or ethical resonance in the datasets used during the lesson inhibited critical reflection. The specific datasets used in the activity (such as the wheel analogy and Taylor Swift lyrics generator) provided no stakes that would motivate students to contest or critically evaluate the AI output's fairness, bias, or reliability. In the post interview, the teacher mentioned that activities with high-stakes scenarios could be more relevant to students' lives and stimulate critical reflection:

> "It doesn't matter (to the students) if a Taylor Swift lyric machine generates something transparent, fair, or reliable. You have to go to another forum ... a language model that is a decision support for the forced removal of children's custody ... that way (you spark dialogue)."

Without the involvement of controversial or personally relevant data to spark critical reflection, the triadic dialogue between student, teacher, and tool remained focused on technical concepts rather than empowered critical reflection in AI literacy. The transition from understanding how the AI system works to critically reflecting on its societal impact remained disconnected because the data lacked resonance and relevance to provoke a response.

4.4.2 Assessment of Different Learning Progress

Heterogeneous learning progress requires more time for formative assessment. In the post-observation interview, the teacher reflected that while some groups grasped the connection between the dataset size and the model's output quickly, others did not show any relevant reflection. This highlights that learning in a critical AI literacy classroom is highly heterogeneous. While dialogue with peers and tools facilitates learning, it does not guarantee that "signs of learning" are visible for every student simultaneously. The teacher believed that more exercises with the analogy might be needed to spark these signs of learning among other students. Reflecting on how to better scaffold this "dialogue with the tool," the teacher noted:

> "If you had more time... you could well have, for example, started with a small dataset and then afterwards have a larger, more nuanced dataset... They didn't have that kind of experience of 'OK if I change the size of the dataset, for example, that something happens.'"

Their reflection also demonstrates the teacher's use of formative assessment to adjust future instruction of similar learning activities, recognising that students' progress at different rates in critical AI literacy, just as they do in traditional subjects.

5. DISCUSSION

This study traced one teacher's journey across a full design-based research cycle, from co-design workshops through independent planning and classroom implementation. The findings elucidate how the teacher developed and implemented strategies for teaching critical AI literacy in a K-12 classroom. In what follows, we discuss the insights revealed from this teacher's journey to address the RQ (Section 5.1) and derive three potential future research directions for how co-design workshops, learning tools, and classroom activities should be developed to better support critical AI literacy in K-12 contexts (Section 5.2). Then in Section 5.3, we discuss the limitations of this study.

5.1 From Co-design to Classroom: What the Teacher's Journey Reveals About Supporting Critical AI Literacy

Our findings demonstrate that the co-design successfully supported the teacher in developing conceptual knowledge of critical AI literacy, planning for dialogic formative assessment, and articulating dual technical and critical learning goals. The brainstorming outputs and high-level lesson plan produced by the teacher during the co-design workshops demonstrated the teacher's contextualised knowledge of his students and classroom realities that the research team could not have independently produced, underscoring the value of co-design as a methodological approach for developing pedagogically situated strategies in collaboration with teachers (Casal-Otero et al. 2023). However, our co-design workshops were insufficient to enable the teacher to facilitate students' critical reflection when encountering classroom realities. By tracing the teacher's journey across the full design-based research cycle, we synthesise three challenges that hindered the support for teachers in teaching critical AI literacy.

5.1.1Co-design Builds Pedagogical Strategies but Not Facilitation Repertoire

The co-design workshops centred on formative assessment principles successfully supported the teacher in gaining knowledge about critical AI literacy and developing high-level pedagogical strategies for teaching it. This directly responds to the call for supporting teachers in developing their own pedagogical strategies for critical AI literacy (Velander, Otero, and Milrad 2024; Veldhuis et al. 2025; Yim 2024). The teacher operationalised critical AI literacy frameworks into learning objectives, naming AI fairness and gender bias as explicit goals, and articulated signs of learning he would look for through dialogue and written responses. Our findings indicate that aligning critical AI literacy frameworks with formative assessment principles, namely mapping learning objectives, learning activities, and assessment strategies, can effectively support teachers in designing high-level lesson plans and teaching strategies during co-design workshops (Biggs 1996; Bennett 2011; Wiliam 2018).

However, when the teacher needed to finalise the detailed lesson plan and implemented in the classroom, a gap emerged: conceptual knowledge of critical AI literacy did not make the teacher confident in facilitating critical reflection in the classroom, and teaching critical AI literacy requires facilitation strategies for heterogeneous learning progress. As our findings showed, the teacher retreated to technical explanations, a familiar domain where answers were relatively standardised, rather than engaging in critical AI literacy practices, which aim to empower learners to develop their own perspectives without imposing a single viewpoint (Ng et al. 2021; Veldhuis et al. 2025). This aligns with prior research demonstrating teachers' pain points in facilitating critical reflection on AI implications in classrooms (e.g., (Kenny et al. 2025; Antle et al. 2021). It suggests that it is important not only to develop teachers' conceptual understanding of critical AI literacy, but also to build a practical facilitation repertoire, in order to provide teachers with specific strategies for eliciting critical reflection, managing heterogeneous learning progress, and recognising non-verbal interaction as valid learning evidence (Wegerif 2013; Linden and Renshaw 2004; Major et al. 2018). Therefore, we stress the importance of exploring ways to facilitate critical reflection during different stages of the design and learning process by combining dialogic approaches (Veldhuis et al. 2025).

5.1.2 Tool Affordances Focus on Technical Understanding Rather Than Critical Reflection

Drawing on dialogic learning theory evidence (Wegerif 2013; Linden and Renshaw 2004), our study showed that the learning tool functioned as an active participant in the dialogue between teacher, student, and tool, and that its affordances determined what kind of learning that dialogue could produce. In our study, the tool effectively mediated technical understanding. Students' hands-on interaction with the word wheels triggered private dialogue for learning about probability and model weights without requiring explicit teacher facilitation. However, the tool's affordances leaned toward the conceptual knowledge of AI concepts instead of critical AI literacy. It actively supported technical understanding but did not aid the teacher's facilitation of critical reflection. This is consistent with other research's findings that current critical AI literacy tools predominantly focus on technical and conceptual skills without rigorously assessing students' outcomes and embedding affordances to scaffold critical reflection (Casal-Otero et al. 2023; Han and Han 2025; Druga, Otero, and Ko 2022). This absence of tools' affordances to support teachers in scaffolding critical reflection compounds the facilitation pressure on teachers identified in Section 5.1.1.

5.1.3 Lack of Relevance and Collaboration Undermined Critical AI Literacy Dialogue

Further, a challenge results from the design of learning activities. Two design choices in our study lacked collaboration and personal relevance, inhibiting the dialogue for critical AI literacy. First, the word wheel and text generation activities were sufficiently simple for students to complete individually, requiring no genuine collaboration or negotiation to progress, eliminating the peer dialogue for critical reflection. Second, the datasets used in the word wheel analogy and Taylor Swift

lyrics machine carried little personal or ethical stakes for students, providing them with little motivation to contest or critically evaluate. As the teacher reflected in the post-implementation interview, it simply did not matter to students whether a Taylor Swift lyric machine was fair or reliable. Without data carrying ethical weight and personal relevance, students were less motivated to engage in critical reflection and inquiry. These insights align with prior research emphasising that critical AI literacy dialogue requires activities grounded in learners' lived experiences and genuine ethical stakes (Veldhuis et al. 2025). For instance, Bilstrup et al. (2020) demonstrated that high-stakes ethical scenarios provoked students' critical reflection, which the Tayler Swift lyrics datasets in our study failed to stimulate. These design limitations indicate that the social and ethical conditions to provoke dialogue for critical AI literacy must be structurally embedded in activity design, rather than assumed to emerge from teacher facilitation alone.

5.2 Future Research Directions for Supporting Teachers in Critical AI Literacy

Building on the challenges identified in teacher-led classroom implementation, we propose potential areas for future work for the HCI and CCI communities to support teachers more effectively in teaching critical AI literacy in K-12 education, drawing on dialogic formative assessment and critical AI literacy research.

5.2.1 Supporting Teachers' Dialogic Facilitation Repertoires

Recent work in AI education emphasises that fostering critical AI literacy requires project-based and contextualised approaches, prioritising teachers' participation in the co-design of didactic proposals together with pedagogues and AI experts (Casal-Otero et al. 2023). However, as we discuss in Section 5.1.1, our findings indicate a persistent gap: co-design workshops support the development of teachers' pedagogical strategies, but not their facilitation repertoires. This observation is consistent with prior research, which reports that teachers feel underprepared to facilitate the complex ethical reflection required for critical AI literacy (Ng et al. 2021; Shokeen et al. 2025). This persistent gap points to a need for future research to develop dialogic facilitation repertoires that integrate dialogic learning approaches and formative assessment, equipping teachers with practical strategies for eliciting, sustaining, and facilitating critical reflection in classroom settings.

Drawing on Vygotsky's notion of private dialogue (1980) and Wegerif's conceptualization of peer-tool interaction as dialogue (Wegerif 2013), dialogic facilitation frameworks should support teachers in recognising students' internal reasoning triggered through tool interaction and externalising it into shared critical inquiry. Formative assessment principles extend this by framing such moments as signs of emerging understanding and critical reflection (Wiliam 2018). In our study, students' interaction with the tool triggered private reasoning about probability and model weights, and the teacher successfully transitioned some of this internal reasoning into classroom discussion. However, this facilitation did not extend toward critical reflection on AI's implications. By combining dialogic and formative assessment lenses, facilitation frameworks could support teachers in identifying these instructional entry points, thereby enabling scaffolding of students' critical reflection.

Building on this, dialogic facilitation frameworks should further support teachers in moving from teacher-led critical reflection to peer-led critical reflection, for instance, using inquiry techniques that encourage students to elaborate on each other's ideas, challenge AI outputs collectively, and connect AI concepts to their own lived experiences (Manalo 2020). This resonates with Veldhuis et al.'s (2025) argument for fostering critical AI literacy by combining dialogic, experiential, and hands-on approaches.

Importantly, developing a dialogic facilitation repertoire cannot be achieved through short-term co-design workshops or lessons. Drawing on dialogic learning literature emphasising the need for teachers to establish a sustained social infrastructure for inquiry (Major et al. 2018), we argue that facilitating critical AI literacy requires long-term cultivation of a classroom culture where students feel safe to express uncertainty, articulate values, and contest AI outputs. Future co-design workshops should therefore be structured as iterative cycles, providing teachers with opportunities to enhance specific facilitation techniques, implement them in the classroom, and return to collaborative reflection on what worked and what required adjustment (Bennett 2011; Anderson and Shattuck 2012).

5.2.2 Designing Critical AI Literacy Tools as Voices in Dialogic Space to Scaffold Critical Reflection

Building on the culturally responsive computing framework (Scott, Sheridan, and Clark 2015) that positions technology as a vehicle for critical reflection on socio-political issues, we argue that critical AI literacy tools must move beyond glass-boxing technical concepts to actively mediate classroom dialogue for ethical reasoning, socio-technical questioning, and collaborative reflection. This extends the argument by Collyer-Hoar et al. (2025) that using tools as part of structured activities can successfully stimulate dialogue and critical thinking. Researchers in HCI also have increasingly sought to foster critical AI literacy through dialogue-based activities, leveraging external scaffolds such as card games (Collyer-Hoar, Rubegni, and

Tomczyk 2026; Ali, Kumar, and Breazeal 2023) and speculative design worksheets and cards (Kenny et al. 2025).

Our findings, however, suggest that when a learning tool and its activities for critical AI literacy focus solely on technical mechanics and provide self-sufficient activities that do not require meaningful collaboration, it closes down dialogic space for critical reflection. The responsibility for achieving empowerment-oriented goals then falls entirely on the teacher. For instance, while the Engine Room tool in our study supported the teacher in explaining technical facts about AI systems, it did not provoke socio-technical questioning and reflection.

To address this issue, future interaction design should pivot away from treating tools as representations of technical AI concepts, and view them as active participants in activities (Wegerif 2013). This framing directly addresses the challenges highlighted by Dangol et al. (2025) who note that the rapid pace of technological advancement means existing AI literacy approaches can inadvertently foster misconceptions about the limitations of newer models. Dangol et al. (2025) argue that educators must equip children with a highly flexible understanding of AI. We contend that to cultivate this flexible understanding of AI, learning tools must act as active participants in the activity that structurally open, widen, and deepen the dialogic space (Wegerif 2013). Grounded in this framework, we propose design opportunities for tools to cultivate this flexible understanding of AI and actively co-facilitate critical reflection alongside teachers.

- **Opening and Deepening Dialogic Space Through Personal Relevance:** To open and deepen dialogic space, interaction design could allow students to explore socio-technical questions that are personally relevant and connect to their lived experiences. Discrepancy between how students personally experience a phenomenon and how the algorithmic systems present it spark reflection and dialogue (Liu, Dybboe, et al. 2026). For instance, activities could foreground ethical aspects by using authentic, high-stakes data, such as AI-driven voter behaviour prediction (Kaspersen et al. 2022) or systems designed to mitigate student loneliness, where more sensitive data enable more powerful systems (Bilstrup, Kaspersen, and Petersen 2020). Such contexts are complex and relevant to learners, fostering collaborative discussion, ethical reasoning, and engagement with bias, fairness, and societal implications.
- **Widening Dialogic Space through Collaboration and Multiple Viewpoints**: Wegerif (2013) posits that the pedagogical purpose of technology should evoke diverse voices and viewpoints in order to widen dialogic space. In the context of critical AI literacy, Veldhuis et al. (2025) similarly advocate for tools that expose learners to multiple viewpoints to form their own perspectives, rather than imposing a single viewpoint. Our findings, however, demonstrate that when a tool lacks structural need for collaboration, it diminishes students' motivation to contest AI's bias and societal implications, which reaffirms that critical reflection must be socially situated and personally meaningful (Scott, Sheridan, and Clark 2015). Future research should explore how to design tools that connect AI concepts to student lived experiences and foster meaningful collaborative dialogue and co-creation. For instance, Morales-Navarro et al. (2024) showed that positioning students as peer auditors created collaborative conditions and successfully engaged them in multiple viewpoints, stimulating critical reflection. By designing tools that actively provoke students to collaborate, encounter, articulate, and negotiate conflicting stakeholder perspectives, interaction design can foster students' critical engagement with AI's societal impacts.
- **Shifting Assessment from Representation to Presentation**: To better capture critical AI literacy, tools should align with Wegerif's notion of shifting from abstract representations of knowledge to presentations of engaged participation (Wegerif 2013, p.107). Interaction design of critical AI literacy tools could realise this by capturing and visualising not only students' digital creations (e.g. alternative dataset inputs or outputs), but also students' collaborative interaction (e.g., their reasoning and negotiation processes). Reflecting collaborative interaction presents students' engaged participation and provides teachers and peers with visible indicators of critical AI literacy and agency for formative assessment (Wiliam 2018). This multimodal indicators of critical AI literacy accommodates diverse classrooms and learning behaviours, ensuring that children remain active meaning-makers rather than passive recipients of technical knowledge (Antle et al. 2021; Iversen, Smith, and Dindler 2018).

Future research must explore how to design critical AI literacy tools that simultaneously scaffold students' critical reflections and support teachers to act as facilitators of in-depth

classroom discussions (Shor and Freire 1987). With these potential design opportunities, our study contributes a research agenda for the HCI and CCI communities to integrate dialogic approaches into interaction design to support teachers in facilitating critical AI literacy. It requires a teacher equipped with dialogic facilitation strategies and a tool that actively co-facilitate critical reflection though personal relevance, structured collaboration, and formative assessment.

5.3 Limitations

As a single-case design focusing on an in-depth analysis of one teacher in the classroom, this study does not seek statistical generalizability, nor does it present a universally successful blueprint for critical AI literacy. While our broader research efforts tracked multiple participating teachers across their implementations, this paper intentionally reports on a single exemplar case study to preserve the qualitative nuance and ecological validity of the trajectory from co-design to the classroom.

A key limitation of this work is that our findings did not support our preliminary hypothesis that short-term co-design workshops would successfully empower a teacher to facilitate a classroom for critical AI literacy. The classroom implementation revealed that the teacher still struggled to facilitate a dialogic classroom for critical AI literacy, frequently reverting to technical explanations. Furthermore, these results are compounded by contextual factors, including the specific classroom context and our choice of the Engine Room tool, which focused mainly on glass-boxing technical mechanics rather than facilitating reflection.

Because our study captured a complex and challenging initial cycle within an iterative DBR process and proved our preliminary hypothesis incorrect, the insights gathered cannot be ratified as concrete design implications. Instead, they should be interpreted as potential design possibilities for future work. These reveal a critical need for the HCI community to consider that future DBR cycles for critical AI literacy must be iterative and to shift the design focus toward creating tools that proactively facilitate dialogue for critical AI literacy.

6. CONCLUSION

This study examines how co-design could support a teacher in developing teaching strategies for critical AI literacy, and how these strategies translated into classroom practice. Tracing one teacher across a full DBR cycle, from co-design workshops to classroom implementation and post-implementation reflection revealed a pedagogical reality: while the co-design workshops effectively supported the development of critical AI literacy knowledge and high-level pedagogical strategies, it fell short of equipping the teacher to facilitate critical reflection in the classroom. Furthermore, we find that when tool affordances prioritise technical mechanics over personal relevance and meaningful collaboration, they hinder the teacher's ability to foster dialogic discussions.

To address these challenges, we propose two research directions: (1) support teachers by developing their dialogic facilitation repertoires for critical AI literacy; and (2) design tools as dialogic partners to scaffold critical reflection on AI's societal implications. By shifting the conversation from idealised implementation for critical AI literacy to the nuanced challenges of the everyday classroom, we provide the HCI and interaction design communities with a starting point for iteratively supporting critical AI literacy in formal education.

ACKNOWLEDGEMENTS

This work is a collaboration of Vindenscenter for digital Teknologiforståelse in Denmark, supported by the Novo Nordisk Foundation (Grant number NNF23SA0087324), the Villum Foundation (Grant number 54607), and the Lundbeck Foundation (Grant number R419-2022-1261). We sincerely thank the dedicated teachers and educators for their valuable participation and contributions to the project. We are also grateful for the reviewers' feedback to improve the paper.

7. REFERENCES

Alexander, Robin J. (2008), *Towards Dialogic Teaching: Rethinking Classroom Talk* (Fourth edition, England).

Ali, Safinah, et al. (2024), 'Constructing Dreams Using Generative AI', *Proceedings of the AAAI Conference on Artificial Intelligence*, 38/21: 23268–75, https://doi.org/10.1609/aaai.v38i21.30374.

Ali, Safinah, Vishesh Kumar, and Cynthia Breazeal (2023), 'AI Audit: A Card Game to Reflect on Everyday AI Systems', *Proceedings of the AAAI Conference on Artificial Intelligence*, 37/13: 15981–9, https://doi.org/10.1609/aaai.v37i13.26897.

Andersen, Lars Bo, et al. (2024), 'Infrastructuring Digital Literacy in K-12 Education: A National Case Study', *International Journal of Child-Computer Interaction*, 42: 100697, https://doi.org/10.1016/j.ijcci.2024.100697.

Anderson, Terry, and Julie Shattuck (2012), 'Design-Based Research: A Decade of Progress in Education Research?', *Educational Researcher*, 41/1: 16–25, https://doi.org/10.3102/0013189X11428813.

Antle, Alissa, et al. (2021), 'Opportunities and Scaffolds for Critical Reflection on Ethical Issues in an Online After School Biowearable Workshop for Youth', *FabLearn Europe / MakeEd 2021 - An International Conference on Computing, Design and Making in Education* (New York, NY, USA), FabLearn Europe / MakeEd 2021, https://doi.org/10.1145/3466725.3466762.

Azungah, Theophilus (2018), 'Qualitative Research: Deductive and Inductive Approaches to Data Analysis', *Qualitative Research Journal*, 18/4: 383–400, https://doi.org/10.1108/qrj-d-18-00035.

Bennett, Randy Elliot (2011), 'Formative Assessment: A Critical Review', *Assessment in Education: Principles, Policy & Practice*, 18/1: 5–25, https://doi.org/10.1080/0969594X.2010.513678.

Biggs, John (1996), 'Enhancing Teaching through Constructive Alignment', *Higher Education*, 32/3: 347–64, https://doi.org/10.1007/BF00138871.

Bilstrup, Karl-Emil Kjær, Magnus H. Kaspersen, and Marianne Graves Petersen (2020), 'Staging Reflections on Ethical Dilemmas in Machine Learning: A Card-Based Design Workshop for High School Students', *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (New York, NY, USA), DIS '20, 1211–22, https://doi.org/10.1145/3357236.3395558.

Bosch, Christina A., et al. (2025), 'Advancing Research on Equitable AI Education Through a Focus on Implementation: Insights from a Middle School Computer Vision Module Beta-Test', *Proceedings of the AAAI Conference on Artificial Intelligence*, 39/28: 29120–7, https://doi.org/10.1609/aaai.v39i28.35184.

Braun, Virginia, and Victoria Clarke (2006), 'Using Thematic Analysis in Psychology', *Qualitative Research in Psychology*, 3/2: 77–101, https://doi.org/10.1191/1478088706qp063oa.

Casal-Otero, Lorena, et al. (2023), 'AI Literacy in K-12: A Systematic Literature Review', *International Journal of STEM Education*, 10/1: 29, https://doi.org/10.1186/s40594-023-00418-7.

Charisi, Vicky, et al. (2020), 'Empowering Children's Critical Reflections on AI, Robotics and Other Intelligent Technologies', *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (25 October), 1–4, https://doi.org/10.1145/3419249.3420090.

Clarke, Sherice N., et al. (2023), 'Developing Shared Ways of Seeing Data: The Perils and Possibilities of Achieving Intercoder Agreement', *International Journal of Qualitative Methods*, 22: 16094069231160973, https://doi.org/10.1177/16094069231160973.

Collyer-Hoar, Gail, et al. (2025), 'Experts Unite, Kids Delight: Co-Designing an Inclusive AI Literacy Educational Tool for Children', *Proceedings of the 24th Interaction Design and Children* (23 June), 852–7, https://doi.org/10.1145/3713043.3731495.

Collyer-Hoar, Gail, Elisa Rubegni, and Ben Tomczyk (2026), 'Supporting Holistic AI Ethics Literacy Education Through Critical Reflection: Three Recommendations for Fostering Children's Ethical Growth', *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (13 April), 1–16, https://doi.org/10.1145/3772318.3791477.

Condens Insights GmbH (2025), *Condens*, Eng (Condens Insights GmbH), https://condens.io/about/.

Connelly, Luke, Karl-Emil Kjær Bilstrup, and Marianne Graves Petersen (2025), 'Beyond LLMs as Black Boxes: Activities and an Educational Tool Supporting Unplugged and Digital AI Learning Activities for K-12 Classrooms', *Adjunct Proceedings of the Sixth Decennial Aarhus Conference: Computing X Crisis* (18 August), 1–4, https://doi.org/10.1145/3737609.3747112.

Dangol, Aayushi, et al. (2025), 'Children's Mental Models of AI Reasoning: Implications for AI Literacy Education', *Proceedings of the 24th Interaction Design and Children* (23 June), 106–23, https://doi.org/10.1145/3713043.3728856.

Dangol, Aayushi, and Sayamindu Dasgupta (2023), 'Constructionist Approaches to Critical Data Literacy: A Review', *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (New York, NY, USA), IDC '23, 112–23, https://doi.org/10.1145/3585088.3589367.

Dindler, Christian, et al. (2023), 'DORIT: An Analytical Model for Computational Empowerment in K-9 Education', *International Journal of Child-Computer Interaction* (NLD), 37/C, https://doi.org/10.1016/j.ijcci.2023.100599.

Druga, Stefania, Nancy Otero, and Amy J. Ko (2022), 'The Landscape of Teaching Resources for AI Education', *Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1* (7 July), 96–102, https://doi.org/10.1145/3502718.3524782.

Flavell, John H. (1979), 'Metacognition and Cognitive Monitoring: A New Area of Cognitive–Developmental Inquiry.', *American Psychologist*, 34/10: 906–11, https://doi.org/10.1037/0003-066X.34.10.906.

Han, Ariel, and Shenshen Han (2025), 'Empowering Children's AI Literacy Through Co-Creating Stories with LLM', *Proceedings of the 24th Interaction Design and Children* (23 June), 994–9, https://doi.org/10.1145/3713043.3731520.

Hautea, Samantha, Sayamindu Dasgupta, and Benjamin Mako Hill (2017), 'Youth Perspectives on Critical Data Literacies', *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2 May), 919–30, https://doi.org/10.1145/3025453.3025823.

Holmes, Andrew G. D. (2020), 'Researcher Positionality - A Consideration of Its Influence and Place in Qualitative Research - A New Researcher Guide', *Shanlax International Journal of Education*, 8/4: 1–10, https://doi.org/10.34293/education.v8i4.3232.

Iversen, Ole Sejer, Rachel Charlotte Smith, and Christian Dindler (2018), 'From Computational Thinking to Computational Empowerment: A 21 st Century PD Agenda', *Proceedings of the 15th Participatory Design Conference: Full Papers - Volume 1* (20 August), 1–11, https://doi.org/10.1145/3210586.3210592.

Kafai, Yasmin B., and Chris Proctor (2022), 'A Revaluation of Computational Thinking in K–12 Education: Moving Toward Computational Literacies', *Educational Researcher*, 51/2: 146–51, https://doi.org/10.3102/0013189X211057904.

Kaspersen, Magnus Hoeholt, et al. (2021), 'VotestratesML: A High School Learning Tool for Exploring Machine Learning and Its Societal Implications', *FabLearn Europe / MakeEd 2021 - An International Conference on Computing, Design and Making in Education* (New York, NY, USA), FabLearn Europe / MakeEd 2021, https://doi.org/10.1145/3466725.3466728.

Kaspersen, Magnus Høholt, et al. (2022), 'High School Students Exploring Machine Learning and Its Societal Implications: Opportunities and Challenges', *International Journal of Child-Computer Interaction* (NLD), 34/C, https://doi.org/10.1016/j.ijcci.2022.100539.

Kenny, Sadhbh, et al. (2025), 'Beyond the Algorithm: Speculative Approaches to Critical AI Literacies with Diverse Youth', *Proceedings of the 24th Interaction Design and Children* (23 June), 89–105, https://doi.org/10.1145/3713043.3727052.

Linden, Jos van den, and Peter Renshaw, eds (2004), *Dialogic Learning: Shifting Perspectives to Learning, Instruction, and Teaching*, 1st ed. 2004 (Dordrecht), https://doi.org/10.1007/1-4020-1931-9.

Liu, Yu-Yu, et al. (2025), 'A Semi-Systematic Literature Review for Improving Assessment Goals, Practices, and Strategies for Promoting Computational Empowerment in K-12 Education', *International Journal of Child-Computer Interaction*, 45: 100766, https://doi.org/10.1016/j.ijcci.2025.100766.

Liu, Yu-Yu, Maja Dybboe et al. (2026), '"Let's Talk about Data'': Co-Designing Critical Data Literacy Tools for K-12 Education through

Dialogic Learning', paper delivered at CHI'2026, Barcelona, Spain, *Proceedings of the CHI Conference on Human Factors in Computing Systems.*, https://doi.org/10.1145/3772318.3791296.

Liu, Yu-Yu, Ane Vielandt Jensen et al. (2026), 'Teachers' Formative Assessment Practices for Computational Empowerment in K–12 Education', *International Journal of Child-Computer Interaction*, 48: 100805, https://doi.org/10.1016/j.ijcci.2026.100805.

Long, Duri, and Brian Magerko (2020), 'What Is AI Literacy? Competencies and Design Considerations', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (21 April), 1–16, https://doi.org/10.1145/3313831.3376727.

Luke, Allen (2014), 'Defining Critical Literacy.', in *Moving Critical Literacies Forward: A New Look at Praxis across Contexts* (USA), 19–31.

Major, L., et al. (2018), 'Classroom Dialogue and Digital Technologies: A Scoping Review', *Education and Information Technologies*, 23/5: 1995–2028, https://doi.org/10.1007/s10639-018-9701-y.

Manalo, Emmanuel, ed. (2020), *Deeper Learning, Dialogic Learning, and Critical Thinking: Research-Based Strategies for the Classroom* (Milton), https://doi.org/10.4324/9780429323058.

Morales-Navarro, Luis, et al. (2024), 'Youth as Peer Auditors: Engaging Teenagers with Algorithm Auditing of Machine Learning Applications', *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference* (17 June), 560–73, https://doi.org/10.1145/3628516.3655752.

Morales-Navarro, Luis, Daniel J. Noh, and Yasmin Kafai (2025), 'Building babyGPTs: Youth Engaging in Data Practices and Ethical Considerations through the Construction of Generative Language Models', *Proceedings of the 24th Interaction Design and Children* (23 June), 1021–6, https://doi.org/10.1145/3713043.3731525.

Ng, Davy Tsz Kit, et al. (2021), 'AI Literacy: Definition, Teaching, Evaluation and Ethical Issues', *Proceedings of the Association for Information Science and Technology*, 58/1: 504–9, https://doi.org/10.1002/pra2.487.

Salac, Jean, et al. (2023), 'Scaffolding Children's Sensemaking around Algorithmic Fairness', *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (19 June), 137–49, https://doi.org/10.1145/3585088.3589379.

Saldaña, Johnny (2021), *The Coding Manual for Qualitative Researchers* (Thousand Oaks, CA).

Schaper, Marie-Monique, Laura Malinverni, and Cristina Valero (2020), 'Robot Presidents: Who Should Rule the World?: Teaching Critical Thinking in AI through Reflections upon Food Traditions', *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (25 October), 1–4, https://doi.org/10.1145/3419249.3420085.

Scott, Kimberly A., Kimberly M. Sheridan, and Kevin Clark (2015), 'Culturally Responsive Computing: A Theory Revisited', *Learning, Media and Technology*, 40/4: 412–36, https://doi.org/10.1080/17439884.2014.924966.

Shokeen, Ekta, et al. (2025), 'What We Talk About When We Talk About K-12 Computational Empowerment Assessment.', Research Paper, International COmputin Education Research (ICER).

Shor, Ira, and Paulo Freire (1987), 'What Is the "Dialogical Method" of Teaching?', *Journal of Education*, 169/3: 11–31, https://doi.org/10.1177/002205748716900303.

Shrivastava, Vaishnavi, et al. (2024), 'Is a Sunny Day Bright and Cheerful or Hot and Uncomfortable? Young Children's Exploration of ChatGPT', *Nordic Conference on Human-Computer Interaction* (13 October), 1–15, https://doi.org/10.1145/3679318.3685397.

Solyst, Jaemarie, et al. (2025), 'RAD: A Framework to Support Youth in Critiquing AI', *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1* (12 February), 1071–7, https://doi.org/10.1145/3641554.3701966.

the Danish Ministry of Education (2026), 'Teknologiforståelse.', https://emu.dk/grundskole/teknologiforstaaelse?b=t5.

Touretzky, David, et al. (2019), 'Envisioning AI for K-12: What Should Every Child Know about AI?', *Proceedings of the AAAI Conference on Artificial Intelligence*, 33/01: 9795–9, https://doi.org/10.1609/aaai.v33i01.33019795.

Velander, Johanna, Nuno Otero, and Marcelo Milrad (2024), 'What Is Critical (about) AI Literacy? Exploring Conceptualizations Present in AI Literacy Discourse', in Anders Buch, Ylva Lindberg, and Teresa Cerratto Pargman, eds, *Framing Futures in Postdigital Education*, Postdigital Science and Education (Cham), 139–60, https://doi.org/10.1007/978-3-031-58622-4_8.

Veldhuis, Annemiek, et al. (2025), 'Critical Artificial Intelligence Literacy: A Scoping Review and Framework Synthesis', *International Journal of Child-Computer Interaction*, 43: 100708, https://doi.org/10.1016/j.ijcci.2024.100708.

Vygotsky, L. S. (1980), *Mind in Society: Development of Higher Psychological Processes*, ed. Michael Cole et al. (n.p.), https://doi.org/10.2307/j.ctvjf9vz4.

Wegerif, Rupert (2013), *Dialogic: Education for the Internet Age* (0 edn, n.p.), https://doi.org/10.4324/9780203111222.

Wiliam, Dylan (2018), *Embedded Formative Assessment* (Second edition, Bloomington, Indiana).

Williams, Randi, et al. (2024), 'Doodlebot: An Educational Robot for Creativity and AI Literacy', *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (11 March), 772–80, https://doi.org/10.1145/3610977.3634950.

Yim, Iris Heung Yue (2024), 'A Critical Review of Teaching and Learning Artificial Intelligence (AI) Literacy: Developing an Intelligence-Based AI Literacy Framework for Primary School Education', *Computers and Education: Artificial Intelligence*, 7: 100319, https://doi.org/10.1016/j.caeai.2024.100319.

A. APPENDIX

A.1 Overview of the co-design workshop activities

Workshop 1		Workshop 2	
(Critical AI Literacy and Computational Empowerment)		(Learning Evidence for Formative Assessment)	
30 min.	Introduction of research project	5 min	Introduction
20 min	Introduction: DORIT model	20 min	Hands-on with the Engine Room Learning tool
60 min	Hands-on with the Engine Room Learning tool	15 min	Recap ideas of lesson plans from Workshop 1
15 min	Introduction to teaching examples of critical reflection	60 min	Developing lesson plans
30 min	Technical insights on LLMs	35 min	Principles of formative assessment
45 min	Drafting ideas for lesson plans	15 min	Teachers' presentation of lesson plans
20 min	Presentation of ideas		

A.2 Pre-implementation Semi-structured Interview Guideline

(1) Can you walk us through the class you are going to teach for this lesson plan? → Goal: To understand the overall learning activities in a linear timeline.

(2) What is the main learning outcome you would like students to achieve by the end of the lesson?

(3) How do you plan to make sure they achieve this outcome during the lesson? How do you observe or measure it? → Goal: To understand their ideas of assessment in the lesson plan.

(4) On a scale of 1-10 (1 is the least confident, 10 is the most confident), how confident do you think you are in executing this plan? Why? → Goal: To evaluate the confidence that we provide for teachers through workshops and to gauge if we should provide additional support before they start teaching, if necessary.

A.3 Post-implementation Semi-structured Interview Guideline

(1) What stands out the most in this lesson? → Goal: To understand their immediate impression of the lesson plan implementation.

(2) How did you think students achieved the intended learning outcome through the lesson? How did you observe or measure it?

(3) Since students may showcase different levels of understanding of the intended learning outcomes, what are your thoughts on this? How do you observe their differences in learning outcomes? → Goal: To understand their practices of assessment in the learning activities.

(4) Have you adjusted your teaching according to how you perceived students' reactions and behaviours? → Yes / No? Can you explain why and how? → Goal: To understand the teacher's formative assessment practices.

(5) If you would like to do this again next time, what additional support do you need to make the teaching and learning experience better? How would you change it?

A.4 The Initial Codebook for Data Analysis

Parent Code	Child Code and Definitions
Formative Assessment: The codes were developed based on Wiliam's embedded formative assessment strategies (Wiliam 2018)	• Eliciting evidence: This focuses on actual events where the teacher gathered insights and understanding from students through various assessment techniques such as inquiry, dialogue, and reflection in the learning process. • Planned formative assessment: This refers to teachers' plans and strategies for engaging students in reflective practices during learning activities to elicit learning evidence. • Feedback exchange: This involves the exchange of feedback between teachers and students to enhance understanding and learning. • Goal orientation: This focuses on leading students to the learning goals, whether through sharing with students directly, or the teacher's reflection on how he facilitates the learning goals. • Assess prior knowledge: The teacher evaluates students' existing knowledge to tailor instruction effectively. • Responsive teaching: The teacher showed or reflected on how they adapted teaching methods based on their assessment of student learning, needs, and feedback.
Dialogic Learning: The codes were developed based on prior work (Wegerif 2013; Flavell 1979)	• Teacher-led scaffolding: Topics are chosen for practicality and feasibility for student activities. • Learner-tool dialogue: It encompasses learning dialogue, whether written or spoken, provoked or mediated by the tool/materials. • Peer dialogue: This encompasses interactions among students that foster collaborative learning and critical thinking. • Reasoning Dialogue: Reasoning dialogue includes exchanges and discussions among students that elaborate on learning concepts or reasons for their actions, choices, and arguments. • Metacognitive talk: This involves student dialogue or learning evidence that shows students' metacognitive abilities, including reflection on their actions and learning.
Critical AI Literacy: The codes were developed based on CE and AI literacy literature (Veldhuis et al. 2025; Long and Magerko 2020; Iversen, Smith, and	• Understanding of AI technical mechanisms: Instance related to comprehension of AI's underlying technical principles and mechanisms. • Construction: Students participated and collaborated in the learning activities to understand the construction of AI mechanisms • AI use: Understanding how to use AI in various contexts.

Dindler 2018; Dindler et al. 2023)	• Societal implications: Reflection on how AI's impact on fairness, values, possibility, hope, solutions, and collective trust in relation to society. • Critical reflection on AI: Reflection of pitfalls about AI, such as bias, unreliability, security, threats, inaccuracy, and misinformation. • Agency/Ownership: Students' confidence, feelings of achievement, and agency ownership in harnessing AI use and its implications in their lives.

A.5 The Inductive Codes

Emergent Codes	**Code Definition**
Constructionism	Hands-on construction, lived experience, and personally relevant experience promote learning, problem-solving, and engagement.
Lecturing explanation	This collection includes teachers' lecturing explanations of various concepts related to critical AI literacy.
Teacher's reflection for improvement	The teacher reflected on how his pedagogical practices could improve the learning process.
Performance assessment	This includes instances where teachers assessed student understanding through observation of in-class actions rather than verbal dialogue. It emphasises interactive tasks and the use of tools to assess learning outcomes.
Teacher concerns	It captures teachers' concerns regarding teaching pressures and the effectiveness of classroom dialogues.
Affective dialogue	Dialogue that fosters positive emotions, curiosity, and active participation among learners.

A.6 The findings correspond to each phase of the teacher's journey

The teacher's journey phases	Co-design workshops	Independent lesson planning	Classroom implementation	Post-implementation reflection
Context	3.3.1 The teacher participated in the co-design workshop on critical AI literacy frameworks and the learning tool, formative assessment principles and learning evidence	3.3.2 In a semi-structured interview, we interviewed the teacher regarding his finalized lesson plan and discussed his pedagogical strategies for teaching critical AI literacy, including the signs of learning he was looking for.	3.3.2 The teacher Implemented the lesson plan at a 7th grade classroom with 16 students (10 male and 6 female). Students worked in pairs organized by the teacher.	3.3.2 In a semi-structured interview, the teacher reflected on his implementation, the students' learning outcomes versus his expectations, and the effectiveness of his formative assessment strategies.
Gains	4.1 Developed high-level pedagogical strategies for teaching critical AI literacy. Felt confidence and interest.	...	4.3.1 Technical Understanding Through Tool Interaction and Teacher-led Inquiry Scaffolding.	...
Challenges	...	4.2 • Tension Between Technical and Critical Goals • Quiet Classroom Culture • Reflection on the learning tool's focus on technical concepts	4.3.2 • Retreat to Technical Explanation • Self-sufficient Activities Undermined Collaboration and Dialogue Opportunities	4.4 • Lack of Capacity for Facilitating Critical Reflection • Neutral Datasets Inhibited Critical Reflection • Assessment of Different Learning Progress

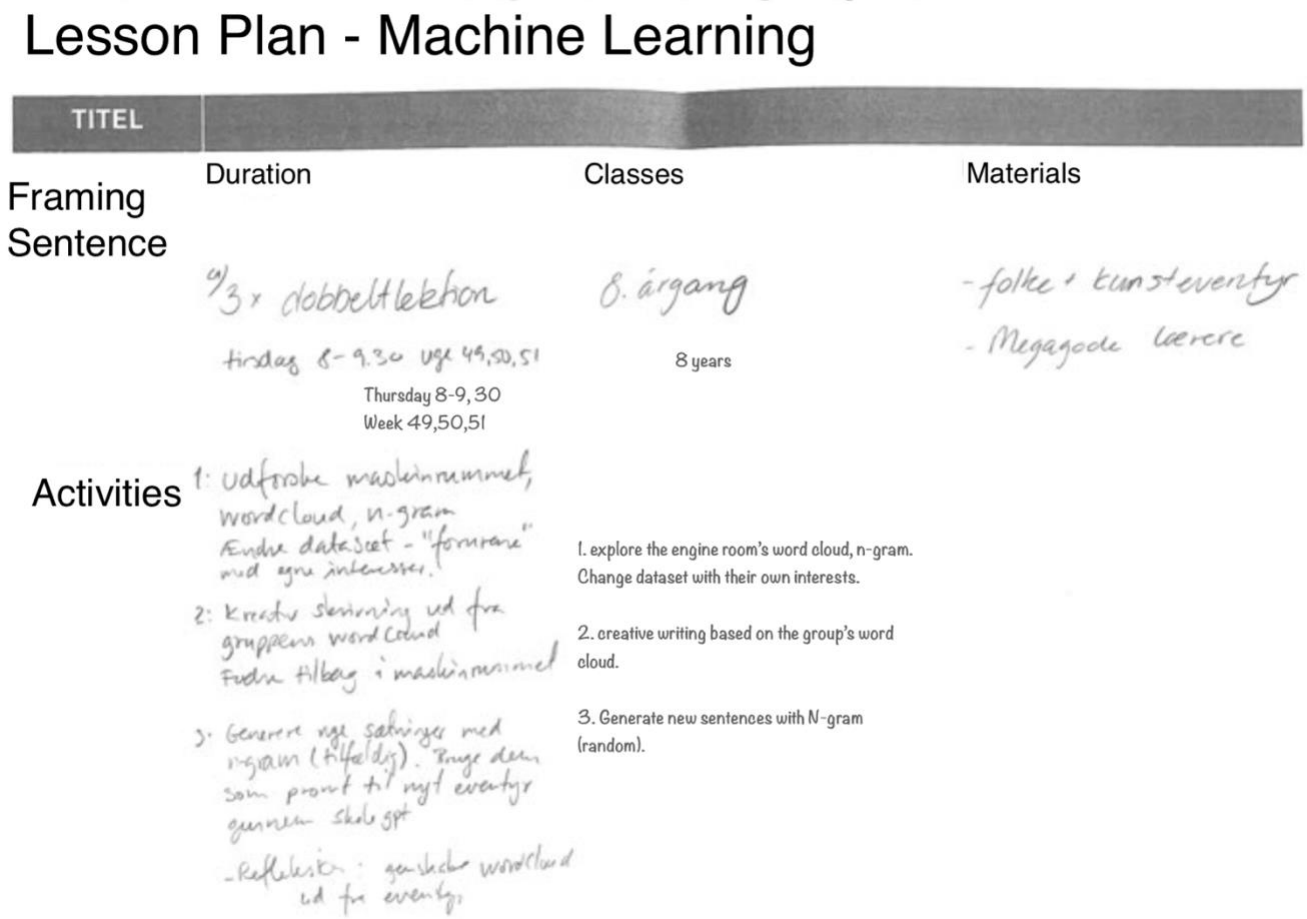
A.7 The inspirational cards

Among the inspirational cards provided to teachers during the co-design workshops, the cards shown in A.7 are those selected by the focal teacher, thereby indicating his pedagogical interests for his classroom practice. As indicated by the five cards on the left, the AI concepts that attracted his attention include N-grams, text generation, and understanding how an AI model was created. In addition, as indicated by the four cards on the right, he showed an intention to encourage students' critical reflection on the consequences of using machine learning technologies, as well as on the underlying intentions and purposes embedded in these systems.

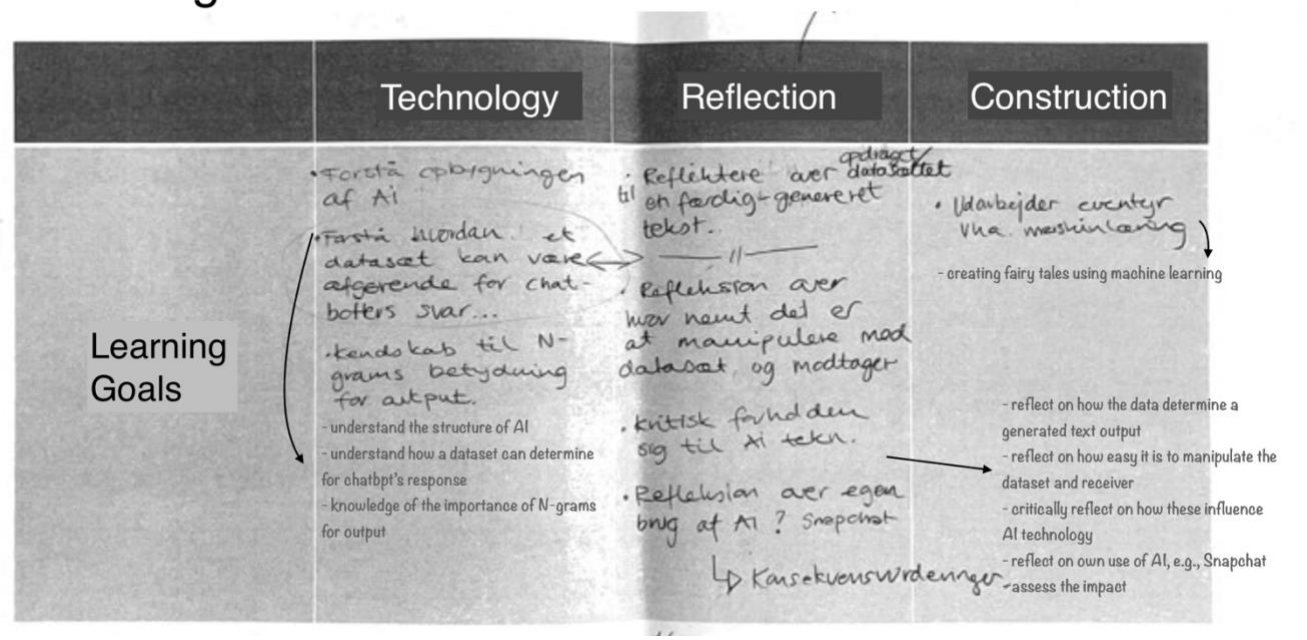


A.8 The figures display the lesson plan templates completed by the focal teacher during Day 2 of the co-design workshop.

(a) Plans for duration, grade level, materials, and learning activities.



(b) Plans for learning objectives, including technical understanding, critical reflection, and construction.



(c) The focal teacher's plan for formative assessment, including learning evidence and facilitation questions.

Formative Assessment

	Technology	Reflection	Construction
Signs of learning	Bruger ~~disse~~ ordene: • n-gram • grådig • vægte tilfældighed • tilfældighed • data (sæt) • wordcloud Use the words: - n-grams - coincidence - dataset - wordcloud	• give bud på "buzz-words" i wordcloud bag historie • datagrundlag bestemmer meget	• hvor "nyt" er deres eget eventyr? • hvor stor grad har egne data overtaget historien
Questions that can be asked of students to clarify signs of learning	• hvordan bli'r det mere åbent/andeleder/grøn • hvordan bliver det mere "klassisk eventyr"?	★ Hvordan kan i som mennesker påvirke MLT (sprogmodeller) ★ fordele/ulempe ved MLT ★ Hvordan påvirker det dig/fællesskabet/samfundet?	- hvorfor afviger det eventyr så meget/ikke fra originalen? Why does your fairy tale differ so much from the original?

KROPSSPROG! AHA' OPLEVELSER! body language! The Aha experiences